



WORKING PAPER ON PUBLIC DATA AVAILABILITY, INTEROPERABILITY AND REUSABILITY

Agustí Cerrillo I Martínez. Universitat Oberta de Catalunya (coord.)

Ramon Galindo Caldes. Universitat Oberta de Catalunya

Javier Miranzo Díaz. Universitat Oberta de Catalunya-Universidad de Castilla-La Mancha

Bruno González. Oficina Antifrau de Catalunya

Grant Agreement number: 101038790 — CO.R.E — ISFP-2020-AG-CORRUPT

This document was funded by the European Union's Internal Security Fund — Police.
The content of this document represents the views of the author only and is his/her sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains



1. Fundamentals of public contract data availability, reusability and interoperability.....	4
1.1 Transparency in public procurement and the fight against corruption in times of emergency	4
A. Transparency in public procurement.....	4
B. Emergency public procurement.....	5
C. Transparency in emergency procurement	7
1.2 The transparency mechanisms in public procurement.....	9
A. The proactive dissemination of public contract information.....	10
B. Access to public contracting information.....	11
C. The re-use of public procurement information.....	11
1.3 Data accessibility, Interoperability and reusability.....	12
A. Data availability	12
B. Data interoperability	23
C. The reusability of information	26
2. Data on contracts, public officers and companies	27
2.1 Availability of the data	27
2.2 Peculiarities of data relating to emergency contracts	30
2.3 Data sources on non-contractual information.....	31
A. Data sources on public officials	33
B. Contractor Data Sources	36
3. Relevant experience in the use of digital tools for the prevention of corruption	40
3.1 Previous experience in the processing and management of information	40
3.2 Corruption Indicators	49
3.3 Indicators in emergency contracts	52
3.4 The identification of relevant information	53
3.5 The possible implementation of AI systems in the creation of indicators and detection of irregularities	54
4. Case study: the availability, quality, reusability and interoperability of data on public procurement in Catalonia	55
4.1 Introduction	55
4.2 Context: procurement activity datafication in the Catalan public sector	56
4.3 Public procurement data sources in Catalonia	60



A. The Public Procurement Services Platform (PSCP)	60
B. The Public Registry of Contracts (RPC).....	73
C. The Open Data Portal of the Generalitat de Catalunya.....	79
5. Bibliography	97
Appendix 1 PSCP data fields list.....	102
Appendix 2.A RPC data list.....	109
Appendix 2.B RPC data field list (search engine)	115



1. Fundamentals of public contract data availability, reusability and interoperability

1.1 Transparency in public procurement and the fight against corruption in times of emergency

A. Transparency in public procurement

Transparency in procurement consists in citizens knowing what is happening inside public administrations when public contracts are being bid for, awarded and executed (Cerrillo i Martínez, 2012b).

According to the expression coined at the beginning of the last century by Filippo Turati, member of parliament, in his speech in the Italian House of Deputies on 17 June 1908, transparency could make the public administration into a glass house. To this effect, sunlight is the best disinfectant (Brandeis, 1914).

Transparency contributes to democratic quality and is a mechanism of good administration. Furthermore, transparency facilitates public administration accountability and control.¹ What is more, transparency is also an effective means to prevent conflicts of interest to combat corruption. In effect, when a representative or a person in the service of the public administration is or can be under public scrutiny, their behaviour will be more integral.

In the last decades, it has been pointed out that transparency can contribute to strengthening public integrity and thus prevent conflicts of interest and combat corruption in public procurement (Kaufmann, 2005). Furthermore, various empirical studies have concluded that higher levels of information mean reduced levels of corruption (Rose-Ackerman, 2004). Oppositely, a lack of transparency, opacity, has been identified as one of the main conditions for corruption to emerge (Kaufmann & Bellver, 2005), although transparency alone is not sufficient (Lindstedt & Naurin, 2010), but must be accompanied by effective control measures and accountability (OECD, 2017). Ultimately, transparency does not only mean a reduction in corruption because it prevents it from happening, but it also allows citizens to become involved in the fight against it (Merloni & Ponti, 2010, 403).

Transparency has a clear impact on preventing conflicts of interest and combating corruption in public procurement. Transparency is one of the mechanisms through which integrity in public procurement is channelled (OECD, 2009, 22). To this effect, adequate decision making is facilitated and a fair and equitable treatment of potential bidders is stimulated, guaranteeing equity among them and competitiveness among their proposals. As summarised in the Court of Justice of the European Union's (CJEU) *Evropaïki Dynamiki* sentence of 12 March 2008 "the principle of transparency is respected, which is its corollary, the essential objective of which is to guarantee that there is no risk of favouritism and arbitrariness on the part of the awarding body. It means that all the conditions and modes of the tender procedure are formulated clearly, precisely and unequivocally in the tender notices or in the terms and conditions (article 144). The Court itself, in its sentence of 12 March 2013 on the *eVigilo* affair, case C-538/13, affirmed that "the essential objective of the obligation of transparency, which is its corollary, is to guarantee that there is no risk whatsoever of favouritism or arbitrariness on the part of the awarding body with respect to certain bidders or offers" (section 34).

¹ In the following pages, we will use the term public administrations as an umbrella to refer to the state, the regional and local authorities, the public law bodies and the associations made up of one or more of these adjudicating powers, according to the provisions of article 2 of Directive 2014/24/UE of the European Parliament and of the Council, of 26 February 2014, on public procurement, and by which Directive 2004/18/CE is repealed.



Transparency must be a particularly relevant value when other mechanisms for strengthening integrity are not available or the existing conditions do not allow all the expected effects to be produced or have the expected impact. This is what happens in situations of emergency when the public administrations must resort to more flexible and streamlined procedures (for example, in the case of urgent or emergency contracting to provide supplies and services). At moments like this, the risks for public integrity are increased (OCDE, 2020); (Group of States against Corruption (GRECO), 2020).

B. Emergency public procurement

Directive 2014/24/UE of the European Parliament and of the Council of 26 February 2014 on public procurement, which repeals Directive 2004/18/CE (henceforth, Directive 2014/24/UE), stipulates that when there are urgent overriding reasons stemming from events that the awarding body could not have foreseen, the negotiated procedure can be used without prior publishing. The use of this procedure is subject to the concurrence of three accumulative conditions (sentence CJEU Commission/German, of 15 October 2009, case C-275/08):

- existence of an event that the awarding body could not have foreseen;
- existence of an extreme emergency that has made compliance with the general deadlines impossible;
- a causal relationship between the unforeseeable event and the extreme urgency.

In Portugal, article 24 of the Decree-Law No. 18/2008, the Public Contracts Code (*Código dos Contratos Públicos, CCP*), states that to the extent strictly necessary and for reasons of imperative urgency resulting from events unforeseeable by the contracting entity, when the deadlines inherent to the other procedures cannot be met, and provided that the circumstances invoked are not, in any case, attributable to the contracting entity, contracts can be directly awarded. During COVID-19 pandemic, the rules applicable to the public contracts and the effectiveness of any public procurement by the Portuguese public administration were modified in March 2020. Since the COVID-19 imposed an enormous risk on the public health system functioning, Decree-Law no. 10-A/2020, of 13 March, and Law no. 1-A/2020, of 19 March, approved exceptional and temporary measures for the public contracting bodies regarding the preparation and realization of public procurement. These urgent changes have also authorized extraordinary expenditure considering the need of an immediate response to the epidemiological situation caused by the coronavirus SARS-CoV-2 and the COVID-19 disease in Portugal. These two legal documents alter procedures and measures for the public procurement regulated by *Código dos Contratos Públicos* (Public Contracts Code) among other acts.

The Spanish Law on Public Sector Contracts (LPSC) has transposed this decision, stipulating that in cases where the public administration must act immediately due to catastrophic events, to situations that pose a serious threat or to needs that affect national defence, contracts can be entered into freely or the execution of the works, service or supply ordered, without the obligation to process the procurement file (article 120.1).² To this effect, when these circumstances concur, the public administrations can enter into contracts without being subject to the stipulated formal requirements.

² In its report 22/2020 on the rendering of accounts for urgently filed contracts to the Council of Ministers, the State's Administrative Contracting Advisory Board excludes the provisions of this article for stable trading companies and state public sector foundations in public hands.



During the COVID-10 health crisis, the European Commission Recommendations on the use of the public procurement framework in the emergency situation related to the COVID-19 crisis, stipulated that to respond effectively to this situation the adjudicating powers must be able to substantially reduce the deadlines in order to accelerate open and restricted procedures. Furthermore, when this measure is insufficient, they must be able to anticipate a procedure negotiated without publication and even resort to the direct adjudication of a pre-selected economic operator when they are the only one that can provide the required supplies in accordance with the technical conditions and time constraints dictated by the extreme urgency.

In this line, Spanish Royal Decree-Law 7/2020, of 12 March, by which urgent measures are adopted in response to the economic impact of COVID-19 (modified by Royal Decree-Law 9/2020, of 27 March), stipulates that the actions directly or indirectly adopted to cope with COVID-19 meet the criteria set out in the LPSC and, therefore, emergency contracting would be instated, without application of the regulation on the guarantees provided for in the LPSC should paying by credit to an account be required for preparatory actions needed to be taken by the contractor (article 16).

In Portugal, according to Article 2.2 Decree-Law No. 10-A/2020, the expenditure limit related to the epidemiological emergency, or the “simplified direct adjustment”, is elevated from €5,000.00 to €20,000.00 for those awards dispensed with the ordinary procedure as Articles 128 and 129 of the CCP determine. In this regard, the preexisting limitations seen in paragraphs 2 to 5 of Article 113 of the CCP, concerning the entities invited to submit a proposal in the prior consultation or direct award procedures, are not applicable in the context of the COVID-19 pandemic. Furthermore, awards shall be communicated by the contracting authorities to the members of the Government responsible for the area of finance and for the respective sectoral area through the public procurement portal. Article 3.1 (a) of Decree-Law No. 10-A/2020 states that the expenditure on contracts or the acquisition of services by the public administration does not require an administrative authorization if the purpose is to conduct studies, elaborate specialized opinions, projects, or any consultancy service. The member of the Government responsible for the sectoral area is the one in charge of the resolution.

In the COVID-19 crisis, the possible national divergences will probably not be due to the concept of extreme urgency, since the health crisis caused by COVID-19 has been accepted without a doubt as an enabling situation for the use of this type of extraordinary procedure³. However, for a contract to benefit from this exception, it had to be specifically linked to the exceptional situation caused by it, and this requirement has caused more problems in practice. For example, in Spain, the Independent Office for the Regulation and Supervision of Contracting (OIRESCON), has warned that some contracts entered into by the public sector through the emergency procedure do not seem to present this relationship with respect to the emergency: concession for the operation of sports facilities services, contracting for the operation of a public facility car park, supply of materials for works on trails or paths, contracting of user satisfaction survey services for a public service (Oficina Independiente de Regulación y Supervisión de la Contratación, 2020). Similar conclusions have been reached by other sub-central supervisory bodies, such as the Audit Office of

³ See the European Commission Guidelines on the use of the public procurement framework in the emergency situation related to the COVID-19 crisis (2020).



Catalonia, which in its report 32/2021 reproached the regional Administration that several files analysed had contract objects that deviated from the measures aimed at protection against COVID-19⁴.

This suggests that, for the analysis and control over the legality of the emergency contracts entered into, an indicator that assesses the degree of linkage of the contract with the health emergency could be relevant. For this, it is possible to count on the CPV lists linked to COVID-19 contracting carried out by previous studies (Abdou, Czibik, Tóth, & Fazekas, 2021), or to consider the use of natural language processing tools to identify unrelated nomenclature in suspicious contract objects. Additionally, other studies have shown the usefulness of assigning any label that includes the term 'COVID' a *dummy* variable, which allows the clear separation of general contracting during the period from the specific acquisition of COVID (Page, McClelland, Oyenubi, Bridgman, & Kollamparambil, 2021).

Finally, the lack of a procedure filed for awarding contracts in situations of emergency grants the adjudicating powers more flexibility and an obvious reduction in deadlines. However, a lesser degree of transparency during the preparation and awarding of these contracts can also be assumed, forcing a later increment in the information disseminated or accessible in relation to the contracts awarded, their execution or modification which, apart from facilitating control and accountability, also enables irregularities and cases of corruption to be detected.

C. Transparency in emergency procurement

Crisis situations facilitate the emergence of cases of corruption (Schultz & Søreide, 2008); (Mihaly Fazekas, Nishchal, & Søreide, 2021). As Rose-Ackerman and Palifka observe, “conflict and its aftermath always generated corrupt incentives and gave national and international stakeholders excuses for turning a blind eye to corruption” (Rose-Ackerman & Palifka, 2019). There are some forms of corruption that are prevalent in these situations (Independent Commission Against Corruption, 2020). To this effect, in the last decades numerous cases of corruption linked to the management of crisis and emergency situations have been identified (for example, in the case of Hurricane Katrina in 2005, and the HIV and Ebola pandemics in East Africa in 2014 (U4. Anticorruption Research Centre, 2015); (OCDE, 2020).

To this effect, while it is evident that transparency emerges as a required mechanism for the prevention of conflicts of interest and corruption in general, it has a specific manifestation in public procurement and, in particular, in emergency situations.

The guarantee of transparency is precisely one of the keys for ensuring that public resources are used efficiently and effectively so that public administrations can deal with emergencies and, ultimately, create trust in how the public institutions respond to them (World Bank Group, 2020b).

Transparency acquires a special significance in the context of the public administrations response to emergency situations, not only to make known the decisions they are making and the resources they are using, but also to communicate why certain decisions are being taken so that they can be subject to public scrutiny. The special circumstances in which public decisions are taken in times of crisis grant transparency a special value. As the president of the Council of Europe's Group of States against Corruption has stated (GRECO), “the need for regular and reliable information from the public institutions is crucial in times of

⁴ Audit Office of Catalonia. Report 32/2021. Administrative contracting derived from the covid pandemic year 2020, resolutions 962/xii and 963/xii of the parliament (2020).



emergency. This refers to the propagation of the risks of the pandemic in themselves, as well as to the emergency measures taken in response to them” (Marin Mrčela, 2020).

However, as already mentioned, in emergency situations transparency in public procurement diminishes considerably during the phases of preparation and awarding of the contract because the filing of a procedure and the publication of a tender notice is not required.

In this line, article 120 LPSC stipulates that accounts must be rendered of the agreements related to emergency contracting adopted by the Council of Ministers or the governing council of the Regional Governments within a period of 30 days. However, this is notwithstanding that the public administrations must publish the contracts awarded and formalised in the buyers profile within a period of a fortnight, as per the general provisions of articles 151 and 154 in relation to 63 LPSC. Nonetheless, as stated by the Administrative Contracting Advisory Board in their 22/2020 report on the rendering of accounts of contracts processed in emergency situations to the Council of Ministers, “the required publication of these contracts must be dictated, however, by what is applicable, considering that there is no preliminary process with the usual procedures”.

The Informative Note on the emergency processing of contracts by which measures to fight COVID-19 are instrumentalised, drawn up by the presidency of the Consultative Committee, recalls that although “the immediacy of the action that gives recourse to emergency processing means exemption from the prior publication of the tender notice”, this does not mean that “exceptions for these contracts in terms of the publishing of the acts of awarding and formalising in the buyer profile of the awarding body with respect to the system of openness generally provided for in articles 151.1 and 154.1 of the LPSC. Openness in these cases must be constrained by applicability, considering that there is neither a preliminary process with the regular procedures, nor is there “any provision that makes an exception for publication within the corresponding official deadlines”.

In Portugal, based on article 127 of the CCP, the publication of any concluded procurement on the public contract portal following a direct agreement remains mandatory. However, it is not a condition for the effectiveness of the respective contract, which may produce all its effects immediately after the contract adjudication. Payments in advance are allowed without the assumptions as provided for article 292 CCP, whenever the guarantee of the availability of goods, services, and the resulting acts assured by the economic operator.

However, in practice, compliance with these obligations has been rather irregular among the different public administrations, some of doing so as stipulated and others not meeting the deadline requirements. In Spain, according to the report compiled by the Independent Procurement Regulation and Supervision Office, “there is asymmetric publication”, having identified that “of the 5,922 contracts processed by the emergency channel linked to COVID-19 published on the Regional Procurement Platforms, 1,667 contracts do not feature among the contracts published by PLACSP in their open data”. In particular, according to OIRESCON, “a series of publications have been identified that have anomalies or unclear concepts, which impedes adequate identification of the individual publication of each contract and, consequently, the information that must be given in the corresponding notice. To this effect, and given the growing and exorbitant volume of emergency procurement due to the COVID-19 crisis, special care and detail is recommended in the advertising of these contracts, given that the later verification of their justification and adaptation to the contracts19 regulation depends on this information, as do the rights citizens may have in relation to their awarding, such as the filing

of an appeal under contentious-administrative jurisdiction” (Oficina Independiente de Regulación y Supervisión de la Contratación, 2020).

In Portugal, due to the enormous risk for the lack of transparency in public contracting during the COVID-19 emergency, in June 2020 the Portuguese Court of Auditors published a report alerting the public administration to the importance of using open data with the objective of reducing opacity for awards. The document refers to the transparency procedures that “must be strictly observed, in an open data regime, so that inspections can be carried out and the due social control can be exercised” (Tribunal de Contas, 2020). Besides the general guidelines, the court also was specific on the production of register, data and information. The document strengthens the need for the “Adequate parameterization of the information systems that support the application of the measures in order to prevent errors and delays and to enable the collection and processing data transparency”. It calls the attention to the necessity of buttressing the computer network security and authentication systems users in particular for decision-makers (Tribunal de Contas, 2020).

1.2 The transparency mechanisms in public procurement

Transparency currently has two manifestations in public procurement.

First, transparency is a principle of open contracting through which the procurement body and, in general, the public administrations can render accounts on public contracts and citizens can thereby have authentic and effective knowledge about the activity and functioning of the public administrations, allowing them to participate in public decisions and monitor administrative activity. Transparency also contributes to strengthening public integrity and preventing corruption. From this perspective, transparency is a manifestation of open governance in public procurement, the aim of which is to strengthen the dialogue between the public administrations and citizens to increase knowledge about contractual activity, participation in public decision making in this area and collaboration in the awarding, execution, monitoring and control of public contracts. In this direction, *Open Contracting Partnership* states that “procurement that is not open undermines shared development. (...) Open contracting strengthens accountability and trust among the different stakeholders in the process and ultimately contributes to better contractual fulfilment and improved development results” (Open Contracting Partnership, 2016, 3).

Furthermore, transparency is a public procurement principle that seeks to boost equity among bidders, promote concurrence in public tenders, guarantee that contracts are awarded to the best offer and the expected results obtained and, ultimately, facilitate integrity. To this effect, the community planning recognises transparency among the principles of public procurement. As the Court of Justice of the European Union (CJEU) highlights in its jurisprudence, beyond its intrinsic value transparency is an instrument to guarantee the principle of equal treatment and an essential requirement of the arm’s length principle. In this regard, the benchmark judgement issued in the *Teleaustria* CASE on 7 December 2000, C-324/98, stated that “the awarding entity’s obligation of transparency consists in guaranteeing, for the benefit of all potential bidders, adequate publicising that allows the services market to be open to competition and impartiality in the awarding processes to be controlled” (section 62). Furthermore, and as recalled later in the *Commission/Cas Succhi di Frutta* judgement of 29 April 2004, case C-496/99, the principle of transparency “means that all the terms and mode of the tender procedure are formulated clearly, precisely and unambiguously in the tender notice in or in the terms and conditions, so that all adequately informed and duly diligent bidders can know its exact scope and interpret it in the same way; and, on their part, so that the



awarding body can effectively check whether the offers submitted by the bidders meet the applicable criteria of the corresponding contract” (section 111). To this effect, and as mentioned previously, according to the CJEU, transparency allows partiality in contract awarding to be controlled and conflicts of interest and corruption in public contract procurement to be avoided. In this regard, the judgement of 7 December 2000 in the *Teleaustriacase* recognised that transparency must facilitate the control of impartiality in contract awarding procedures (section 62); and later the judgement of 12 March 2008 in the *Evropaïki Dynamiki* case affirmed that the objective of the principle of transparency “is essentially to guarantee that there is no risk of favouritism or arbitrariness on the part of the awarding body” (section 144).

Transparency is currently channelled through different mechanisms, the difference between them to do with the role played in them by the public administrations and citizens. In this regard, first we can refer to the dissemination of public information (active transparency), then to access to public information (passive transparency) and, last, to the re-use of public information (collaborative transparency).

The universalisation of the use of electronic means has had a significant impact on the different transparency mechanisms (Cerrillo i Martínez, 2012a). First, it has facilitated the creation of channels for access to and the dissemination of information. Second, it has increased the volume of information and data available for the use of online means in the processing of public procurement procedures. Last, it has increased the uses for public information, both for the public administrations and for citizens.

A. The proactive dissemination of public contract information.

The dissemination of public information consists in proactively making information that could be of interest to citizens related to the public administrations available to them.

In particular, the dissemination of information about public contracts provides knowledge about the decisions taken in this regard and the reasons for them, the resources allocated to them, how the contracts are awarded and the control and rendering of accounts. Furthermore, and as mentioned previously, the dissemination of information about public contracts promotes equality and concurrence among bidders in the procurement process.

The proactive dissemination of public information is provided for in general terms in the transparency acts that have been approved by Member States (for example, Law 19/2013, of 9 December, on transparency, access to information and good governance also Decreto Legislativo 14 marzo 2013, n. 33, Riordino della disciplina riguardante il diritto di accesso civico e gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazioni).

And regarding information about public procurement in particular, its dissemination is provided for in Directive 2014/24/UE, which establishes the obligation to publish notices related to public contracts and the possibility of diffusing this information in the buyer profile. Based on the provisions of the EU regulation, the different Member States have regulated these proactive information dissemination channels for public procurement (for example, LPSC, articles 63, 135, 151 o 154).

Beyond this legal provision, the proactive dissemination of public information on the part of the public administrations is in response to the transparency policies currently promoted in most Member States for the purpose of preventing and combating corruption. Notably in this regard, article 10 of the UN Convention against corruption stipulates that each state must adopt the means necessary to boost transparency in their



public administrations, including regarding their organisation, functioning and decision-making processes, where applicable.

B. Access to public contracting information

Access to information is usually defined as a subjective right, sometimes with a fundamental nature (for example, Charter of fundamental rights of the European Union), of all persons be they a bidder or not, thereby not having to have or accredit a specific interest in knowing the information about public contracts. Both EU institutions and most Member States currently have a regulation about access to public information (Dragos, Kovač, & Marseille, 2018); (Worthy, 2020).

Parallel to this general right, bidders in a procedure have the right to access information about contracts, for example, the terms and conditions and complementary documentation, and the information related to the contract awarded. For example, this is provided for in article 53.2 of Directive 2014/24/UE, which establishes that the awarding bodies must provide bidders with access to the information about the terms and conditions and any complementary documentation, and to wait six days before the deadline set for receiving offers.

The state transposition regulations also include this provision. For example, the LPSC provides that the public sector entities will facilitate the free access to information about public contracts (article 28.2 LPSC). To this effect, as observed in resolution 21/2017, of 1 February, of the Commission to Guarantee the Right of Access to Public Information, “the entry into force of the new transparency legislation has strengthened the right of access for bidders to the contract file, in general, and to the documentation submitted by the other bidders, in particular”. In fact, as recalled in the cited resolution, Directive 2014/24/UE expressly obliges the state regulation on citizens’ right to access to be taken into account when the confidentiality of the data provided by bidders and candidates and is regulated.

To this effect, the Spanish regulation stipulates that those interested in submitting a bid can request access to the terms and conditions and complementary documentation (article 138.1 LPSC); information related to the contract awarded or to the file prior to the filing of the special recourse in matters of procurement (article 52.1 LPSC). However, generally in these cases access is only recognised for candidates and bidders that have taken part in a procedure, and those that have been rejected from it.

The right to access must generally be exercised following a procedure through online channels that have been set up for this purpose on transparency portals or in the buyer profile. In some countries, specific guarantee mechanisms have been established as a result of public administration non-compliance (for example, in Spain the possibility of filing a claim before the Transparency and Good Governance Council or, in the case of the Regional Government of Catalonia, the Commission to Guarantee the Right of Access to Public Information).

C. The re-use of public procurement information.

As stated in the manifesto of the European Data Strategy [COM(2020) 66 last], “the use of data can give EU companies and the public sector the means to make better decisions”.

In particular, the re-use of information allows citizens to collaborate with public transparency and, where relevant, the analysis of public information that allows contracting activity to be monitored and controlled (Granickas, 2014, 4). In effect, public collaboration in public transparency leads to collaborative

transparency, which is the transparency channelled by citizens through the re-use of public information. Collaborative transparency is a vital aspect in any anti-corruption strategy (Beke & Blomeyer&Sanz, 2015, 31); (Web Foundation, 2018). As Kaufmann states, this makes members of the public into thousands of auditors (Kaufmann, 2002, 19).

As we will see later, to facilitate the re-use of contractual information the public administrations must disseminate the data in open format and under legal conditions that allow its re-use.

1.3 Data accessibility, Interoperability and reusability

For transparency to be an effective mechanism to guarantee integrity and prevention of corruption, different elements that provide effective knowledge of the contractual activity and allow its monitoring must concur.

The characteristics the data must have to be able to meet these purposes varies. From a general perspective, reference is made to the initiative of the FAIR principles; in others words, the data must be available, accessible, interoperable and reusable (Wilkinson et al., 2016).

To our understanding, in the area of public procurement the data characteristics that must concur so that they can be used to control contractual activity and thereby help combat corruption is that public contracts must be available, interoperable and reusable (DIR).

These characteristics must concur independently even though they are closely related, as we can see below.

Availability	Accessibility	Completeness
		Ease of access
		Understandability
	Quality	
	Openness	
Interoperability		
Reusability		

Table. Data elements

A. Data availability

Availability is the first characteristic that data about public contract procurement must have and it refers to the data being ready to be used. In this regard, Soylu recalls that “transparency and accountability require giving citizens and companies much more data with the possibility of easily connecting relevant data sets (e.g., spending and company data), both within and beyond national borders and languages, allowing extended and deeper analyses” (Soylu et al., 2022).

Availability requires that the data have three characteristics: accessibility, quality and openness.

a) Information accessibility

Information accessibility refers to information being complete and relevant, available through an easily locatable and accessible channel and easily understandable.



- Information completeness

First, public administrations must guarantee that all the information necessary for citizens to know their activity is at their disposal, in general, and at bidders' and contractors' disposal, in particular.

For this, the information must be complete; in other words, it must make reference to all the aspects linked to the life cycle of the contracts, and to all the people taking part in it. The information should only not be facilitated when its knowledge can cause harm to certain property or rights explicitly protected by the regulation in force (personal data, the confidentiality of the information, intellectual property, public security, etc.). In relation to this matter, the public administrations must ponder whether the information must be transparent or not.⁵

The information about public contracts that the public administrations disseminate must be relevant to guarantee the transparency of the contractual information. To this effect, the procurement bodies must disseminate all useful information to guarantee the transparency of public procurement and to facilitate the supervision and control of the procurement activity of the public administrations.

The current legislation aims to guarantee the completeness and relevance of the information about public contracts that is disseminated, identifying the information that must be disseminated. However, beyond this information the procurement bodies can publish other information that may be of interest, particularly to detect conflicts of interest and cases of corruption (for example information about the meetings held by those responsible or the staff at the service of the procurement body and the bidders and contractors; information about gifts or courtesies they may have received, and so on).

To this effect, the transparency laws provide for the dissemination of a large volume of data on public procurement. In the case of Spain, both Law 19/2013, of 9 December, on transparency, access to public information and good governance, and the regional acts that have developed the state regulation, provide for the dissemination of information about the preparation, awarding and execution of public contracts through the transparency portals.

<p>Data that must be dissemination in relation to all contracts:</p> <ul style="list-style-type: none"> - Object of the contract - Duration of the contract - Amount of the tender and awarding - Procedure used for its execution - Instruments used for advertising, where applicable. - Number of bidders taking part in the procedure - Identity of the awarding body - Contract modifications - Decisions regarding rejecting or renouncing contracts. <p>Statistical data about public contracts</p> <ul style="list-style-type: none"> - Percentage in budgetary volume of contracts awarded through each of the procedures provided for in the public sector contracts legislation.

Table. Data to be published in the transparency portal (ES)

⁵ These aspects will be covered in Working Paper 2.



Specifically, in relation to contractual information, Directive 2014/24/UE provides for all the information that must be disseminated. In this regard, the reading of Annex V of Directive 2014/24/UE stipulates the data that must be published in the different notices, which is shown in the following table in summary form.

<p><u>Information about the awarding body</u></p> <ul style="list-style-type: none"> - Name, ID number (where required in the national legislation), address, included NUTS code, telephone and fax number, email and website address of the awarding body and, if different, of the service by which complementary information can be obtained. - Type of awarding body and main activity carried out. - CPV codes. - Website address of the “buyer profile” (URL). <p><u>Information about the contract</u></p> <ul style="list-style-type: none"> - Type and scope of the works, type and quantity or value of the supplies, type and scope of the services. If the contract is divided into lots, this information will be provided for each lot. Where relevant, description of possible variants. - Total estimated order of magnitude of the contact or contracts: when the contracts are divided into lots, this information will be provided for each lot. - Calendar for delivering supplies or works or for providing services and, where possible, the duration of the contract. - Specific conditions to which execution of the contract is subject. - NUTS code of the main location of the works, in the case of works contracts, or NUTS code of the main delivery location or place of execution in the case of supply or services contracts. - Value of the offer(s) selected or the highest and lowest values of the offers being considered for awarding the contract(s). - Where relevant, for each contract award, the value and proportion of contracts intended to be subcontracted to third parties. - Information about whether the contract is related to a project or programme funded by the EU. - Description of the contract before and after its modification; type and scope of the works, type and quantity or value of the supplies, type and scope of the services. - Price increment due to the modification. - Description of the circumstances that necessitated the modification. <p><u>Information about notices</u></p> <ul style="list-style-type: none"> - Date of the notice in terms of the publication of a preliminary informative note sent to the buyer profile. - Date notice sent. - Dates and references of previous publications in the Official European Union Newspaper relevant for the contract(s) being notified.
--

Information about the procurement procedure

- Type of procurement procedure (restricted procedure, including or not a dynamic acquisition system, or procedures with competitive dialogue).
- Conditions for participation
- Brief description of the criteria that will be applied to award the contract.
- Total estimated magnitude of the contract(s); when the contract is awarded in lots, this information is provided for each lot.
- Information about the lots
- Specific information related to restricted procedures, procurement with negotiation, competitive dialogue or association for innovation.
- Deadline for receiving manifestations of interest.
- Deadline for receiving offers or applications to take part.
- Address to which manifestations of interest and offers and applications to take part must be sent.
- Language(s) authorised for submitting candidatures or offers.
- Where relevant, indications as to whether: a) offers and applications to take part will be required or accepted by electronic means; b) electronic requests will be used;
- c) electronic invoicing will be used; d) electronic payment will be accepted.
- Information about whether the contract is related to a project or programme funded by the EU.
- Number and address of the body responsible for filing appeals and, where applicable, for mediation. Instructions on deadlines for appeal proceedings or, where necessary, the number, address, telephone and fax numbers and email of the service via which this information can be obtained.
- Criteria established in article 67, which will be used for awarding the contract(s). Where relevant, indication as to whether an online auction was used (in the open, restricted and procurement with negotiation procedures).
- Date of awarding the contract(s) or the framework agreement(s) following the decision for the awarding or the event.
- Number of offers received with respect to each award and, specifically:
 - the number of offers received from economic operators that are SMEs; number of offers received from another Member State or third country;
 - numbers of offers received electronically.

Information about the bidders selected

- Name, address, included NUTS code, telephone and fax number, email and website address of the bidder(s) selected, specifying:
 - if the successful bidder is an SME;
 - if the contract has been awarded to a group of economic operators (company in participation, consortium or other).

Table. Information about public contract procurement according to Directive 2014/24/UE

Despite all the information that must be published via the Official Journal of the EU, Fazekas has observed that while some of this information is provided for almost all contracts (for example, the awarding body, other information is often lacking (for example, the value of the contract) (Mihály Fazekas, 2017, 6). As this author states, “While TED data is impressive in both size and scope, it typically captures high-value



contracts above the mandatory reporting thresholds which means many smaller or poorer regions and localities have little contracting activity to analyse” (Mihály Fazekas, 2017, 38)

As previously mentioned, while emergency contracting can impact on publicising prior to the contract, it must not intervene with respect to the information related to the awarding and formalisation of the contracts (Oficina Independiente de Regulación y Supervisión de la Contratación, 2020). However, and as likewise mentioned previously, during the COVID-19 pandemic, compliance with these obligations has been very irregular among the different public administrations.

- Ease of access

Second, the public administrations must guarantee that all the information is disseminated via channels that allow citizens to generally recognise it and which, therefore, are easily locatable and allow fast and easy access to the data. Thanks to this, “data on economic activity and transactions, especially adjacent to the public domain, are becoming increasingly well-organized and available” (Villamil, Kertész, & Wachs, 2022).

Regarding the channel, the transparency regulation covers the creation of transparency portals, online advertising platforms via which the public administrations disseminate information to comply with transparency legislation. The transparency portals disseminate all the information in an aggregated and organised way. On occasion, the public administrations disseminate information in graphic format or using language that is accessible to citizens, facilitating their access to it and understanding of it.

In the area of public contract procurement, the main publicity mechanism considered by Directive 2014/24/UE are the notices related to public contracts (generally the preliminary informative notices, procurement notices, contract awarding notices, and notices related to modifying contracts, the awarding of social services and other specific services contracts or project tenders) (articles 48, 49, 50, 72, 75 y 79 Directive 2014/24/UE, respectively). These notices are elaborated by the awarding bodies and are sent by electronic means to the European Union’s Publications Office, which will be responsible for their publication within a period of five days from receiving them (article 51.2 Directive 2014/24/UE). These notices must be published in full in the official language(s) of the EU institutions chosen by the awarding authority, in addition to a summary of the important points of each notice in the other official languages of the EU institutions (article 51.3 Directive 2014/24/UE). The notices must be available for the stipulated period (for example, preliminary informative notices for 12 months or until notice that the contract has been awarded is received (article 51.4 Directive 2014/24/UE).

All this information is accessible through the online version of the EU’s Official Journal Supplement, dedicated to European public contracting, Tenders Electronic Daily (TED). Hundreds of thousands of contracts awarded in the different Member States are published there every year, amounting to more than 2,500 procurement notices every day. The information is organised by country, regions and activity sectors. The information is published in the 24 official EU languages. TED offers some added value services for registered users (for example, adapted search profiles depending on needs; email notifications sent based on search profiles; personalisation of RSS sources on websites; access to automatic translation of all the notices).

Together with this mechanism, Directive 2014/24/UE stipulates that the awarding authorities can publish this information at national level, in theory after their publication at European level. To do so, the awarding bodies must create a buyer profile (article 52). Aside from the notices, the awarding body must guarantee access to the terms and conditions of the contract via open, direct, complete and free electronic means (article 53).

The information must be easily locatable in the different channels. For this, the information must be suitably organised and make use of management systems that facilitate its localisation (for example, via



search engines). Unique identifiers can also be used (of the awarding authority, of the bidders and contractors, and on the contracts), which facilitate the traceability of the information.

Some transparency regulations specifically stipulate that the information must be easily locatable. To this effect, Law 19/2013, of 9 December, on transparency, access to public information and good governance stipulates that the information must be disseminated in a structured way.

However, the regulation on public contract procurement does not dictate the required characteristics, the contractor profile or the public sector contracting platform.

- Understandability

Third, the information must be understandable. Understandability requires that the information facilitated must be written at different levels depending on who the recipients of the information are. At a first level, the information must be disseminated in summary form and with graphics that facilitate its understanding by citizens. At second level, the possibility of accessing the complete information must be given, where the data are presented in a more detailed way.

Understandability is closely related to accessibility of information in the strict sense, the purpose of which is to guarantee that the information can be received and recognised by anybody irrespective of their personal circumstances. In relation to this element, notably Directive (UE) 2016/2102 of the European Parliament and of the Council, of 26 October 2016, on the accessibility of public sector organisations' websites and mobile applications, stipulates that accessibility must be addressed in an integral way in the design, management, maintenance and updating of website and mobile device content; that the public authorities must offer a mechanism by which suggestions and complaints can be filed; they must inform of any possible non-compliance with the requirements of accessibility, and that the public administrations must disseminate a detailed, exhaustive and clear accessibility statement on complying with the provisions of their current website and mobile application regulation.

Some transparency regulations specifically stipulate that the information must be easily locatable. To this effect, Law 19/2013, of 9 December, on transparency, access to public information and good governance stipulates that the information must be disseminated clearly and in an understandable way for all interested parties. Once again, the public contract procurement regulation says nothing to this effect.

b) Quality of the information

Second, information about public contract procurement must be of quality; in other words, it must be easily usable so that users and recipients can meet the objectives foreseen and sought with its consultation.

The quality of the information is closely related to its accessibility. In effect, quality information is complete, precise, current, consistent and accessible (Cichy & Rass, 2019):

- Completeness: the data have the width, depth and scope required for their purpose.
- Precision: the data are correct, reliable and certified.
- Up-to-dateness: the data must be current.
- Consistency: the data is presented in the same formats and are compatible with previous data.
- Accessibility: measure in which the data are available, or are easily and quickly recoverable.



Compliance with these characteristics guarantees the quality of the data and, therefore, the possibility of monitoring and evaluating the contractual activity. Contrarily, “Without high-quality data providing the right information on the right things at the right time; designing, monitoring and evaluating effective policies becomes almost impossible” (United Nations, 2014).

Special mention must be made of updating the information. Updating means that the information disseminated by the public authorities is current at the time of it being facilitated, and its dissemination is maintained during the period that the information remains of interest and of use. The public administrations must identify the periods in which the information must be disseminated, the time the information will remain disseminated and when the information will be updated.

The public administrations must have mechanisms and methodologies to know the level of quality of their data. All the data must be displayed publicly so that the re-users of the data can also know the utility of the data disseminated. These methodologies must allow the level of precision, up-datedness and absence of erroneous data to be evaluated.

From this perspective, the information must be published regularly and be kept updated, periodically withdrawing any obsolete data. To guarantee quality, it is important that the regularity with which the data will be updated is decided.

According to the World Bank, “information about the volume of spending and the awarding of grants and loans to persons and companies must be shared daily and in a way that is accessible, using multiple means to ensure maximum reach” (World Bank Group, 2020a). In the same line, Transparency International has made a similar statement, considering that “it is important, especially in this extraordinary situation, that public information is available in real time and that it is contrasted, hierarchised and evaluated so that citizens and different interest groups can easily and clearly access its content” (Transparencia Internacional España, 2020).

One of the main problems re-users of information have is its low quality, for example it contains errors, it is not current or it is biased. As the (Comisión Nacional de los Mercados y la Competencia, 2019), has observed “there remains another fundamental problem for quantitative analysis in this area: data quality. The information published on the different procurement platforms contains a number of inconsistencies and missing data that are not negligible”.

The absence of quality in the information can impede its analysis or make the results obtained in the analyses carried out not reliable (Soylu et al., 2022). These problems can be magnified when we refer to large volumes of data, or big data (Juddoo, George, Duqueno, & Windridge, 2018).

Regarding public contract procurement, this problem has been highlighted by several authors who have attempted to re-use contractual information to analyse whether there are any irregularities or cases of corruption. Fazekas has observed that despite the data published in the TED being revised by the EU’s Publications Office, “there is a non-negligible amount of missing or nonsensical data. Our analysis shows that data errors tend to be concentrated in selected countries and procuring bodies (see more on data quality and improvements below). The contract-level public procurement database used in this analysis can be downloaded at digiwhist.eu/resources/data” (Mihály Fazekas, 2017, 5). However, the level of transparency does not only affect the information published in TED, but contractual information in general. In line with this author, “Such deficiencies in data quality and scope are by no means limited to TED; many national public procurement systems are weak, especially among the richer Member States (Cingolani et al., 2015)” (Mihály Fazekas, 2017, 38), the situation varying significantly among the different European countries (Mendes & Fazekas, 2018). For example, according to Soyly, in the UK “the number of contract award notices that include a zero value in the Contracts Finder site has exceeded 10% in every year since its launch” (Soylu et al., 2022). “According to a survey, approximately 10 percent of published contracts in Slovakia have at least one piece of key information missing”, while “another study found that a quarter of contracts had a “subject” missing,

12 percent were missing price-related information, and in 4 percent of contracts the name of the counterparty was redacted” (Clare, Sangokoya, Verhulst, & Young, 2016).

In this regard, the first challenge to be addressed to advance in the re-use of contractual data is to guarantee its quality. Kutlina-Dimitrova agrees, pointing out that “the first challenge is related to the lack of reliable, detailed data and internationally agreed methodology for standardising and collecting government procurement contract award data” (Kutlina-Dimitrova, 2018).

This often happens in public contract procurement and especially in emergency situations. For example, OiRESCON has observed that “there are still fields with key information for procurement files which are either not of mandatory compliance, or can be erroneous or incoherent, given that there is no quality control or verification of public information (...). One of the major handicaps in analysing the information on the different platforms was the lack of uniform and homogenised criteria, which directly affects the quality of the data and, among other things, impedes the massive exploitation of public contract procurement data” (Oficina Independiente de Regulación y Supervisión de la Contratación, 2021). Fundación Civio has highlighted the same when re-using public procurement data, pointing out that “we have had to extract, clean and structure and complete data for months and we have found information gaps, errors, incoherent codes, delays and differences in the ways that each body publishes its data. But it was the only way to have as complete a vision as possible: the Transparency Portal, for example, includes only contracts from the General State Administration-and not all of them-and confuses emergency with urgency”.⁶

In light of all the above, not only must the importance of quality data be recognised. but also that the public administrations must have the mechanisms needed to guarantee quality and to prevent the information disseminated on contracts from being incomplete or erroneous (Mendes & Fazekas, 2018)

c) Data openness

Public procurements data openness seeks to guarantee that the public administrations facilitate information to citizens and companies that can be easily used without technical or legal obstacles, without glitches that can hinder or make their re-use difficult, to create new innovative information or services for private or commercial ends, but also for general interest purposes.

Through data openness the public administrations can place huge volumes of data related to public procurement within reach of companies, civil society and the scientific community. With data openness regarding public contract procurement not only can the economy of data be boosted, but advantages for citizens are reported, facilitating their involvement in public affairs, for example, supervising and controlling public procurement.

Data openness has stood out as an anti-corruption strategy with a significant impact on public integrity (Haute Autorité pour la transparence de la vie publique, 2016). In the same line, la National Commission on Markets and Competition (CNMC) affirms that “the construction and maintenance of global, interoperable and specifically designed databases at the most disaggregated level possible can enable the supervision of economic efficiency by specialist bodies” (Comisión Nacional de los Mercados y la Competencia, 2015). In effect, as recognised by the G20 Principles on open data against corruption, open data can contribute to preventing, detecting, investigating and reducing corruption.

From a general perspective according to the *Sunlight Foundation*, open data are data that are complete, primary, opportune, physically and electronically accessible, electronically processable, non-

⁶ Accessed at: <https://datos.civio.es/dataset/todos-los-contratos-de-emergencia-desde-2020-publicados-de-todas-las-administraciones-publicas-espanolas/> (last accessed: February 2022).

discriminatory, use open standards in the dissemination of the information, unlicensed, permanent and without cost.⁷

These characteristics have been decided in the area of open procuring through different initiatives like that promoted by the Alliance for Open Procurement in collaboration with the World Wide Foundation and the World Bank in the open procurement data standard. This is a non-proprietary structured standard referring to the entire procurement cycle, which allows its users to publish shareable, reusable and machine readable data and to design applications for analysing and sharing data.⁸ Its use has been identified as a useful instrument to detect fraud and cases of corruption and, ultimately, to achieve greater concurrence and profitability in public procurement. The standard specifies a methodology for identifying data that must be disseminated, a structure for how it must be disseminated and some common patterns of publication for open data. Furthermore, from the Open Contracting Alliance, a community for ongoing improvement and support in the implementation and exchange of experiences has been facilitated.⁹

In addition to the open contracting data standard, they can also refer to the Guidelines to Open Data in Contracting, produced by the *Sunlight Foundation*,¹⁰ which state the principles that must characterise the dissemination of data related to public procurement.

The procurement process

1. All laws and policies related to public procurement must be accessible to the public.
2. Tenders must be available to the public and must remain available even when the deadline has ended.
3. Data related to direct procurements and other contracts that do not require public tender must also be published.
4. The documents in each contract and the contact information of all the bidders must be made public.
5. Communication between businesspersons and public employees during the tender process must take place in a public forum.
6. Contract awards and their justifications and awarding criteria must be publicly disseminated as soon as they take place.

⁷ Accessible at: <https://sunlightfoundation.com/policy/documents/ten-open-data-principles/> (last accessed: February 2022).

⁸ Accessible at: <http://standard.open-contracting.org/> (last consulted: February 2022).

⁹ One of the aspects on which both the open contracting data standard and the Guidelines on Open Data insist is the need to identify each contract by means of a unique identifier (Open Contracting ID, OCID), which allows the life of the contract to be monitored in its different phases, and links the data that can be found in different databases. The unique identifier is based on an international system that seeks to create a unique system that is coordinated at world level. Its use facilitates the monitoring of contracts. The possibility of using a unique identifier can also be used to identify the bidders and the awarding bodies.

¹⁰ Accessible at: <https://sunlightfoundation.com/procurement/opendataguidelines/> (last consulted: February 2022).

7. The complete text of any contract awarded and any modification made to it must be published proactively in order, facilitating minimum information, such as that related to the awarding body, the amount, the date, etc., in addition to structured data.
8. Information related to the procedures to resolve conflicts must be available during all the phases of the procurement procedure.

The participants in the procurement procedure.
9. The databases that contain current and historic information about the execution of contracts must be publicly available.
10. Information related to the status of contracts must be available to the public at all times and even after the contract is finalised.
11. If a large part of the contract is subcontracted, the available data for the contractor must also be available for the sub-contractor.
12. Contracts must have a unique identifier for ongoing use in the contracting process.
13. The awarding bodies must be identified by a non-propitiatory, publicly available unique identifier. How to free this information.
14. Wherever possible, governments must aim to adopt existing regulations for freeing contracting data to promote its trans-border interoperability.
15. All the aforementioned data must be available as structured data in machine readable formats. They must also be available online and on paper.
16. All the aforementioned data must be available without cost and without requiring registration, either online or in person.
17. All data must be published at the opportune moment and continuously throughout the fiscal year. Delayed dissemination to protect the confidentiality of the information must be an occasional occurrence.

Guidelines on Open Data in Contracting

Some of these elements are already reflected in the legislation.

In the last years, the regulations governing the re-use of public information has also advanced towards data openness. In the last decades, this evolution has been boosted by technological development, the strengthening of relations between citizens and the public administrations, international promotion and, in particular, by Directive (UE) 2019/1024 of the European Parliament and of the Council, of 20 June 2019, relating to open data and the re-use of public sector information.

According to consideration 16 of Directive (UE) 2019/1024, open data are “the data in an open format that anybody can freely use, re-use and share for any purpose”. For this, the data must be facilitated using



open formats; in other words, “in a file format independent of platforms and places at the disposal of the public without restrictions that impede the re-use of the documents” (article 2). To facilitate their re-use, the data must be disseminated using machine readable formats, or in other words in a “structured file format that allows IT applications to easily identify, recognise and extract specific data” (article 2).

To facilitate their re-use, the procurement bodies must elaborate and facilitate documents according to the principle of “open documents by design and by default” (article 3). To this effect, the provisions of the G20 Principals on open data against corruption adopted during the Turkish presidency in 2015 are complied with, which recognise that to be able to combat corruption open data must be open by defect.

In this line, Directive (UE) 2019/1024 stipulates that “public sector bodies and public undertakings shall make their documents available in any pre-existing format or language and, where possible and appropriate, by electronic means, in formats that are open, machine-readable, accessible, findable and re-usable, together with their metadata. Both the format and the metadata shall, where possible, comply with formal open regulations” (article 5). However, this provision does not mean that it recognises the obligation of public bodies “to create or adapt documents or provide extracts in order to comply with that paragraph where this would involve disproportionate effort, going beyond a simple operation”, nor does it require that they “continue the production and storage of a certain type of document with a view to the re-use of such documents by a private or public sector organisation” (article 5.4)

Furthermore, to facilitate access to data that someone wants to re-use, Directive (UE) 2019/1024 stipulates that public sector bodies “shall make dynamic data available for re-use immediately after collection, via suitable APIs and, where relevant, as a bulk download” (article 5).

Furthermore, when the datasets have a high value, in other words when they are “documents the re-use of which is associated with important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs, and of the number of potential beneficiaries of the value-added services and applications based on those datasets” (article 2), they will be made available to re-users in a machine readable format via suitable APIs and, where relevant, as a bulk download”. The high-value datasets will be established by the commission by prior consultation with operators and experts and having carried out an impact evaluation and analysed the complementarity with current regulations, in view of the fact that they can generate important socioeconomic and environmental benefits and innovative services, befitting a large number of users, in specific SMEs; contribute to generating income, and be combined with their datasets from among the planned thematic fields in Annex 1 (geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, mobility) (article 14).

To facilitate the re-use of data, Directive (UE) 2019/1024 also stipulates that the Member States create practical devices that facilitate searching the available documents (lists of documents with metadata, portals connected to decentralised lists, etc.). Likewise, where possible this will facilitate the linguistic search of documents in different languages (article 9.1). It will also simplify access to datasets, for example, by creating a unique access point to all the documents to which the directive applies in accessible formats, easy to localised and reusable by electronic means



Last, it is interesting to highlight that beyond what is provided for in the regulations through which the Member States have transposed Directive (UE) 2019/1024, some regulations on public procurement have explicit stipulated that the information on public contracting is disseminated in open formats.

In this line, we can highlight the LPSC, which stipulates that “all the information contained in the buyer profiles of the contractor will be published in open and reusable formats” (article 63). Likewise, in the case of the Public Sector Contracting Platform, it is stipulated that the information will be published in open and reusable standards (article 347).

In practice, it is observed that the level of data openness of public contacts is low, by inference of the Global Open Data Index.¹¹

There has also been an evolution in the openness of the data published in the TED Before 2015, the notices published in the TED were in PDF or HTML formats. Despite the information also being stored in XML, it was difficult to process the information because, among other reasons, the data were organised hierarchically. From 2015, the EU’s Publications standardised XML files, meaning that files in CSV format can now be easily downloaded (Duguay, Rauter, & Samuels, 2020).

The reality among the contracts that are not published at European level is very different (Duguay et al., 2020). In the case of Spain, the LPSC stipulates that both the information in the buyer profile and the information disseminated in the public sector contracting platform must be disseminated in open format (articles 63.1 and 346.8 LPSC).

However, (Mendes & Fazekas, 2018) observed that the current situation in Europe with respect to the reusability of information is problematic. In particular, these authors observe that in only three countries and in the TED are the data on contracts fully machine readable. However, even in these cases downloading them is complex and costly. In a further 26 countries the data is only partially machine readable because although the data can generally be downloaded, this is not consistently the case and is not always structured. In 11 countries, the data are not even partially machine readable.

To advance in the openness of public contracting data, different initiatives have been promoted, such as #DatiBeneComune, a campaign promoted to advance the openness and accessibility of the data related to the management of the pandemic.

The future European data strategy can have an impact on the openness of public contracting data, with the creation of a federated public contracting data space.

B. Data interoperability

Interoperability is the capacity of information systems to share data and exchange information. Interoperability allows the applications used by the public administrations or by citizens to exchange information and mutually use interchanged information. As indicated in the European Interoperability Framework, interoperability “allows administrative bodies to electronically exchange among themselves and with citizens and companies, information with meaning and in a form that is understandable for all parties” (European Parliament Communication to the Council, the Economic and Social European Committee and the Regions Committee COM (2017) 134 last).

¹¹ Accessible at: <https://index.okfn.org/dataset/procurement/> (last consulted: February 2022).



To be able to analyse data from different public administrations or information systems together the data need to be interoperable; in other words, they must comply with some regulations, criteria and recommendations that allow information to be interchanged and to mutually use the interchanged information.

The lack of interoperability in public contracting data negatively affects its analysis. Unfortunately, as Soyly et al. point out, "increasingly more open data is being published in the public sector; however, these are created and maintained in siloes and are not straightforward to re-use or maintain due to lack of quality, insufficient metadata, missing links to related domains, as well as the technical heterogeneity." (Soyly et al., 2021)

Following the European Interoperability Framework, we can understand that "interoperability is the capacity for organisations to interact with a view to achieving common objectives that are of mutual benefit and have been previously and conjointly agreed, based on sharing information and knowledge among organisations, through the business processes that support them, by means of the exchange of data among their respective IT systems".

If, in the strict sense, interoperability is related to technology, in the wider sense, other factors, such as social, political and organisation factors, must also necessarily be considered (Gottschalk & Solli-Sæther, 2009). In effect, interoperability has a polymeric nature (Cerrillo i Martínez, 2010). Beyond technology, interoperability has different dimensions whose concurrence is necessary to achieve the finalities of digital administration.

In this direction, interoperability is manifested in different planes: the technical dimension, the semantic dimension and the organisational dimension, to which the European Framework adds the legal dimension.

The technical and semantic dimensions make up the technological aspect of interoperability. Technical interoperability refers to the technical aspects linked to the relationship between computer systems and services. This is the set of characteristics and elements of a technological nature that physically permit information systems of the participating entities to be able to interact among themselves (Jiménez, Criado, & Gascó, 2011).

Semantic interoperability seeks for the exact meaning of the information interchanged to be understandable by any other application that was not originally developed for this purpose. It is concerned with the use of data and information, guaranteeing that any application can understand the precise meaning of the information. The final objective is that, by means of the combination and processing of data from different sources, interpretation is unambiguous. For this, tools such as classification systems, thesauruses, metadata and ontologies are used.

Organisational interoperability, for its part, includes the definition of the objectives of the processes and service of the organisations in the provision of electronic services or of cooperation initiatives or integration of black offices. This dimension of interoperability enables coordination and alignment of the administrative procedures to be ensured.

Last, legal interoperability seeks to guarantee that the organisations that operate within different legal, political and strategic frameworks can work together. This dimension has a clear manifestation when the public administrations that interchange information belong to different states or are subject to various legal regulations.

Beyond the technical dimension, interoperability must be understood as organisations' ability to interact with agreed and common objectives, for the purpose of obtaining mutual benefits.



The European Interoperability Framework is currently the document that outlines the agreed directives at European level that must concur to guarantee European public service provision in an interoperable way. These directives are adopted in the Framework in the form of common principles, models and recommendations for the following purposes:

- to guide the European public administrations in their efforts to design and consistently provide other European public administrations, citizens and companies with European public services that are, as far as possible, digital, trans-border and open;
- to guide public administrations in elaborating and updating their national interoperability frameworks or policies, strategies and national orientations to promote interoperability;
- to contribute to the creation of a single digital market for providing European public services, promoting trans-border and intersectoral interoperability.

The European Interoperability Framework aims to be a general framework applicable to all EU public administrations.

The European Interoperability Framework considers the need to adopt an interoperability governance referring to the decisions adopted to guarantee and supervise interoperability on a national and EU scale (agreements, organisational structures, responsibilities, and so on). The governance of interoperability provided for in the European Interoperability framework includes the Interoperability Action Plan and the European Interoperability Architecture (EIRA).

The interoperability of open data is an important element to be able to take the greatest benefit from their re-use. This is recalled in Haute Autorité pour la Transparence de la Vie Publique, which states that “the use of open data in matters of public integrity often entails the re-use of data that are not strictly related to the action of public officials” (Haute Autorité pour la transparence de la vie publique, 2016).

In particular, in relation to public sector contracting information, different authors observe the lack of interoperability. To this effect, (Mendes & Fazekas, 2018) state that “quite problematically, several countries as well as the EU-wide TED portal use a large number of different templates — more than 15 or even 20 — which adds an unnecessary layer of complexity to understanding and reusing the data”. Soylu et al. also point out that among the problems related to public contracting data that must be resolved are “Data heterogeneity including structured data (e.g., statistics and financial records), as well as unstructured data (e.g., text and social media content) sources in various languages with their own vocabulary and formats, such as PDFs, APIs, CSVs, and databases”. To this effect, they propose “Transforming this large and heterogeneous set of data sources into an interconnected knowledge organisation structure using standardised vocabularies and sustainable knowledge integration and sharing approaches, which could be analysed in depth to detect patterns and anomalies” (Soylu et al., 2022)

Different initiatives have been promoted to resolve this issue and promote interoperability. First, we can again refer to the open contracting data standards promoted by Open Contracting Partnership. Other initiatives mentioned by Soylu et al are OpenPEPPOL,¹² and CEN BII, TED eSenders¹³ (Soylu et al., 2022).

¹² Accessible at: <https://peppol.eu/aboutopenpeppol> (last consulted: February 2022).

¹³ Accessible at: <https://simap.ted.europa.eu/web/simap/sending-electronic-notice> (last consulted: February 2022).

C. The reusability of information

Collaborative transparency is closely related to the possibility of citizens using public administration information. The re-use of information is one of the mechanisms through which the principle of participation in open government is exteriorised (Cerrillo i Martínez, 2017).

Public contracting information can be re-used to detect irregularities, conflicts of interest or cases of corruption. (TACOD, 2015); (G20 & OECD, 2017, 11). To this effect, the re-use of public information allows citizens to be involved in the analysis and monitoring of public activity, thereby increasing citizens' trust in procuring bodies.

Re-use requires that the information is disseminated in technical formats and under legal conditions that facilitate this. In this line, for example, the European Commission recalls the need to “to make better and more accessible data on matters of procurement available because this opens up a whole range of opportunities to improve evaluation of the performance of policies related to procurement, and to optimise interaction between the public procurement systems to shape future strategic decisions” (Comisión Europea, 2017).

For this, as we saw in the previous paragraph, the information must be disseminated in open formats. Furthermore, to facilitate re-use some legal conditions need to concur that facilitate this and have been defined, at European level, in Directive (UE) 2019/1024 of the European Parliament and the Council, of 20 June 2019, regarding open data and the re-use of public sector information, with the aim of fomenting the use of open data and stimulating product and services innovation.

This regulation has made significant modifications to its predecessor, Directive 2003/98/CE of the European Parliament and of the Council, of 17 November 2003, regarding the re-use of public sector information, widening the scope of application of the European regulation (for example, including the information generated during the provision of general interest services by public companies and the data generated in research financed by public funds), encouraging Member States to promote the creation of databases that are in principle open in their design and by defect ad facilitate the re-use of dynamic data of high economic value.

According to Directive (UE) 2019/1024, re-use consists in the use by natural or legal persons of documents that intervene in the power of public sector bodies and public companies, with commercial or non-commercial aims that are different from the initial proposal assigned to these documents in the public service mission or in the provision of general interest services for which they were produced (article 2).

Directive (UE) 2019/1024 does not set out an obligation to facilitate re-use, although it stipulates that Member States should ensure that documents can be re-used for commercial and non-commercial ends (article 3).

In general terms, re-use will not be subject to conditions. However, when the public administrations set a condition it must be objective, objective, proportionate, non-discriminatory and justified on grounds of a public interest objective, and shall not unnecessarily restrict possibilities for re-use and shall not be used to restrict competition (article 8).

re-use will be open to everybody. To this effect, exclusive agreements are generally not allowed (article 12). However the public administrations can attribute an exclusive right to re-use when it is necessary



for the provision of a public interest service. In this case, periodically, and obligatorily every three years, the motive that justifies the granting of this exclusive right must be reconsidered.

The public administrations can subject re-use to obtaining a license. In this line, Directive (UE) 2019/1024 stipulates that the Member States use standard licenses that are available in digital format and can be processed electronically (article 8). Directive (UE) 2019/1024 establishes the principles that must govern the procedure to authorise re-use. Among other aspects, it stipulates that where possible electronic means will be used to submit applications and for delivering the documents in a reasonable time, and in any event within 20 working days of receipt (article 4.2). It also establishes that where re-use is denied, in addition to communicating to the applicant the reasons for this denial, they will also be informed about the possibility of filing an appeal to an impartial review body with the appropriate expertise (article 4.4).

In principle, the re-use of documents should bear no cost. However, Directive (UE) 2019/1024 provides for the possibility of the public administrations recovering the marginal costs incurred for the reproduction, provision and dissemination of documents as well as for anonymisation of personal data and measures taken to protect commercially confidential information (article 6). In this case, the income obtained cannot be more than the cost of its collection, production, reproduction, dissemination and data storage, increasing the investment by a reasonable profit margin. Prior to the re-use of data, the public administration must set their tariffs and publicise them by electronic means wherever possible and opportune.

2. Data on contracts, public officers and companies

The available data on contracting is concentrated especially in the web portals and databases of public contracts that have been referred to in the previous paragraph (such as the Tender Electronic Daily, TED). There are also other European and international initiatives, both public and private, with information on public procurement.

Along with them, it is important to keep in mind that there are other sources of information that disseminate data whose analysis may be of interest and utility in the prevention and fight against corruption.

In this section we will focus our attention especially on data related to the people involved in public tenders as well as in the performance of contracts. In this sense, we prospectively address different possible sources of data on public officers and companies. This is data that does not appear in the contracting portals or databases but that offers valuable information on a possible conflict of interests between the people who participate in the contracting process and the contractors.

These sources of information are very useful and valuable to prevent corruption in public procurement, particularly, to detect conflicts of interests. Even so, most of these sources present problems that up until now are difficult to solve. Especially, difficulties in accessing information -because this information is not structured or access is not allowed- or obstacles to its free access. Consequently, these data do not generally meet basic interoperability and reusability criteria. We will hardly be able to use such data within the framework of this project, although we must emphasize the need for this information to be accessible in the future.

2.1 Availability of the data



The EU institutions have been warning for years about the lack of data and about the need to achieve solid and useful databases (Comisión Europea, 2017). This need has also been revealed in the studies on the subject that can be found both on national (Valcárcel Fernández, 2019); (Thibous, 2019); (Mantini, 2015); (Miranzo Díaz, 2020) and European levels (Mihály Fazekas & Kocsis, 2020); (Arosa Otero, Arvelo Flores, Cano Rodríguez, Colomer Pedrosa, & García Rodríguez, 2021); (Kayte & Schneider-Kamp, 2019); (Prier, Prysmakova, & McCue, 2018), whether they are legal or technological environments. On a European level, the Tender Electronic Daily (TED) platform collects data in two different ways: 1) manual publication directly on the web through online forms and 2) through automatic exchange with the contractor's profiles from the different member states, through the eSender mechanism. Although TED is a great source of public procurement data, increasingly nourished due to the obligations of its use, there are major problems of data quality, as well as a lack of harmonisation in the data published between the different member states of the EU (Arosa Otero et al., 2021). This means there are some methodological problems that can arise in the use of self-reported information – entered into the system not automatically but manually by each employee or contracting authority – such as that published on TED, as the main source of (uncontrolled) procurement data (Czibik, Tóth, & Fazekas, 2015). These types of problems, moreover, are repeated in a large part of the national platforms (Arosa Otero et al., 2021). There are currently studies under implementation aimed at optimizing the quality of TED data (Halsbenning, Scholta, & Niemann, 2020), and the E-forms initiative is working to provide Member States with open standards for the publication of public procurement data, that is, a common framework of standards and terminology, which can significantly improve the quality and analysis of public procurement data.¹⁴

On the national level, the Member States have invested various efforts in the creation of information systems on public procurement in recent years, which gives rise to the availability of information that is generally of quality, but varied and fragmented. In addition, the countries show different degrees of centralisation of information on public procurement, and on the subnational level there are also different constitutional, territorial and socio-political organisational structures that often result in even more fragmented information. The DIGIWHIST study categorised information into 4 types of dispersion, showing that in some cases there were many difficulties in finding the data due to the duplication of platforms on different administrative levels. Specifically, this is the case, at least, for Belgium, France, Germany, Iceland, Italy, Malta, and Spain (Mendes & Fazekas, 2017). For the specific case of emergency contracts, on the national level some supervisory bodies, such as the case of Spain, have shown how the advertising obligations associated with emergency contracts were often carried out only on sub-central platforms, which makes the data collection process difficult (Oficina Independiente de Regulación y Supervisión de la Contratación, 2020).

With regard to the availability of data, however, various initiatives that have created data repositories should also be highlighted, which can help to overcome some of the problems posed by official platforms or to complement their information. On the international level, the World Bank's **Public Accountability Mechanisms Initiative** (PAM) stands out, whose objective is to provide timely information and assistance to the World Bank's operational teams working on issues of transparency and accountability. It does not only focus on elements of public contracting, but on auditability of the public sector in general. Currently the database is disabled but it has been integrated, as far as Europe is concerned, into EuroPAM, a tool resulting

¹⁴ Accessible at: https://ec.europa.eu/growth/single-market/public-procurement/digital/eforms_en (last consulted: February 2022).



from the DIGIWHIST project, and which on its own page is defined as a continuation of the World Bank project.¹⁵

Among the previous studies and initiatives in the matter, the aforementioned European project DIGIWHIST stands out¹⁶, which made important advances in the identification of data sources, their limits and processing capacities. This project also published its own databases –although some of them are no longer updated¹⁷–. In the preparation of these databases, this project revealed some of the main shortcomings of contractual data in the EU Member States, such as the instability of the regulatory framework or its lack of definition, the variety of formats in which information is found, the absence of protocols for the use of the platforms and the introduction of information by the managers (for example, with multiple awardees with similar names), the existence of "free" fields that give rise to multiple formats of information (quantities that use semicolons indistinctly, the absence of forms or templates for publishing the information in the tender notices, contractual modifications or resolutions (even within the same country) (Mendes & Fazekas, 2017), information in PDF, HTML or XML with problems of conversion or collection, the existence of duplicate information, etc. (Czibik et al., 2015); (Tirado, Serban, Guo, & Yoneki, 2016).

At this point, it is worth drawing attention to the existence of tools developed by international entities such as the **Open Contracting Partnership**, which allow cleaning and converting unstructured information into structured and usable data. This is the case of the Flatten Tool for OCDS, which allows converting the unofficial CSV serialisation to a JSON format and vice versa.¹⁸

For its part, the aforementioned **EuroPAM** initiative, also linked to DIGIWHIST, assigns a risk score to different items regarding contracting, conflicts of interest and integrity. Although the evaluation and the indicators used serve to detect deficiencies on the national level, and not on the contractual level, there are important elements that can be useful for the project. Specifically, although carried out in 2015, the DIGIWHIST project produced a *Database of legal and regulatory standards*¹⁹, which can be useful as a starting point for other analyses. For example, in its last update of 2020, the available database²⁰ already contains systematised data on the national thresholds on which the ordinary advertising rules begin to apply, focusing on traditional contracting authorities (public sector, central government) and the three types of contracts: works, services and supplies / goods. In this project they must consider what subjective limits apply and, in the event that they are extended with respect to the DIGIWHIST project, keep it in mind to expand and refine the information in case of using it.

In this project, the information used on public procurement was extracted using scrapers from 20 national web portals and from the European platform TED. For the HTML data, a set of web crawlers was used that extract raw data in HTML format and processed the data to obtain structured information on

¹⁵ Accessible at: <http://europam.eu/?module=about> (last consulted: February 2022).

¹⁶ The Digital Whistleblower: Fiscal Transparency, Risk Assessment and the Impact of Good Governance Policies Assessed (EU Grant Agreement number: 645852)

¹⁷ Accessible at: <http://digiwhist.eu/resources/data/#public-procurement> (last consulted: February 2022).

¹⁸ Accessible at: <https://flatten-tool.readthedocs.io/en/latest/usage-ocds/> (last consulted: February 2022).

¹⁹ Accessible at: <https://digiwhist.eu/publications/database-of-legal-and-regulatory-norms/> (last consulted: February 2022).

²⁰ Accessible at: <http://europam.eu/?module=data-downloads> (last consulted: February 2022).



contracting authorities²¹. It might be considered whether, for those countries in which a multiplicity and/or duplication of platforms have been detected on the sub-central level, it would be possible or potentially interesting to try to extract information from these official regional sources as well.

But in addition, the DIGIWHIST project gave rise to other databases or data repositories directly or indirectly related to the project, and other available databases generated or processed by other private or civil society initiatives should be explored. This is the case, for example, of the Government Transparency Institute (GTI), a think tank that investigates good governance, and which has processed and structured data from the TED platform prior to 2016²². And specifically, perhaps the most interesting database of those mentioned so far in the work is the one that was created, and that continues to be updated, on the **Open Tender** platform²³, from which the contracting data of all EU countries can be downloaded.

Likewise, on the national and sub-national levels, some extremely interesting private initiatives for data collection and processing can be identified, which can offer additional data sources for the development of the project. This is the case of the **Abgeschlossene Projekte**, managed by the Open Knowledge Foundation of the Netherlands, and which allows downloading in open and reusable formats of all the data available from the Dutch platforms, the records, and other information collected in the project.²⁴ Similarly, the Hungarian **Red Flags** project, developed by K-Monitor, PetaByte and Transparency International Hungary, with the support of the European Commission, aims to improve the transparency of public procurement in Hungary and support the fight against corrupt hiring. On its website, it provides an interactive tool that allows the monitoring of contracting processes and their implementation by citizens, journalists or even public officials and detects risks of fraud at different stages of the contracting process. In this case, the Red Flags API tool makes the data collected by the RedFlags project for Hungary available in a structured and actionable format automatically outside of the RedFlags website²⁵.

In Italy, the Autorità Nazionale Anticorruzione (ANAC) has developed an open database called **Banca Dati Nazionale dei Contratti Pubblici**, to directly evaluate data on public procurement, based on its main source of public procurement data that contains more than 52 million records in the last 15 years, and that in addition to offering an interactive search engine, allows downloading all data groups in reusable formats²⁶. In Spain we must highlight the **Contratosdecantabria** initiative²⁷, a civil initiative that has generated its own database, which allows browsing the details of more than 20,000 contracts that exceed 1,300 million euros in the region of Cantabria (Spain), and downloading all the information in a structured way for its processing. It also has processed and reusable information on budgets or companies in the region. The utility of these types of private databases should be explored for inclusion in the analysis.

2.2 Peculiarities of data relating to emergency contracts

²¹ For more information on the particularities found in each country in terms of the availability of information, you can consult: (Cingolani, Fazekas, Martinez Barranco Kukutschka and Tóth, 2015).

²² <http://www.govtransparency.eu/category/reports/> (last consulted: February 2022).

²³ Accessible at: <https://opentender.eu/start> (last consulted: February 2022).

²⁴ Accessible at: <https://okfn.de/projekte/#open-data> (last consulted: February 2022).

²⁵ The data can be accessed here: <http://api.redflags.eu> (last consulted: February 2022).

²⁶ Accessible at: <https://dati.anticorruzione.it/superset/dashboard/appalti/> (last consulted: February 2022).

²⁷ Accessible at: <https://contratosdecantabria.es/wtf/> (last consulted: February 2022).



In CO.R.E., given its focus on contracting in times of a pandemic and other extraordinary periods, a systematisation of the national requirements for access to emergency procedures must be carried out. Although the regulations regarding extraordinary emergency procedures are homogeneous above the European thresholds, the works related to the quality of information in non-competitive procedures suggest that there is little availability and standardisation in the publication of direct awards or non-competitive procedures of contracts, especially in awards below the thresholds (Prier et al., 2018). Additionally, below these European thresholds there are a variety of national enabling budgets to resort to these procedures that can add elements of differentiation, and there is currently no systematised information on this aspect – on the true degree of differentiation between national emergency regulations and others–, nor about the obligation to publish this type of contract, the emergency processing agreement, if the publicity criteria are subsequently applied in the case of modifications, etc. An analysis in this direction would be appropriate for an optimal conceptualisation of the phenomenon and the problem.

The use of this type of awards, moreover, even above the thresholds, based on the legal authorisation of article 33 of Directive 2014/24/EU to execute direct awards in "extremely urgent" cases as we have already seen, is a true deactivation clause of the contracting regulations (Sanchez-Graells, 2020) that, however, does not offer any type of community development, and that may therefore have been interpreted in a heterogeneous way in the different Member States. The first problem in this sense is, therefore, that there is no clear concept of what an extreme situation is.

2.3 Data sources on non-contractual information

To identify the possible risk, it is necessary to connect the data of the contracting party (individual or legal entity) with the people (elected, members of the contracting bodies) who participate in the award process. This does not necessarily imply the existence of corruption but rather the risk of a conflict of interest. The first task is to identify the sources of information, the limitations to its access and the possibility that this information can be reused.

The Anti-Fraud Office (OAC) of Catalonia published the report *Risks for integrity in public contracting* (January 2021) (Oficina Antifrau de Catalunya, 2021) in which different risk factors in contracting are identified. Anti-Fraud performs an X-ray of the risks and the identification of potentially conflictive situations. On the other hand, it highlights that the lack of aggregate data is a difficulty in controlling situations that may entail a risk of corruption.²⁸ The accessible -and free- sources are the official bulletins, and lists and files published in a fragmented way, either in transparency portals, open data²⁹ portals or public websites. The registers –as will be seen in the case of the commercial register- present difficulties of access, and above all of cost and reusability.

²⁸ Thus, it is stated that an obstacle to a good study of public contracting in Catalonia is the lack of aggregate publication and in open data format of information as basic as the complete lists of contracting bodies in Catalonia (autonomous, local and university administrations), the level of subjection to the LCSP with which they are contracting, the budgeted expense for purchase and investment that they manage, the dimensions of the respective personnel templates, etc. This makes it difficult to study the degree of impact of the institutional problems detected in this report. Therefore, the effective resolution of the problems that generate, enhance and perpetuate the risks inherent in contracting depends on the willingness of public entities to self-assess their risks and start designing plans with the most suitable preventive measures in each case. *Íbid.*, p. 29

²⁹ The datos.gob.es portal is an example of this, the data is very fragmented, with different degrees of updating, formats, etc.

The following approximation is made not only from the point of view of the Public Administration - and therefore of the information strictly on public contracting- but also in the possible connections between elected officials and public employees and companies and individuals. In the first case, the existence of reasons for abstention or incompatibilities may affect their impartiality or may give cover to corrupt practices. Connecting public officers with possible contractors could be useful in preventing inappropriate, inconsistent, or illegal conduct and, particularly, conflicts of interests. The following graph illustrates the need to consider both sides and therefore sources of information and data from different sources.

This would allow us to observe the conflict of interests from both points of view. We have a clear example in the Spanish case, both in the regulation of public procurement and in the incompatibilities of elected and public officials. On the one hand, the prohibition of contracting (article 71 LCSP) by Public Administration with certain natural or legal persons, on the other hand, the incompatibilities that both elected (LOREG) and civil servants have to contract with the Administration.

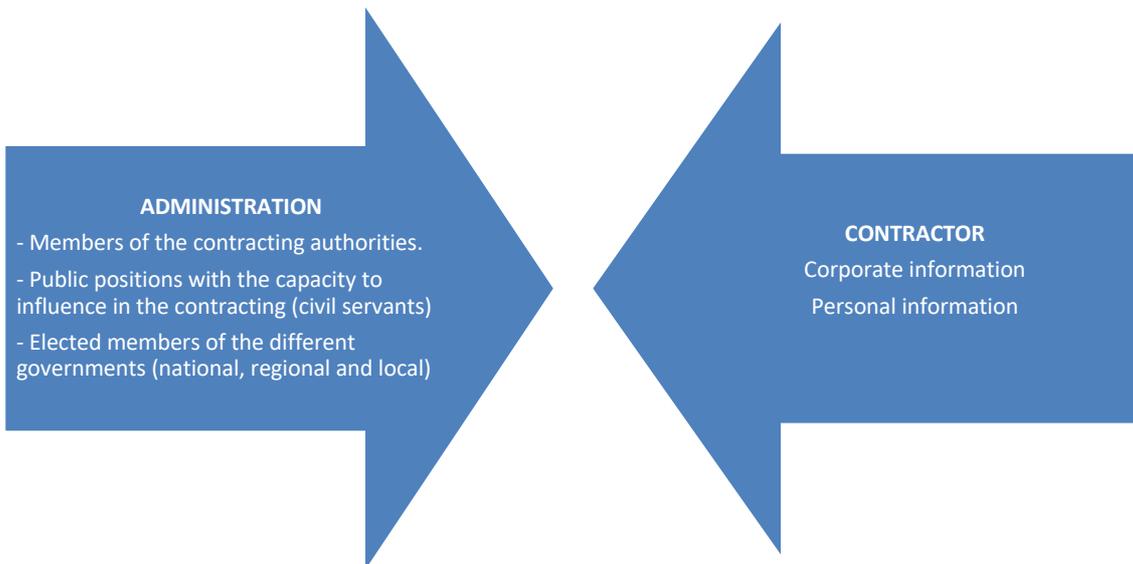


Figure 1: Data sources on non-contractual information

For the detection of possible conflicts of interest sources that include reliable, accessible, reusable and interoperable data are necessary. Beyond the possible risks identified from the data on public procurement, it would be a question of connecting the information on decision-makers in the Public Administration –elected positions and officials- with the information related to contractors. In this sense, it has been recommended that governments link public procurement data and related datasets by using common contract and organisation identifications in different data systems, such as public procurement, payments, company registration or court rulings (Mendes & Fazekas, 2018).



Different categorisations of sources are possible³⁰ In our case we differentiate between sources with information on public positions and employees and those related to companies. In the first case, we should have data on the positions and officials, their responsibilities in the hiring process and possible interests/incompatibilities. In the second, the connections of the contractors with said employees, through family relationships, service relationships or common interests. It is not about directly identifying corrupt practices, but about the existence of risks that compromise impartiality or objectivity and/or facilitate favourable treatment towards a contractor (concession, access and use of confidential information, etc.). The data should be public and reusable or, failing that, be able to be recovered and processed in an automated way (for example, through scrapping systems such as those mentioned).

A. Data sources on public officials

The DIGIWHIST project identifies the main sources of information: contracting authorities, public officials and budget information. Regarding the former, the project refers to lists of contracting authorities, information that in the case of Spain can be obtained through public procurement portals. It is more difficult to obtain information on elected officials and public employees who participate in public contracting processes.

For example, in Spain we do not find any record on contracting authorities, nor structured and/or reusable information in this regard, although it is possible to obtain such information on public procurement portals, although often indirectly. There is no registry of members of the contracting administrative units, although it is possible to verify their names in the case of the permanent contracting tables (on the corresponding platform or Official Gazette) or individually in the administrative clause specifications or other procurement administrative documents. In any case, they are not mandatory bodies according to the LCSP – they are in open, restricted and negotiated procedures with publicity-

Both elected and public employees can be part of the contracting bodies:

a) Regarding **elected members (especially local councillors)**, the first step would be to have a directory, database or registry that allows identifying the person, position and responsibility in the contracting process.

- Elected members register. Basic information -and perhaps a starting point for the rest- is that relating to the elected members (national and regional representatives, mayors and local councillors). In Spain, this information is available on the Portal of Local Entities³¹. It is possible to download –province by province- the lists of elected Mayors (municipality, name, date of election, political party), but not of the councillors. It is an updated list and in Excel format. However, there is no general registry or database of elected officials. The electoral lists are public in each of the elections -national, regional and local-, in the corresponding official gazette. However, it is debatable that being part of a list supposes a risk *per se*, since most of the members are not in starting positions

³⁰ Vasconcelos and Cavique (2022) divide them into four domains: Corruption Domain (C) aggregate data corresponding to illegal acts committed by civil servants or militaries or companies that they are owners; Employment domain (E) provide servant's registrations from Human Resources Management System like income and number of coordination roles; Political Domain (P) covers data related to political activities; and Business Domain (B) presents company features that civil servants and militaries are owners. In their case, they use different records on people sanctioned, databases of public employees, electoral data, etc..

³¹ <https://concejales.redsara.es/consulta/> (last consulted: February 2022).



and simply support a political option, without expectations of holding public office. Unfortunately this situation is also similar in other EU countries.³²

- Contracting bans. It supposes the impossibility of contracting directly or indirectly with Public Administration in which services are provided. On a general level, these situations are regulated in Sapin in article 71 of the LCSP and more specifically in other regulations.

- Declarations of interests. The information on declarations of interests and incompatibilities allow the identification of possible conflicts of interest between the public officials and the contractor. Elected members may maintain activities in the private sector that conflict with the defence of the public interest as elected members. In Spain, declarations of interests and asset declarations must be published³³. The usual publication format (scanned document) does not facilitate its reuse. The situation is similar in the rest of the EU countries. There is an obligation to make such declarations (assets declarations) and in general these are public, although there are no centralised databases. An exception is the Lithuanian Chief Official Ethics Commission (COEC)³⁴, which manages a database of declarations of interest. Kotlyar and Pop propose a declaration automation system that can help detect illicit enrichment, conflicts of interest, incompatibilities or other illicit conduct, and even the construction of risk indicators (red flags) that show the existence of a risk, not only in contracting but in any area of public activity. The authors themselves admit limitations –alerts need verification- and the difficulty of their construction. However, the potentialities regarding contracting are evident: “Risk indicators aimed at detecting potential conflicts of interest would be harder to develop and use, although not impossible (for example, by comparing data in the declaration about business interests of the declarant and family members with any public procurement database)” (Kotlyar & Pop, 2021).

- Previous activities in the private sector. It would make it possible to detect whether previous activity in a company or sector can lead to favourable treatment in a contractor with whom the elected can have a service relationship currently or may have a future expectation. This would mean connecting the information of the elected positions with the companies’ registry, not only directly by position-company but also indirectly (connections between companies of the same group). Also detect a history of corruption that may affect the elected positions.

- In this sense, Wensink & de Vet identify some practices that can help prevent and detect predictive “red flags”, such as “proper screening of contractors and beneficiaries, especially their ultimate beneficiary owners;” and “pre-employment screening and periodical in-employment screening of all those involved in public procurement: public officials as well as temporary staff and external parties hired to facilitate public procurement” (Wensink & de Vet, 2013). The difficulties lie in the non-existence of databases - much less open - that provide such information. The same authors confirm that in most EU countries there is no national database on corruption cases (in the first case), while the prior screening of the people involved in the contracting processes is generally prior to access to Public Administration or based on declarations that are not contained in databases either, even though they are public (Wensink & de Vet, 2013).

³² It is interesting to point out particular initiatives that try to gather this information. For example, the Open Council Data UK portal (<http://opencouncildata.co.uk/>) collects information on local councilors in the United Kingdom, and offers the information in CSV format. However, the author himself warns that this information has been built through other databases -not accessible and not updated- and webscraping, which means that the information is not totally reliable.

³³ In the case of local entities, the annual declarations of assets and activities must be public (article 8.1 h) of Law 19/2013, of 9 December on transparency, access to public information and good governance)

³⁴ <https://vtek.lt/en/home/> (last consulted: February 2022).



- Institutional agendas. The agendas can be available on transparency portals and allow the elected official and the people with whom they meet to be identified.
- Gifts or benefits obtained from private subjects. The mere fact that these are public and limited discourages them from being an effective means of influencing subsequent contracting decisions. This information is also found on transparency portals.

b) Another group of people is that of the members of contracting administrative units who are not elected members, but rather **civil servants**. In the same way as in the case of the elected, it is necessary to have personal data of the civil servants who intervene in the contracting processes.

In Spain it is the responsibility of each Public Administration to maintain a register of its employees³⁵. On the part of the Central Administration, only aggregated statistical information is offered³⁶. The most useful information comes from Public Administrations themselves, which have to create it. However, the information contained in a personnel record has a function of custody and accreditation of the administrative situations of the employee, and does not "connect" with other registries or databases.

- Background (sanction records): The sanctions imposed on civil servants are found in the personnel register of each Administration, whose information is not public.
- Contracting bans. In the same way that has been indicated in the case of elected members, it implies the impossibility of contracting directly or indirectly with Public Administration in which services are provided. In the case of civil servants in Spain, this limitation is contained in Law 53/1984, of 26 December on Incompatibilities of Personnel in the Service of Public Administrations –and regional regulations-.
- Previous activities in the private sector. Although the usefulness of this information is very evident, there is no record of it. It would be possible to obtain this information -partially- through the company registry, although only in the event that the activity is as a director of a company. It would not be possible to obtain information in the event that the previous relationship refers to a service provision relationship or an employment relationship in a contractor company.
- Authorisations of second activities (incompatibilities). These authorisations are public in Spain, in accordance with article 8.1 g) of Law 19/2013, of 9 December on transparency, access to public information and good governance (LTAIPBG): *Resolutions of authorisation or acknowledgement of compatibility that affect public employees as well as those that authorise the exercise of private activity upon termination of senior positions in the General State Administration or assimilated according to regional or local regulations*. It therefore affects both positions and officials. The fulfilment of this obligation is variable and the information is generally not reusable -or not easily so-. However, this information would be useful for the preventive detection of possible conflicts of interest or incompatibilities.

³⁵ The article 71 of Royal Legislative Decree 5/2015, of 30 October (TRLEBEP) regulates the Central Registry of Personnel

³⁶ Secretary of State for Public Administration: <https://www.mptfp.gob.es/portal/funcionpublica/funcion-publica/rcp.html> (last consulted: February 2022).



- Equity increases (tax data, property registration). This is information that can hardly be used in the project to create an alert, despite the fact that it could indicate a rapid increase in the assets of officials.

c) In both cases, family or friendship relationships are an obvious risk of conflict of interest. They are also difficult to define. In the case of financial information, it has been affirmed that it is necessary not only that which affects the position or official (public official) but also that of their immediate family members³⁷, in order to compare the information available with external sources, such as records³⁸.

The difficulty of building that circle of family, friendship or relationship that could lead to a situation of risk is manifested in Spain in the LCSP itself. Indeed, in terms of contracting, article 71.1 LCSP includes various prohibitions on contracting that extend to *spouses, people related to an analogous affective cohabitation relationship, ascendants and descendants, as well as second-degree relatives by consanguinity or affinity of the persons referred to in the preceding paragraphs, when there is a conflict of interest with the head of the contracting body or the heads of the bodies to which the power to contract has been delegated or those who substitute the former.* (article 71.1 g) LCSP). Said circle of prohibition is not strict, since even being in this situation does not necessarily lead to a prohibition of contracting, although it does make it possible to establish an alert. However, materialising the existence of said situation is complex, since these situations are not reflected in any complete and reliable database or source. It is the contracting authorities that must assess the existence of the prohibition (article 72.1 LCSP), but it is difficult for them to do so if they are not previously aware of said possible incompatibility. This situation is not exclusive to Spain, but in any contracting regulation this is an aspect that is difficult to define,

B. Contractor Data Sources

As we have advanced, there are data sources available through which an alert mechanism could be established, and where appropriate, a system of indicators. The problem is not so much in the existence or completeness of the data, but in the difficulties of access to it and especially of interconnection with data from other sources -especially with the information of the people -elected positions or officials- who participate in the contracting process. Also information on who contracts with the administration, to prevent conflicts of interest between the people who participate in the contracting process and the contractors³⁹. For

³⁷ *Many systems require not only the disclosure of a public official's financial interests but also information about the identity of a public official's immediate family and other dependent persons. When that is the case, information gathered generally includes each of those individual's full name and ID details, date of birth and residence details. Disclosure of the identities of relatives is usually limited to the persons living in the same household as the official; although, if conflict-of-interest risks encompass a broader definition of close relatives and associates then the identity of those individuals may also be required* (OECD, 2020), p. 49. The possibilities of information are diverse: *assets, financial assets and investments, securities and stocks, trusts, beneficial ownership or control of companies, income, gifts, sponsored travel, intangible assets, liabilities, expenses and transactions, memberships and positions, outside activities, pre-tenure employment and activities, post-employment work and activities, government contracts* *ibid.*, pp 59-52.

³⁸ It may require a comparison of the data in the financial disclosures with the external data sources (e.g., public registers of companies, real estate and vehicle ownership, procurement awards, licenses and permits sought and issued, etc.) as well as knowledge of official records and processes. The primary focus of the verification/ review, however, should be the detection of potential COI [conflict of interest] and then subsequent management to prevent their escalation to real conflicts of interest *ibid.*, p.53.

³⁹ (...) *it will not only be necessary, but essential, that on the one hand the administration has a database as complete as possible of those companies and individuals with whom it has an economic relationship in each and every*



this, corporate information and personal information about the contractors with Public Administration (or people who have a direct relationship with contractors of the Administration) are necessary. The main sources can come from registers of companies and individuals. Specifically, the Mercantile Registry, and to a lesser extent the Property Registry.

a) Companies registers in Europe

Many of the European company registers are public and offer different information, although a minority provide access to their databases, and even fewer offer such information free of charge. From different organisations –such as Access Info, which has tried to obtain this information through a consultation of 32 European countries⁴⁰- it has been claimed that company registers should be fully accessible with open data formats.

Different civil society organisations -such as the Tenders Guru consortium⁴¹- have repeatedly demanded the opening of the company registry data, in line with the Directive 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public sector information,⁴² which considers commercial registries and registry identifiers ("Companies owned by companies»), a "high value dataset" (Cons. 66), and high-value datasets should be made available for reuse with very few legal restrictions and at no cost.⁴³ In Spain CIVIO has demanded in this sense that the BORME information be open, free [and downloadable] and with a search engine that makes it possible to find, for each company and business, information such as its company name, address, sector of activity, previous names of the company, past owners and all new acts: appointments and dismissals of directors and administrators, status, date of delivery of accounts and document filing history, as well as associated documents⁴⁴.

In the case of the United Kingdom, the Companies House⁴⁵ offers the possibility of finding out information about companies and their accounts. Company information (among other things, registered address and date of incorporation), current and resigned managers, names of previous companies, or

one of its forms (invoicing, bidding,...), including not only the data collected by the administration itself in its files, but also other data existing in other databases (LibreBORME, Mercantile Registry,...). And on the other hand, of those algorithms that allow the detection of certain elements related to malpractice or anomalies (Calabuig Rodríguez, Falciani, Ferrer Sapena, García Raffi, Raso, Sánchez del Toro and Sánchez Pérez, 2018)

⁴⁰ <https://www.access-info.org/es/2021-06-03/open-company-registers-eu-at-risk/>; <https://www.access-info.org/es/registros-mercantiles-2/> (last consulted: February 2022).

⁴¹ <https://tenders.guru/> (last consulted: February 2022).

⁴² Also Directive 2019/1151 of the European Parliament and of the Council of 20 June 2019, which modifies Directive (EU) 2017/1132 with regard to the use of digital tools and processes in the field of Company Law aims to "provide exhaustive and accessible information on companies" (Cons. 2) and registers must be interconnected (cons.29).

⁴³ The Directive states that In order to ensure maximum impact and facilitate reuse, high-value datasets must be made available for reuse with very few legal restrictions and at no cost. They must also be published via API. However, this does not prevent public sector bodies from charging for services they provide in relation to high-value data sets in the exercise of their public authority, in particular by certifying the authenticity or veracity of documents. (Cons. 69)

⁴⁴ <https://civio.es/novedades/2019/11/04/queremos-que-el-registro-mercantil-sea-abierto-y-gratuito/> (last consulted: February 2022).

⁴⁵ Companies House: <https://www.gov.uk/government/organisations/companies-house> (last consulted: February 2022).



insolvency information may be obtained. There are only costs in the case of certain more specific information. Greater transparency in company data has been particularly highlighted in this case⁴⁶.

More focused on the financial aspect, the Tax Justice Network has analysed two interesting aspects in our case, such as the transparency of companies with respect to the people involved and about their businesses –in seven European countries: Belgium, Denmark, Latvia, Romania, Slovenia, Sweden and the United Kingdom- (Heitmüller, Harari and Meinzer, 2019). With somewhat uneven results, it seems clear that these should include minimal information. Not all information needs to be public, but it should allow a person to be clearly identified (full name, document of address, document of date of birth, and a number issued by the company registry (Knobel, Meinzer and Harari, 2017).

b) The companies register in Spain

The access and reuse of companies register data is at the heart of the debate. The companies register can offer general data of a company, its social capital, the social representation -administrators and proxies-, the list of registered acts published in the Official Gazette of the Companies Register (BORME), the list of deposited accounts and the legalised books. The following table lists the possible certifications with company data that can be issued.

<ul style="list-style-type: none"> ▪ Certification of a specific position: This certifies the current position that a specific person holds in the organic representation of the company. ▪ Certification of validity and positions: This certifies the existence of the company and of all the current positions that make up the representative body of the company. ▪ Certification for the issuance of the electronic certificate of legal entity representative: This certifies the current position that a certain person holds in the organic representation or that they are a proxy, in order to request the electronic signature certificate of the company. ▪ Certification of powers: This certifies that a certain person is the representative of the company and what their powers are. ▪ Certification of corporate bylaws: This certifies the literal content of the company's current bylaws. ▪ Board meeting regulation certification: Certifies the literal content of the current board meeting regulations of the company. 	<ul style="list-style-type: none"> ▪ Council regulation certification: Certifies the literal content of the current council regulations of the company. ▪ Certification of parasocial agreements: Certifies the literal content of the current parasocial agreements of the company. ▪ Certification of powers of the CEO: Certifies the current position of Chief Executive Officer of a specific person, stating the form of action (several or joint) if there are several, and the powers delegated. ▪ Real Ownership Certification: Certifies the person or persons who are beneficial owners of the company, by personal contribution of more than 25% of the share capital (direct beneficial owner) or through another entity (indirect beneficial owner) or, failing that, the person or persons who are part of the administrative body (assimilated real owner), indicating their name and surnames, ID no., date of birth, nationality, country of residence, and, if
--	---

⁴⁶ Companies House's requirement to verify the identity of new company directors, persons of significant control and those incorporating companies, to help in the fight against money laundering or misuse of companies has been described as very positive (Hardman, 2021).



	applicable, the percentage of real ownership that they hold in the company.
--	---

Source: <https://sede.registradores.org/site/mercantil#/>

It is possible to access BORME free of charge to access information on the constitution of new companies and on appointments and dismissals of administrators, although with a time limit (2009) and with a cost.

The article 23.1 of the Commercial Code establishes that the Companies Register is public and that the publicity will be made effective by certification of the content of the entries issued by the Registrars or by simple informative note or copy of the entries and documents deposited in the Register, without motivation or interest being necessary for access. However, the Commercial Code⁴⁷ itself establishes that the certifications have a cost (article 23.2) and refers to the Mortgage Law for electronic advertising of the content of the Companies and Property Registers (article 23.4)⁴⁸. The Mortgage Law⁴⁹ in its article 222.2, establishes that said procedure must be carried out by a registrar, so that the possibility of advertising without intermediation is effective, ensuring, at the same time, the impossibility of its manipulation or remote emptying. The Companies Register Regulation⁵⁰ also establishes that the person responsible for processing the data is the registrar and that they must prevent the impossibility of its manipulation or remote emptying⁵¹.

It has been understood that the purpose of these records is not the identity of the persons or their assets, but rather the existence of certain movable and immovable property and the legal acts and transactions carried out on them. The regulations that regulate these registers try to make advertising compatible with the principle of quality and proportionality, preventing it from involving indiscriminate and excessive processing for the purpose of the Register (Troncoso Reigada, 2009). The legitimacy for access to information or data protection have been arguments to prevent access and reuse of such data.

It seems evident that there is useful and relevant data on potential contractors with the administration, their corporate structure, etc. and there is a position of the State in this regard. This is reflected, for example, in the IV Open Government Plan of Spain (2020-2024), which aims to promote the opening of data and indicates as regulatory objectives the transposition of Directive (EU) 2019/1024, regarding data and the reuse of public sector information, and the transposition of EU Directive 2019/1151 (Directive for the digitalisation of companies), which will *allow a clear improvement in access to data from the Companies Register*. In the axis on transparency and accountability, the adoption of *good practices in sectoral areas is pointed out: economic, budgetary, public employment, companies register*. This action is

⁴⁷ Royal Decree of 22 August 1885 by which the Commercial Code is published.

⁴⁸ Section added by Law 24/2001, of 27 December on Fiscal, Administrative and Social Order Measures.

⁴⁹ Decree of 8 February 1946 approving the new official wording of the Mortgage Law.

⁵⁰ Royal Decree 1784/1996, of 19 July approving the Companies Register Regulations.

⁵¹ The Companies Register is public and the Companies Registrar is responsible for the professional processing of the content of the register entries, so that their direct advertising is made effective and, at the same time, the impossibility of their manipulation or remote emptying is guaranteed (article 12.1). In 1998, 5 sections were added (arts. 12.4 to 8) on the creation of databases, which were deleted by STS on 24 February 2000.



indicated in the section on "reform of the regulatory framework" as one of the Plan's commitments to the SDGs of the 2030 Agenda (SDG 16).⁵².

In the same sense, the Independent Review Mechanism⁵³ highlights the Action Plan: Spain 2020-2024, which also includes an initiative to open the Companies Register and make its enquiries free of charge, in accordance with the transposition of European directive 2019/1151. (p. 13). The report affirms that "The opening of the companies register and it being free of charge would represent a substantial advance in meeting a recurring demand from various anti-corruption CSOs and defenders of integrity, since this would remove an important economic barrier (non-existent in other European countries that allows access to this information without payment or fee). This is key information to identify private interests (of individuals or business networks) that have or may have a direct relationship with decision-making, the management of public money, or public contracting processes. Although there is a private initiative that offers this service based on available public data, if it were done from public institutions, the commitment would provide a guarantee of the sustainability of the companies register as a public service".

Despite all this, initiatives to create databases on companies and contracts are private. Either they are part of business services for the provision of consulting or advisory services or they are individual initiatives. The two best known are LIBREBOR⁵⁴, which compiles the daily publications of BORME, and Contracts of Cantabria⁵⁵, which includes information on contracts and bidders from the Regional Government of Cantabria.

The Spanish Government has recently expressed its desire to open the Register of Beneficial Ownership (RETIR), a centralised database that makes it possible to find out who owns the companies. Its access will not be free and a digital signature will be necessary for its consultation. Although the opening of this source is positive, the situation in practice is the one described for the companies register.

3. Relevant experience in the use of digital tools for the prevention of corruption

3.1 Previous experience in the processing and management of information

At a comparative level, in Europe and the rest of the world, until approximately 2010 there were no major projects to implement data analysis systems for the prevention of corruption in public procurement. However, especially during the last decade, a series of interesting initiatives have been developed that must be taken into account in this work, since much of the work carried out and the experience gained can be useful, either to replicate methodological experiences of success or to avoid undertaking repetitive tasks and objectives that could reduce the scope of the project. In this regard, efforts have developed rapidly in the last decade, as governments adopt digital strategies and take advantage of open data, big data and data analysis:

⁵² Specifically on the Companies Register (action 2.4.3)

⁵³ Independent Review Mechanism. *Action Plan Review: Spain 2020-2024*

⁵⁴ Accessible at: <https://librebor.me/> (last consulted: February 2022).

⁵⁵ Accessible at: <https://contratosdecantabria.es> (last consulted: February 2022).



- **DIGIWHIST**, an acronym that comes from the expression Digital Whistle-blower, was born with the aim of empowering society to fight against corruption in the public sector. The role of the initiative is to systematically collect, analyse and disseminate information on public procurement for the entire EU. In addition, the platform also collects data to compare, contrast and assess potential conflicts of interest within the public procurement system. This initiative uses TED and EU Member State datasets to produce a consolidated database of contracts on the European level, both above and below thresholds. The experience of the Digiwhist project can be useful to review in order to understand how it identifies elements of risk, especially those that can also be analysed in emergency contracts, such as unrealistic or outlier contract values, the existence of multiple CPVs, etc.⁵⁶ Other related initiatives such as OpenTenders Portal or the Government Transparency Institute are born from this Project, which will be analysed later in the work.

The initiative includes in its research on public contracts various countries from the European Union (EU), Economic European Area (EEA) and European Community (EC). External links are used to facilitate the access to data, but the project also makes available 2009-2012 data series on public procurement for Czech Republic, Hungary, Slovakia, Macedonia, Switzerland and other European states. The territory of these countries is georeferenced nationally with the variable “announcing body location” and pinned at the NUTS 3 Level. Neither map nor other sort of graphic representation has been used to plot the data by the DIGIWHIST’s researchers on the portal. **DIGIWHIST has been given public funds from the European Union’s Horizon 2020** to conduct its research and propose innovative tools for the prevention and fight against corruption. As part of a consortium, universities and non-governmental organizations, such as the University of Cambridge and Government Transparency Institute, are part of its network. The Project DIGIWHIST shows two types of data on its webpage, i.e., *csv* and *dta* formats. **No data on public contracting during the emergency time caused by the Covid-19 was found.**

Regarding the re-usability of the quantitative data, the format *csv* and *dta* are easily downloadable with not cost, however, not so easy to be opened and understood. The Excel file should be transformed into *xlsx* or other readable extension while the *dta* only through specific programs. We used R Studio, but the *dta* files are originally designed for Stata software. There are documents of metadata containing an explanation about the list of variables employed, but the codes and keys are highly technical. The figure shows how the data on public procurement for Czech Republic, Hungary and Slovakia are visualized. Those files with special characters in Hungarian or Czech are more difficult to mine since it is not obvious for non-experienced users to automatically prepare their machines with utf-8 charset. Other data sets are more general without revealing the type of contracting or sector body, for example. The data series on EU, EEA, Macedonia and Switzerland.

⁵⁶ Accessible at: <https://ec.europa.eu/digital-single-market/en/content/digiwhist> (last consulted: February 2022).



	A	B	C
1	anb_id	anb_name	anb_nuts
2	SK00634816	Da=òov© riaditefæstvo Slovenskej republiky	
3	HU10856417	MvÁV Magyar vÁllamvasutak Zv°rtkvðr?en M?kvðd? RV©szvv©	HU101
4	CZ61389030	vöSTAV EXPERIMENTVÁLNvç BOTANIKY AV fâr, v.v.i.	CZ0100
5	HU309765	Orszv°gos Onkolvzgiai Intv©zet	HU101
6	SK00634816	Da=òov© riaditefæstvo Slovenskej republiky	
7	HU10856417	MvÁV Magyar vÁllamvasutak Zv°rtkvðr?en M?kvðd? RV©szvv©	HU101
8	HU10856417	MvÁV Magyar vÁllamvasutak Zv°rtkvðr?en M?kvðd? RV©szvv©	HU101
9	HU303082	ADvi- vâS PvâNZvúGYI ELLEN?RZVâSI HIVATAL (SZVÁMvçTVÁST	HU101
10	HU730886	Kiskunlachv°za Nagykvðzsv©g vñkormv°nyzat	HU102
11	HU10856417	MvÁV Magyar vÁllamvasutak Zv°rtkvðr?en M?kvðd? RV©szvv©	HU101
12	HU303082	ADvi- vâS PvâNZvúGYI ELLEN?RZVâSI HIVATAL (SZVÁMvçTVÁST	HU101
13	HU730886	Kiskunlachv°za Nagykvðzsv©g vñkormv°nyzat	HU102
14	HU10856417	MvÁV Magyar vÁllamvasutak Zv°rtkvðr?en M?kvðd? RV©szvv©	HU101
15	HU10856417	MvÁV Magyar vÁllamvasutak Zv°rtkvðr?en M?kvðd? RV©szvv©	HU101
16	HU10856417	MvÁV Magyar vÁllamvasutak Zv°rtkvðr?en M?kvðd? RV©szvv©	HU101
17	HU10856417	MvÁV Magyar vÁllamvasutak Zv°rtkvðr?en M?kvðd? RV©szvv©	HU101

Figure 2: Reproduction of an Excel Screenshot for Public Procurement at a National Level

Yet on quantitative data and its re-usability, there is a second section called EuroPAM refers to the European Public Accountability Mechanisms as a legal and regulatory norm database for 34 European nations. As an extension of the World Bank initiative, Public Accountability Mechanisms (PAM), it is designed to generate assessments of in-law and in-practice efforts to strengthen transparency in public administration as well as accountability of public officials. The supportive work of the DIGIWHIST has systematized information in *xlsx* tables and *pdf* files applying questions on political financing, financial disclosure, conflict of interests, freedom of information, and public procurement. According to a ranking methodology, they classified EU28, Armenia, Georgia, Iceland, Norway, Serbia, Switzerland with a score varying from 0 to 100 in 2012 and 2015.⁵⁷

The third section is called Budget Data and contains a list of civil society or governmental initiatives.⁵⁸ The projects referred produce, mine and reshape quantitative-qualitative data from national and regional public procurements in Germany, Italy, Poland, and Slovakia. The fourth and last section is about civil society organizations, think tankers, and non-governmental actors producing knowledge or sharing mostly qualitative information supported by facts, statistics and numbers related to the prevention and fight against corruption. The centers, partnerships and networks collaborating are Center for Open Data Enterprise, Czech Supreme Audit Office Nejvyšší Kontrolní Úřad, eGovernment4EU, Open Government Partnership, OpenCorporates and Spend Network.

- **OpenTED** is a now-defunct initiative that aimed to make TED information more easily accessible to journalists and researchers. Although the initiative is no longer operational, the source code of the

⁵⁷ The data made available by the DIGIWHIST initiative is divided into four sections. They combine qualitative and quantitative data. See <https://digiwhist.eu/resources/data/>.

⁵⁸ Barometr Ryzyka Nadużyć, Boost Initiative, World Bank, Dati.Piemonte.it, OffenerHaushalt, Open Bilanci, Open Public Procurement – Slovak tenders since 2009, Open Spending WG, PublicSpending, and Soldi Pubblici



developed solution is accessible through GitHub⁵⁹. The project team will review the data processing methodology to assess whether any practice can be reused.⁶⁰

OpenTED aimed at promoting access to the EU Tender Electronic Daily and use the TED's archive for journalistic and analytical purposes. The package monnet was developed and shared with the OpenInterests.eu project using the same database.⁶¹ It was conceived to be an open and free source with a code to mine the data on public procurement. **So far, what we can see is basically metadata and not including public contracting information. Figures on Covid-19 crisis were not found since the time series mined information from 2008 to 2015.**⁶² **There is no graphic representation of the data and the re-usability is just useful with basic knowledge in programming languages, map elaboration and the manipulation of data frames.** No reference about who financed the initiative was found. It seems the project is deactivated. No library monnet has been traceable in python repository for packages.

- The **“Red Flags” project**, launched in 2015, aims to improve the transparency of public procurement in Hungary and support the fight against corrupt procurement. It provides an interactive tool that allows the monitoring of contracting processes and their implementation by citizens, journalists or even public officials and detects fraud risks at different stages of the contracting process. The Red Flags tool automatically checks Tenders Electronic Daily (TED) acquisition documents and filters risky acquisitions through a special algorithm. Its database is available.

Red Flags is an API dedicated to Hungarian public procurements. The project follows the standards of EU regulations, and recommendations published by the Organization for Economic Cooperation and Development (OECD) as well as the European Bank for Reconstruction and Development (EBRD).⁶³ The warnings on public procurement are given through updated data of public procurement procedures highlighting individual risks and generally an overall threat category regarding public contracts.⁶⁴ The API requires a personal registration and, only after this first step, the user will receive a token in order to be able to run the Red Flag's code. **The project has had financial support of the *Prevention of and Fight against Crime Programme from the European Commission*. No data on public procurements specifically on the Covid-19 has been found.**

Concerning the re-usability of the quantitative data, Red Flags uses two file formats being one of them csv and the other JSON. However, to have full employment of the API built up by the initiative

⁵⁹ Accessible at: <https://github.com/opented/opented> (last consulted: February 2022).

⁶⁰ Accessible at: <https://github.com/Yannael/OpenTED> (last consulted: February 2022).

⁶¹ Accessible at: <https://github.com/opented/opented> (last consulted: February 2022).

⁶² Accessible at: <https://github.com/pudo/ted> (last consulted: February 2022).

⁶³ See the website of the initiative <http://api.redflags.eu/> (last consulted: February 2022).

⁶⁴ “Within the framework of the project ‘Prevention and detection of corrupt procurements through analysis, red flags and follow up’, the European Commission provided grant to Transparency International Hungary (TI), K-Monitor and PetaByte to create a warning system which accelerates the fight against corruption in the European Union (EU) by developing practical solutions that help the control of public procurement procedures and the identification of risks in the early stage of procedures. The project’s objective is to develop a risk-assessment methodology, which allows for the creation of a creative, innovative and interactive online monitoring tool that may be used in all member states of the EU”. Red Flags Project, *New Warnings System for the Identification of Red Flags in Public Procurements*, p. 4. See <https://www.redflags.eu/files/redflags-summary-en.pdf>.

one must be fluent in JSON and must bear in mind the utilization of code in order to visualize the figures on public procurements through HTML language. In short, the API has the purpose of intermediating the information published on TED’s portal and the use of the data by civil society, that is to say, **it is mostly an interoperable application.**

Figure below reproduces some of the columns of the most recently first 1,000 public procurements being the column “flagCount” the one indicating how many red flags were detected corresponding the winners in the column “winnerids”. Among the variables selected to generate the project’s indicators, we can find the *id* of the announcement, the *document family*, *date*, *contracting organization id*, *contracting name id*, and *deadline*. It is also possible to check the *type of contract*, *estimated amount*, *currency*, *winner name*, *winner id* and *procedure type*. **The result is the number of red flags related to each contract based on 31 indicators for contract notice and 9 indicators for contract award notice.**

	P	Q	R	S	
1	flagCount	contractType	procedureType	winnerids	winnerNa
2	2	Services	Open procedure	ORG-759070f3fb6fb8bb5e6419dd886a2b67	SZENTINE
3	2	Services	Open procedure	ORG-759070f3fb6fb8bb5e6419dd886a2b67	SZENTINE
4	1	Services	Open procedure	ORG-6b130872680fbbbc4209b892e6f5147c	DEBRECEI
5	1	Services	Open procedure		
6	2	Services	Open procedure	ORG-d998de765cd94dca080d31d1df81fa7d	Affidea M
7	0	Services	Open procedure	ORG-c6d0f9aaa75348024a9356c4d53c5de4	KROMAT
8	0	Services	Open procedure	ORG-40a566c8ac197c0f01c92e0669c4ea46	F=EBER N
9	0	Services	Open procedure	ORG-094c3cab23eada21d3a8787994b9370f	MIDDLE C
10	0	Services	Open procedure	ORG-ca8569df1abfd6bb56bc7b29c41ef87	Amanda C
11	0	Services	Open procedure	ORG-fc8a78b17893c97f44389bb691dad140	KEREKERI
12	0	Services	Open procedure	ORG-4573607f07cbd542af39e258ed8208cc	SAGRA vâ
13	0	Services	Open procedure	ORG-8729adbbc119029218cb87c7c67979d	P. DUSSV
14	0	Services	Competitive procedure with negotiation	ORG-7a368102a0da9286164b0980ee35ecb9	Julienne S
15	0	Services	Competitive procedure with negotiation	ORG-9197fd07d0d36debdebc41060ab9f32b	Julienne K
16	0	Services	Open procedure	ORG-8525dce4d116351ae214375774924191	UVATERV
17	0	Services	Open procedure	ORG-485087f532c10b1633cde32f32dac6d2	Create Va
18	0	Services	Open procedure	ORG-efe7da690961363ca83026dc460751e	LOUNGE I
19	0	Services	Open procedure	ORG-04b3d09c5d1ad9996561bdd8d1c8677f	KOSZ-FAL
20	0	Services	Open procedure	ORG-61f000d95256d7cbf2e4688685527688	Nett Proj
21	0	Services	Open procedure		
22	0	Services	Open procedure		
23	0	Services	Restricted procedure		
24	0	Services	Open procedure		
25	0	Services	Open procedure	ORG-7d8e43b766214ea0709a9104375fad5e	S&T Cons

Figure 3: Reproduction of an Excel Screenshot for Red Flags in Public Procurement

The sources of the databank made available by the initiative come from the Tenders Electronically Daily: “Indicators were listed and the expressly negative risk warnings, the red flags were developed based on the above and the contract notice data in TED”.⁶⁵ No graphic visualization is available for the users.

⁶⁵ Red Flags Project, *New Warnings System for the Identification of Red Flags in Public Procurements*, p. 9. Accessible at: <https://www.redflags.eu/files/redflags-summary-en.pdf> (last consulted: February 2022).

204983-2022	Contract award notice Személy és vagyonőr - Paks II - 1.1.2.-es épület Hivatkozási szám: EKR000745...	9 M HUF	
Apr 15, 2022	Contracting authority: Paks II. Atomerőmű Zrt. Award winners: SZENTINEL Vagyonvédelmi, Biztonság-Technikai Szolgáltató Kft.		
204982-2022	Contract award notice Személy és vagyonőr - Paks II. - Dankó P. - 2021 Hivatkozási szám: EKR001152...	9.5 M HUF	
Apr 15, 2022	Contracting authority: Paks II. Atomerőmű Zrt. Award winners: SZENTINEL Vagyonvédelmi, Biztonság-Technikai Szolgáltató Kft.		
204955-2022	Contract award notice DEK-1077: Rendezvényszervezési szolg. (saját) Hivatkozási szám: EKR000255532...	600 M HUF 600 M HUF	
Apr 15, 2022	Contracting authority: Debreceni Egyetem Award winners: DEBRECENI CAMPUS Szolgáltató és Tanácsadó Nonprofit Közhasznú Kft.		
204879-2022	Contract notice Személy- és vagyonvédelmi feladatok ellátása Hivatkozási szám: EKR0003446520...		
Apr 15, 2022	Contracting authority: Budapest Környéki Törvényszék Deadline: May 4, 2022 (19 days after publication)		
204835-2022	Contract award notice Diagnosztikai szolgáltatás beszerzése OSEI részére Hivatkozási szám: EKR0007...	240.8 M HUF	
Apr 15, 2022	Contracting authority: Országos Sportegészségügyi Intézet Award winners: Affidea Magyarország Egészségügyi Szolgáltató Kft.		
204779-2022	Modification of a contract/concession during its term Analitikai készülékek karbantartása, javítása Hivatkozási szám: EKR00033332...		
Apr 15, 2022	Contracting authority: Nemzeti Szakértői és Kutató Központ Award winners: KROMAT Műszerforgalmazó Kft.		

Figure 4: Reproduction of Red and Pink Flags in Public Procurement

On the other hand, the project offers qualitative information through the system of pink flags. In cooperation with K-Monitor, Red Flags has created a pink-flag signal with company or public administration references found on newspaper articles. That indicator is called “other risks” for public procurement in Red Flag’s methodology. Figure above illustrates how these two alerts go hand in hand.

Concerning the re-usability of the Red Flags’ information, the data is open and free. Nevertheless, some knowledge on programming languages is needed. In other words, it is not so easy for journalists, academics and citizens with non-data science experience to re-use the API code created by the project. Moreover, the number of columns or folders with more or less complete information may vary from one country to another on the website of Tenders Electronic Daily (TED). In other to do so, the user of the API algorithm must depurate the database working on, for example, the same number of columns or folders. Figures show how the data available on TED’s portal may differ from country to country affecting, consequently, the re-use of data or interoperable APIs such as Red Flags.

Original language | Data

15/04/2022 S75
I. II. IV. V. VI.

Norway-Trondheim: Environmental management
2022/S 075-206231
Voluntary ex ante transparency notice
Services

Legal Basis:
Directive 2014/24/EU

Section I: Contracting authority/entity

I.1) Name and addresses
Official name: Miljødirektoratet
National registration number: 999 601 391
Postal address: Brattørkaia 15B
Town: Trondheim
NUTS code: NO Norge
Postal code: 7010
Country: Norway
Contact person: Hilde Nystad
E-mail: hilde.nystad@miljodir.no
Telephone: +47 95411606
Internet address(es):
Main address: www.miljodirektoratet.no

Figure 5: A Snapshot of a Norwegian Call with two tabs

Current language | Original language | Summary | Data | Document family

Display compact view

15/04/2022 S75
I. II. IV. V. VI.

Hungary-Budapest: Repair and maintenance services of motor vehicles and associated equipment
2022/S 075-206024
Contract award notice for contracts in the field of defence and security
Services

Directive 2009/81/EC

Section I: Contracting authority/entity

I.1) Name, addresses and contact point(s)
Official name: Védelmi Beszerzési Ügynökség Zártkörűen Működő Részvénytársaság
National registration number: AK27368
Postal address: Schweidel utca 2-4.
Town: Budapest
Postal code: 1118
Country: Hungary
Contact person: dr. Fail Kinga vezérigazgató
E-mail: beszerzes@vbuzrt.hu
Telephone: +36 13819400
Fax: +36 13819401
Internet address(es):
General address of the contracting authority/entity: <https://www.vbuzrt.hu/>
Address of the buyer profile: <https://www.vbuzrt.hu/>

I.2) Type of the contracting authority

Figure 6: A Snapshot of a Hungarian Call with four tabs

- **Open Spending EU Vision.** It is a collaborative initiative between non-governmental and professional organisations to promote more open information management in Europe. Its



manifesto raises some problems and some associated measures to facilitate data collection, such as the consideration of public procurement as a high-value data set in Annexe 1 of the Open Data Directive; EU-wide implementation of open contracting data standards, including Europe's daily electronic tenders; more transparency in sub-threshold recruitment across the EU; public lists of excluded bidders: mandatory use of the European Commission's data mining tool; stricter transparency requirements within the regulations of future EU funds; transparency of the company and of the real beneficiaries of public aid; full disclosure of company ownership information under the Open Data Directive; Full publication of beneficial ownership registers under the Anti-Money Laundering Directive⁶⁶.

Open Spending EU Vision is a collaborative non-government network dedicated to ensure governments use public money fairly, openly and efficiently in Europe. **Different from DIGIWHIST, OpenTed and Red Flags, it scans those public procurements derived from the EU public funds.** It also pushes “for full, open and free access to all company registration data, including beneficial ownership and company ownership registers within the EU”.⁶⁷ Although public contracting is not covered in all Europe, various countries like Austria, Hungary, Italy, Lithuania, Netherlands, Poland, Portugal, Romania, Slovakia, and Spain have observers.

The databases used by the project comes from tools containing indicators and indexes from other partners such as Tenders Guru, AskTheEU.org, ManoValstybe, Defence Elvis, Libellula, Red Flags, Covid-19 Explorer, and Transparência (Transparency International Portugal). On the other hand, the main role of project is to advocate: “for user-friendly, accessible information across the entire procurement process by European governments from the planning, to the tender and award of contracts, to their delivery”. **In short, Open Spending EU Vision does not produce data, but written and quantitative visual such as graphics or tables on public procurement and public funds related to EU.** Its sources are based on what other initiatives generate in terms of **quantitative-qualitative evidence. No icon, image or text about how the Open Spending EU Vision is financed has been found.**

In order to achieve its goals based on the principle of transparency and participation, formal statements, calls, recommendations, and guidance are published on its official website. The network makes also public short analysis on how the EU public funds are spent using concise textual information and graphics. **Particularly about the emergency context triggered by the Covid-19 pandemic, a report was launched to communicate how and where the €672.5 billion in loans and grants from the EU Recovery and Resilience Facility (RRF) have been used by governments.**⁶⁸

⁶⁶ Available at: <https://www.open-spending.eu/vision/>

⁶⁷ See <https://www.open-spending.eu/vision/>.

⁶⁸ See https://www.open-contracting.org/wp-content/uploads/2021/06/RRF_report.pdf.



There are other experiences in the processing and management of information which are briefly analyzed below:

- **Public Procurement Indicators (PPI)**⁶⁹ is an initiative of DG GROW with the aim of presenting the main indicators on public procurement for each Member State during a specific year, using the available TED data. It is of utmost relevance for this project to compare the methodology used by the PPI initiative for the management of certain problems faced during this project, such as the calculation of missing values, the methodology of revision of outliers, the generation of contractual values, etc.
- **Poland, Romania, and Hungary Tender Tracking**, are a set of platforms developed by the Government Transparency Institute and Corruption Risk Centre in Budapest, thanks to funding from the Open Society Institute, which calculates the Corruption Risk Index (CRI) of tenders.
- **Public procurement Due Diligence Tool**, prepared by GAN Integrity (and therefore a private initiative) published on the Business Anticorruption Portal, is a tool developed for the evaluation and prevention of corruption risks in public procurement, consisting of three documents and an interactive tool. Based on the questions provided by the tool, the different phases of the public procurement procedure must be analysed and, based on the answers given, an evaluation matrix identifies the degree of risk for transparency and the procedure.
- **Arachne** is a specific data mining tool offered by the Commission to detect projects that may be at risk of conflict of interest, which aims to increase the effectiveness of project management and selection controls and help strengthen the identification, prevention and detection of fraud. Developed by the European Commission and OLAF, it can be used by any entity that manages Structural Funds (ESF and ERDF). Arachne's risk scoring tool identifies more than 100 risk indicators, which are grouped into 7 risk categories, such as acquisitions, contract management, eligibility, performance, concentration, other reputation and fraud alerts.
- **DATAACROS** – Developing A Tool to Assess Corruption Risk factors in firms' Ownership Structure is a research project co-financed by the European Union Internal Security Fund, aimed at developing a prototype tool to detect anomalies in the ownership structure of companies that may indicate high risks of collusion, corruption and money laundering in the single European framework. The DATAACROS project had a duration of 24 months, ending in February 2021. The second phase of the project, DATAACROS II, was scheduled to start in January 2022, with the aim of improving the tool and optimising its application to detect anomalies in the ownership structure of companies.⁷⁰
- Outside the EU, other interesting projects can be identified such as the **Corruption Risk Index in Lebanon**⁷¹, which brings together qualitative and quantitative data that helps to try to assess the points of risk in the Lebanese public sector, using a series of indicators applied, among others, to

⁶⁹ Accessible at:

<https://ec.europa.eu/docsroom/documents/38003/attachments/1/translations/en/renditions/native> (last consulted: February 2022).

⁷⁰ Accessible at: <https://www.transcrime.it/datacros/> (last consulted: February 2022).

⁷¹ Accessible at: <https://www.corruptionriskindex.com> (last consulted: February 2022).



public contracting and conflicts of interest. And in the same vein, the Indonesian tool⁷² **Corruption Watch** can be highlighted, a Platform that analyses national procurement data and publishes its findings, which is proving to be a powerful tool for tracking irregularities and ensuring that publicly accessible data is translated into the public understanding of public spending. Finally, **M&E Platform**, a Vietnamese platform developed by Development Gateway that extracts data from the country's eGP system and flags suspicious behaviour, can be highlighted.

Others could be added to these instruments that, indirectly, have contributed to progress in the implementation of this type of technique. This is the case of the various guides that develop check-lists of indicators, the public and private initiatives that have collected data and show it in a structured way, or others that develop processing techniques that can be incorporated into the supervision process. Some of them will be mentioned in the following pages, which will address how taking advantage of these experiences can serve to improve the availability of data, compile indicators that have already been tested, or implement management techniques.

3.2 Corruption Indicators

The inability to normatively capture all the possible behaviour that can cover up a case of corruption make the indicators a key element for an effective management of the problem. Corruption indicators can be understood as common symptoms that contracts affected by corruption frequently show⁷³; in other words, their use and design has focused on understanding what is normal in a given context in an attempt to detect anomalies. Such anomalies are "red flags" of suspicious behaviour, warranting further investigation. The use of these tools, which therefore facilitate early and effective detection of anomalies, began to be explored especially at the end of the 1990s (Poerting & Vahlenkamp, 1998). Thus, the indicators were used in the area of fraud as early as 2002 by the *American Institute of Certified Public Accountants (AICPA)*⁷⁴, and later, in 2007, the World Bank adopted a new Governance and Anticorruption Strategy in which it recommended the use of red flags to locate cases of corruption in the projects it finances (World Bank, 2007). Since then, the use of this type of indicator has been spreading and developing in an extensive scientific literature with the aim of minimising the effects of irregularities in contracting procedures and in other areas of public activity (Kane & White, 2009); (Di Nicola & McCallister, 2006).⁷⁵

In this way, the European Commission prepared, already in 2009, an informative note on the indicators of corruption and fraud in public procurement procedures.⁷⁶ Subsequently, in the report *Identifying and Reducing Corruption in Public Procurement in the EU* commissioned by the Commission and published in 2013, 27 red flags were identified that were considered especially relevant when detecting irregularities and corruption (Wensink & de Vet, 2013). For its part, the OECD developed, among other documents and guides (i.e., (OECD, 2003); (OECD, 2005); (OECD, 2007), an indicator implementation tool, in

⁷² Accessible at: <https://opentender.net> (last consulted: February 2022).

⁷³ OLAF. Detection of conflicts of interest in public procurement procedures within the framework of structural actions. Practical guide for those responsible for management. Brussels: European Commission, 2014, p. 25 et seq.

⁷⁴ AICPA. *Consideration of Fraud in a Financial Statement Audit*. 2002. Section 316, 1719 – 1770. Access on 19 October 2017. Available at: www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-00316.pdf

⁷⁵ The Open Contracting Partnership has even produced a non-exhaustive list of reports and bibliography on the matter: <https://docs.google.com/spreadsheets/d/12PFkUIQH09jQvcnORjcbh9-8d-Nnluk4mAQwdGiXeSM/edit#gid=245962708>

⁷⁶ EUROPEAN COMMISSION. *Information Note on Fraud Indicators for ERDF, ESF and CF. EC DG REGIO, COCOF 09/0003/00-EN*. 2009. Access on 08 October 2017. Available at: www.eufunds.bg/document/271



which it identified a series of corruption indicators that it made available to contracting entities for implementation⁷⁷. In recent years, as stated above in this work, a whole series of initiatives are being developed that are testing and refining these indicators⁷⁸, among which the one carried out by TI Hungary can be highlighted, with the collaboration of the European Commission and OLAF, in the so-called Red Flags project, the initiative has led to the creation of a red flags alert computer program that analyses the presence of indicators in the contracts collected on the EU's Tenders Electronic Daily (TED) electronic platform. This platform, although it is aimed especially at those in charge of supervising the procedures, is open access and can be used by the rest of public employees as well as private companies and the civil population in general (Anita & Tünde, 2015).

In short, throughout this study process on the subject by the most important organisations and experts in the fight against corruption, a large number of indicators have been developed, which, however, vary in their degree of usefulness and impact. Thus, KENNY and MUSATOVA classified in 2010, in a study for the World Bank, the existing indicators to date into four types (Anita & Tünde, 2015); (Kenny & Musatova, 2010):

- *Unobservable indicators*: these are the red flags that are difficult or impossible to assess, even by carrying out “on-the-spot” supervision of the contract. This is where pressure is put on the members of the contracting board, falsification of CVs in the provision of services, etc.
- *Indicators that cannot be collected*: are those that, although they can be observed in a standard supervision or control process, do not leave a palpable trace in the contracting documents, and are difficult to identify in an “office” examination, These include, among others, the elaboration of technical specifications or selection criteria in excess limiting the competition.
- *Irrelevant indicators*: this type of red flag is found in a large number of contracts, both corrupt and intact, and although they may be indicators of corruption, their relationship with the phenomenon is diffuse and a clear relationship cannot be confirmed. E.g.: the lack of publicity of certain aspects.
- *Relevant, detectable and collectible indicators*: these types of indicators are those that must be especially taken into account when evaluating the possibility of corrupt acts in a contract, since they are considered directly related to the phenomenon and are easily observable.

As is evident, although other types of indicators can complement the study of a contract suspected of corruption, the truth is that in order for these to be effective, so that the supervision and control agencies can select the contracts to be investigated, special attention should be given to visible and relevant red flags. This type of classification, however, is not static. Today, the existence of numerous indicator systems, together with the implementation of digital systems, enable a greater rationalisation of efforts, attending to risk areas and not merely legal or cyclical criteria. The flow of information must adequately allow risk situations to be identified on the micro level (contract by contract) and global evaluations to be carried out on the macro level (in the contracting system on the national or European level). These tools can turn certain indicators that until now were considered unobservable or uncollectible into observable indicators. Thus, those indicators that are difficult or impossible to assess even by carrying out an "on-the-spot" supervision

⁷⁷ OECD PP *Toolbox: Tool: indicators of procurement risk* (2009). Available at: <https://www.oecd.org/governance/procurement/toolbox/search/indicators-procurement-risk.pdf>

⁷⁸ In addition to those already presented in the initial part of the work, an interesting compilation of these indicator tools has been prepared by (Tátrai and Németh, 2018)



of the contract, as is the case of pressure on the members of the contracting board, falsification of curriculums in the provision of services, etc., or those that do not leave a palpable trace in the contracting documents, and are difficult to identify in an "office" examination, such as the preparation of technical specifications or excessive selection criteria that limit competition, can now be identified if certain technological tools are applied.

In any case, in recent years there has been constant work, both by institutions and by academia and science, to develop more precise, more easily observable and compiled indicators. Thus, there is wide scientific literature and numerous experiences on the matter, from which some characteristics that have a greater specific weight as a sign of irregularities can be identified (non-exhaustive list)⁷⁹:

- a) The deadline for submitting bids is excessively short or does not meet the minimum standards.
- b) The use of the *urgent procedure* or other non-competitive procedure.
- c) Contracts entered into subtly *below thresholds* that impose additional transparency or competition obligations.
- d) *Limited number* of bidders. Those procedures in which only 3 or fewer bids are submitted offer greater facilities when manipulating the procedures.
- e) The conclusion of *framework agreements*, due to the characteristics of these procedures, which limit competition and generate a lasting relationship between the contractor and the contracting party.
- f) There is a large percentage of *bids declared excluded* that can cover up cases of collusion and distortion of competition.
- g) Excessive *time spent evaluating bids*.
- h) The *technical or solvency characteristics* that define the object of the contract are adjusted only to one of the supplier companies in the market. The World Bank provides some sub-indicators, such as the presence of unnecessary or inappropriate specifications, vague and/or scarce descriptions and information, or the exclusion of a large number of bidders.
- i) The *size* of the contract seems excessively large and/or includes very diverse benefits.
- j) The *duration* of the contract. If it is renewed periodically indefinitely or recurrently.
- k) There have been *appeals* or other types of claims by the bidders regarding the impartiality of the contracting entity.
- l) There are *relationships* between the bidding companies – organisational or otherwise – that could have jeopardised effective competition between them.
- m) There are *strange behaviours between the submitted offers*: they are very close or very far from each other, there are abnormally high offers or with round figures (which may imply false offers to meet

⁷⁹ Sources and studies in which more developed indicators and red flags can be found include: International Anti-Corruption Resource Centre (IACRC). *Red Flags of Corruption, Bid Rigging, Collusive Bidding and Fraud*, available at <https://guide.iacrc.org/the-red-flags-of-corruption-bid-rigging-collusive-bidding-and-fraud/>; (Németh and Tátrai, 2015); GAN integrity. *Public procurement Due Diligence Tool*, <https://www.ganintegrity.com/portal/tools/due-diligence-tools/public-procurement-due-diligence-tool/>; Corruption Risk Index (CRI): <https://www.corruptionriskindex.com/>; Open Contracting Partnership. *RED FLAGS for integrity: Giving the green light to open data solutions*, 2016. Specifically, the list of indicators can be found at: <https://docs.google.com/spreadsheets/d/12PFkUIQH09jQvcnORjcbh9-8d-Nnluk4mAQwdGiXeSM/edit#gid=2027439485>; (Bernstein, Recanatini and Georgieva-Andonovska, 2018); (Carloni, 2017); (Kenny and Musatova, 2010); (Wensink and de Vet, 2013); (Integrity Vice Presidency, 2009); (Ferwerda, Deleanu and Unger, 2017).



- the competition requirements), rotation of the winners in successive contracts, losing bidders become subcontractors, etc.
- n) The same company obtains *successive contracts* concluded by the same contracting entity.
 - o) Contract award documents are not given adequate *publicity* or all participants are not informed equally.
 - p) Substantial *modifications* have occurred within a short period of time after the award of the contract, or in a way that is clearly beneficial to the bidder (reduction of services without reduction in price, increase in personnel without increase in services or materials, etc.).
 - q) *Deficiencies are observed in the execution of the contract* or it is not executed.

These and other indicators that can be found in the aforementioned literature have been tested in different studies, so that their effectiveness in one context or another, or at a micro and macro scale, can also be evaluated in terms of their use and incorporation (Charron, Dahlström, Fazekas, & Lapuente, 2017); (Mihály Fazekas & Kocsis, 2020). For this, some strategic decisions about the scope of the project already mentioned above will have to be taken into consideration, such as: the thresholds above which information will be collected and the indicators that will be applied; the territorial scope of the sample; the potential user of the system of indicators created (because a system for use by citizens, external supervisory bodies, or the contracting body itself and the contract managers will be different. These issues must be taken into account, together with the particularities of emergency contracts, when adapting the above indicators and those used in the scientific literature cited in this study.

3.3 Indicators in emergency contracts

Additionally, for the purpose of this project, it must be taken into account that, when resorting to an emergency procedure, the procedure is often dispensed with almost entirely, and many of the indicators indicated above are somewhat unusable due to their close relationship with the procedural characteristics or formal requirements. Consequently, the use of or promotion of other possible specific indicators of the contracts entered into in times of emergency. Some of these indicators are specific to these exceptional situations and have been identified by scientific work on the subject, and specifically, the following can be highlighted (Abdou et al., 2021):

- *Market change.* Non-health providers who provided services in the medical products sector during the emergency. These providers can hide irregular practices of abuse of the changing regulatory framework during the emergency by changing the domain of the activity.
- *Newly created companies or risk companies.* Identifying the companies that had not previously contracted with Public Administration is relevant, as they could hide hidden changes in activity – taking advantage of the fact that in these cases, solvency requirements related to previous experience are not required, as a general rule. There are other indicators traditionally used to identify potential risk contractors, such as operations in tax havens or other risk factors.
- *Size of the company.* Although in ordinary contracting it does not have to be understood as a relevant indicator, the contracting of large international amounts of supply with small companies can hide irregularities.
- *Geographic proximity or distance.* In emergency contracts, local suppliers as a risk category, since the geographical proximity between the supplier and the buyer can be indicative of collusion in contracting. Conversely, distant contractors (especially those located in more corrupt areas or tax

havens) who win small-scale contracts can target corruption (Mihály Fazekas, Cingolani, & Tóth, 2018); (Caneppele, Calderoni, & Martocchia, 2009).

- *Relations between contractors and public employees of the contracting entity.* This indicator may be interesting because it does not depend on the type of procedure used, and would remain valid in the use of extraordinary procedures. However, it should be kept in mind that the personal connections between political office holders and private companies are diverse (Mihály Fazekas et al., 2018). It is a complete phenomenon that occurs differently at different central or peripheral procurement levels, in different sectors, and often has a legal treatment that varies from one Member State to another (Miranzo Díaz, 2020).
- Type of owner/s of the contractor company. The lack of information about the actual ownership and management of the company can point to corrupt behaviour (Aziani, Ferwerda, & Riccardi, 2020).
- Finally, irregularities caused by repeated hiring of the same specific health material have also been identified. This is reiteration that would weaken the extraordinary nature of the emergency situation, and consequently, a second indicator could be proposed on the use of these extraordinary procedures, related to the reiteration or chaining of contracts with the same object.

3.4 The identification of relevant information

So that the exposed indicators or others that may be developed can be properly evaluated, it is also necessary to identify what data may be relevant for each of the indicators, as well as their availability and possible collection problems. At a comparative level, some studies have proven the relevance of certain procedural data that, not being direct indicators, can act as proxy or indirect indicators, or in any case, offer relevant information for subsequent data crossing and thus be able to predict or identify possible risks of corruption in this type of procedure (Auriol, Straub, & Flochel, 2016). This is the case of the initial value of the contract, the date of approval and the duration, the publication of the tender notice, the formalisation agreement or other documents, the existence of extensions and/or modifications, the nature and identity of the contracting entity, the address of the contractor (to identify, for example, whether or not it operates in a tax haven), the percentage of expenditure executed in favour of the same contractor (Abdou et al., 2021), and the type of process used to select the contractor (Hyytinen, Lundberg, & Toivanen, 2007), if there is a direct award, etc. (Gallego, Prem, & Vargas, 2020).

In the same sense, DIGIWHIST identified as relevant data sources those related to company registration information (company name, ID, date of incorporation, address, company size, etc.), financial data (volume of annual turnover, profit rate, liabilities, etc.), owner and manager information (Fazekas et al., 2018). Other recent studies have shown how other organisational information from Public Administration (Miranzo Díaz, Cerrillo i Martínez, Castro, & Galindo Caldés, 2022) or financial data (Decarolis & Giorgiantonio, 2022) can contribute to creating proxy indicators of corruption. Similarly, an important element identified and often left in the background, but which can also become more relevant in emergency situations, is the identification of the buyer and the seller (Mendes & Fazekas, 2017). The pandemic has given rise to contracts in which the identity of the successful bidder, their address or other data has not been published⁸⁰, which is a fundamental element. In the same way, it is sometimes problematic to identify, when

⁸⁰ In the case of Spain, the following cases can be seen: LA RAZÓN: “The suspicious contracts signed by the Illa Ministry of Health in the first months of the pandemic” <https://www.larazon.es/sociedad/20220218/3gwp4mro7nez5ihugdgsmsneiei.html> ; COLL, B. “35 million picked out for an unknown company: the "reckless" contract of the Generalitat that questions Antifraud”, <https://elpais.com/espana/catalunya/2022-02-22/35-millones-a-dedo-para-una-empresa-desconocida-el-contrato->

faced with similar names, when we are dealing with the same contracting authority or a different one (Mendes & Fazekas, 2017). The use of identifiers for public purchasers would greatly contribute to sometimes avoiding the duplication of contracting entities that are actually the same, allowing greater monitoring of trends and possible purchasing irregularities.

At this point it should be mentioned that there are also certain shortcomings identified in regard to the extraction of the relevant data. First, although national business registries do exist, they are not always free to use and often contain only a limited set of information. The DATACROS project has carried out an aggregated analysis of company ownership anomalies in the EU-27 and has designed a prototype tool for the risk assessment of legitimate companies, capable of detecting anomalies in the ownership structure of companies that can signal high risks of collusion, corruption and money laundering⁸¹. However, it has not been possible to locate on the website that the database used is available, nor that the tool can be used in other environments or in conjunction with other tools.

Likewise, another element that has been pointed out as problematic in other works is the lack of information in relation to the execution of the contract in most European countries. The regulatory frameworks stop the contracting requirements—except for some issues related to modification—in the award of the contract. Information on the execution of contracts is often recorded in other administrative systems and sometimes a more or less differentiated legal framework is applied to them, making monitoring and analysis highly complex. Even on many occasions, modifications and unsuccessful bids are not properly recorded, or at least there is a suspicion that the available data is not reliable (Mendes & Fazekas, 2017). This execution data seems especially important in contracts concluded by procedures without publicity, with emergency procedures or through direct awards, since in this type of contract the elements to be controlled in the previous phases of the procedure are especially scarce (Cingolani et al., 2015).

3.5 The possible implementation of AI systems in the creation of indicators and detection of irregularities

To all the previous experience, we should add a differentiating element that can play an important role in the design and management of indicators and corruption risks in the future: AI systems. There is not yet extensive experience in the implementation of this type of technology, since most of the systems and tools exposed so far cannot be considered AI, or at least not in the most advanced versions of this technology such as Deep learning. However, yes, some interesting scientific and practical initiatives can already be ruled out. On the science side, some studios are experimenting with new ways in which AI can help design new indicators or identify complex anomalies, such as collusion or bid rigging, more effectively than traditional systems (Rabuzin & Modrusan, 2019); (Huber & Imhof, 2019).

In the case of practical experiences, the clearest example of AI implementation to date is the DoZorro tool, developed by TI Ukraine and other development cooperation organisations such as the EBRD, which presents significant technological advances with respect to the rest of the algorithms analysed until now, which make it particularly interesting, but which at the same time bring new challenges and legal risks (Köbis, Starke, & Rahwan, 2021). Unlike the rest of the aforementioned analysis systems, *DoZorro* artificial

[temerario-de-la-generalitat-que-cuestiona-antifraude.html](https://www.elmundo.es/economia/2022/02/18/620fbc6ce4d4d8a3188b458b.html) ; SEGOVIA, C. “The Court of Auditors certifies lack of control in express purchases of medical supplies from 13 State agencies in 2020”, available at <https://www.elmundo.es/economia/2022/02/18/620fbc6ce4d4d8a3188b458b.html>

⁸¹ Available at: <https://www.transcrime.it/datacros/> (last consulted: February 2022).



intelligence is very different from traditional risk indicators, since in these cases there are not previously definitively established indicators.

It should be noted that the Ukrainian Administration had previously developed an analysis system based on pre-established indicators called *ProZorro*. In it, the State Audit Service of Ukraine managed since 2018 the automatic verification of offers in *ProZorro* on the basis of 35 risk indicators. They include violations of the bidding procedure, bidders not having the power to participate, absence of digital signatures, violations of review deadlines, absence of the concluded agreement, absence of electronically uploaded bidding documentation, absence of the description of the purchased item, a large number of lots included in a tender, etc. The more indicators that are identified by the program concurring at the same time in a contract, the greater the assigned risk will be, and therefore the greater the possibility that the offer will be inspected by the auditors.

The *DoZorro* system, however, is more flexible and does not have an exhaustive list of indicators. It is a machine learning neural network system—in other words, what has been considered AI, strictly speaking—that began learning to—or being trained to—identify procurement violations in July 2018 (Petheram, Pasquarelli, & Stirling, 2019). The developers sent 20 experts around 3,500 offers to analyse. They were asked to answer a single question, whether they are at risk or not. The experts did not know the amounts or the names of the contracting entities so that their evaluation was as objective as possible. Finally, all responses were fed into the AI algorithm. In a deep learning system like *DoZorro*, therefore, the system is "trained" and learns experimentally, developing its own indicators and criteria.

The use of deep learning and machine learning techniques, either supervised or autonomous, can be considered in this project for different reasons. In the first place, due to the scarcity of indicators currently identified by practice and academia and the lack of empirical results on their application due to the lack of practical testing. Secondly, due to the non-formalist nature of emergency adjudications and the potential lack of information, which make it especially difficult to find clear elements of risk, such as lack of competition, lack of information, excessively complex technical documents or simple etc. An AI system could identify new patterns that are currently unobservable or uncollectible from the traditional pattern identification perspective. At this point, however, the degree of feasibility with respect to the application of machine learning techniques must be analysed, and where appropriate, their compatibility and integration with other previously related analysis techniques.

4. Case study: the availability, quality, reusability and interoperability of data on public procurement in Catalonia

4.1 Introduction

The conceptual framework, the guiding principles and the regulations governing the availability, quality, interoperability and reusability of data have been presented and analysed in the previous sections of this study.



We are now going to change our point of view, and try to move from a theoretical perspective to a practical one. Hence, we are going to analyse the availability, quality, reusability and interoperability of procurement data of the public sector in Catalonia.

There are two reasons to focus on this specific case.

On the one hand, as we shall see shortly, the Generalitat de Catalunya —the Catalan regional government and administration— has been a pioneer in the digitalisation of public procurement activity. Therefore, the data, the raw material that feeds the corruption risk indicator systems such as those that are the object of the CO.R.E project, have been generated in an operational context and on largely digital supports. This favours a priori their availability, accessibility, interoperability and reusability. We will have the opportunity to analyse to what extent this *a priori* actually do occur.

On the other hand, one of the main objectives of the Data Analysis Team of the Anti-Fraud Office of Catalonia, partner of the CO.R.E. project, is focused at present specifically on the collection, structuring, analysis and exploitation of public procurement data of the Catalan public sector, to support the main missions of this Office: «To study, promote and foster the application of good practices in relation to transparency in public management and the prevention of and fight against fraud» in public administrations —article 3 a) of Act 14/2008, of 5 November, on the Anti-Fraud Office of Catalonia—. The operational work which is being carried out by that Team constitutes an ideal test bench in which to verify to what extent the theory presented up to this point materialises in practice.

4.2 Context: procurement activity datafication in the Catalan public sector

As mentioned above, the field of **public procurement** is probably one of the areas of public sector's activity in which **digitalisation has developed most intensively**, and this is due to a double regulatory imperative, already pointed out:

- a. of the European regulations: Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC —known as the *classic* public procurement Directive—, and
- b. of the Spanish national regulation transposing the abovementioned Directive: Act 9/2017, of 8 November, on Public Sector Contracts (LCSP).

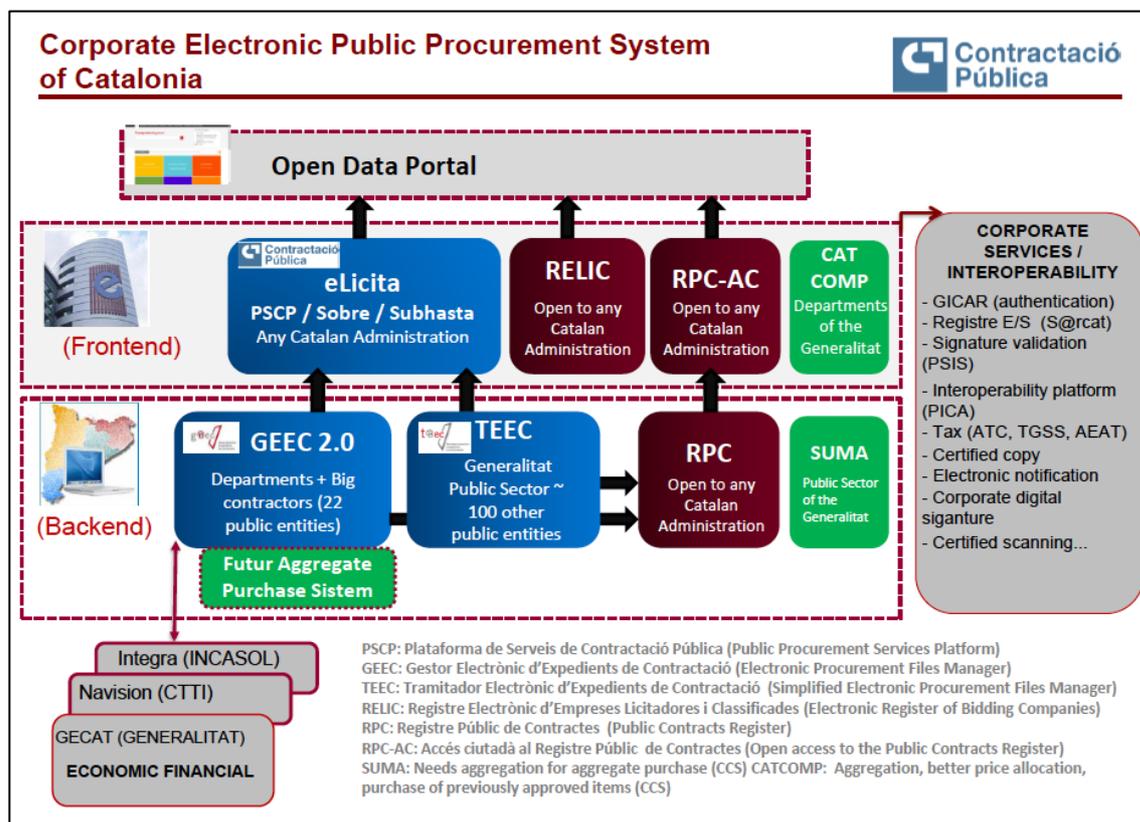
This is the current regulatory framework, but the truth is that the digitalisation of the public procurement activity began much earlier.

The Catalan Government was a pioneer in this area. Thus, even before the classic Directive previously in force —Directive 2004/18/CE— was enacted, in Catalonia the Decree 96/2004, of 20 January regulated the use of electronic, computerised and telematic means in public procurement. This decree constituted the regulatory framework in which the Corporate Electronic Public Procurement System of Catalonia began to be built⁸². This system has been integrating the various elements that make possible, gradually more and

⁸² A description of the current configuration of the System and its individual constituent elements can be found at: <https://contractacio.gencat.cat/ca/contractacio-electronica/> (last consulted: March 2022).

more intensely and profoundly, the complete digitalisation of the procurement activity of the Catalan administrations: the electronic registry of tenderers, the electronic manager of contracting files, the contracting platform, the public registry of contracts, etc.

Figure 1 shows a global vision of this Catalan public procurement ecosystem.



Source: Generalitat de Catalunya

Figure 7: Structure of the Corporate E-Procurement System of Catalonia

As can be seen, the main elements of the system are:

a. for the **handling of the procedures**:

- i. on the frontend: the Public Procurement Services Platform (PSCP, by its Catalan acronym), which hosts the buyer profiles of the Generalitat and its public sector, as well as the majority of local Catalan entities; and
- ii. on the backend: the electronic procurement file managers, GEEC and TEEC (by their Catalan acronyms); and

b. for the **storage of contract data**, the Public Registry of Contracts (RPC), which is, in its current version, fully interoperable, in an automated way, with the GEEC.



It must be noted though that this is the situation on the level of the regional administration, the Generalitat de Catalunya and its public sector. In the local administrations, the digitalisation of public procurement has not yet reached such an advanced level of development. Thus:

- a. regarding the handling of the procedures:
 - i. most of the Catalan local entities, as we have just seen, host their buyer profile in the PSCP⁸³;
 - ii. however, the use of electronic procurement file managers is far from being widespread on the local level⁸⁴; and
- b. regarding the registration of contracting data, the level at which the local authorities inform the RPC—in a non-automated way, due to the lack of electronic procurement file managers interoperable with the RPC, or directly due to the lack of any electronic procurement file manager—is still lower than the level of the Generalitat.

Obviously, the creation, by article 42 LCSP of the buyer profile as a tool for dissemination and publicity on the Internet of public procurement notices, information and data of any contracting public entities, was a milestone of great relevance, at the time.

But if the previous Directive 2004/18/EC decidedly opted for the intensive use of electronic means in procurement procedures, the regulation established in the current Directive 2014/24/EU has represented a true qualitative leap forward in this area, insofar as it established, once a transitional period of 30 months had elapsed, the widespread implementation of fully electronic tender process: the *e-procurement*, that is, the use of electronic means in all phases of the procurement procedure⁸⁵—contract notices, availability of procurement documents, request to participate and bids submission, resolution, etc.—. As indicated in paragraphs (52) and (80) of the preamble to Directive 2014/24/EU, the objectives are to simplify the publicity of tenders and make the procurement procedures more efficient and transparent, facilitating access by the various economic operators and, especially, SMEs.

The evolution in this field from the Directive of 2004 to the current one of 2014 is extremely significant. One of the explicit objectives of the Directive of 2004 was to place the electronic and the traditional means of communication on an equal footing—paragraph (35) of the preamble—, assuming that the latter were the usual means of communication between economic operators and contracting authorities.

⁸³ The PSCP hosts a total of 1,912 buyer profiles, of which 274 are form bodies and entities of the Generalitat and its public sector and 1,565 from local entities, among others—data accessible at: https://contractaciopublica.gencat.cat/ecofin_pscp/AppJava/ca_ES/cap.pscp?reqCode=start (last consulted: March 2022).

⁸⁴ Proof of this is the initiative of the Localret Consortium which, in mid-2021, carried out a preliminary market consultation to prepare the subsequent tender for a contract to provide local authorities in Catalonia with a public procurement monitoring, control and planning tool. As a result of the said preliminary market consultation, the Localret Consortium began, at the end of 2021, the procurement procedure for the conclusion of a framework agreement for the acquisition of a computer solution for the monitoring, control and planning of public procurement.

⁸⁵ In the words of the European Commission presentation of its e-procurement policy, on its official website—accessible at: https://ec.europa.eu/growth/single-market/public-procurement/digital-procurement_es (consulted on 28/03/2022)—: «Public procurement is undergoing a digital transformation. The EU supports the process of **rethinking public procurement process with digital technologies in mind**. This goes beyond simply moving to electronic tools...».

In Directive 2014/24/EU though the use of electronic communication means —defined in its article 2.1.19— becomes the general rule — article 22—.

The imperatives of the European legal framework, its transposition by the Member States and, in Spain, its development by regional Parliaments, have determined the progressive extension in the public sector of the phenomenon known as datafication⁸⁶ of public procurement activity, which could be defined as «the process of capturing all aspects of life and transforming them into data»⁸⁷.

It is well known that the phenomenon of datafication has occurred and spread in most areas of individual and collective life —social, economic and political— in our society: in private life it is a direct consequence of the more and more intensive use of electronic means of communication —paradigmatically smartphones which, beyond their original function as a communication tool, have ended up being one of the main channels of access to social networks and to the vast number of telematic services of the most varied nature: private, economic, administrative and financial, etc.—.

But datafication is also taking place, increasingly rapidly and intensely, in the field of public sector activity, driven by two factors:

- a. On the one hand, directly, the electronic processing of procedures in general and, in particular, of procurement procedures, involves the direct generation of data stored —in a more or less automated way and more or less easily recoverable and, therefore, exploitable— in the electronic file itself and in the electronic documents included in it, as well as, eventually, in the management applications that support and facilitate such electronic processing.
- b. And on the other hand, indirectly, the imperatives derived from transparency, and the progressive implementation and development of the open data paradigm oblige public sector organisations and entities to collect certain data in electronic form —at least that which must be actively publicised— even if the data is not in an electronic file or document.

Both in one case and the other, the end result is that each time the activity of the public sector and, in particular, its contractual activity, generates more data in electronic form which is, therefore, easily collected, stored, processed and exploited for multiple purposes, including obviously the fight against corruption and fraud in public procurement and the promotion of integrity. At the same time, the last decade has seen an exponential increase in the technical capabilities for storing and exploiting this enormous amount of data, at an increasingly reduced cost.

In turn, the datafication of the Catalan public sector's procurement activity has determined that there are different sources of public procurement data, more or less interrelated and interconnected, and with a greater or lesser capacity to provide the necessary data for the effective implementation of the risk indicators that are of interest in this project.

⁸⁶ Cukier and Mayer-Schoenberger are usually considered introducers of the neologism *datafication*, in their article *The Rise of Big Data* published in the May-June 2013 issue of *Foreign Affairs*.

⁸⁷ We quote here the introductory work on data science (O'Neil & Schutt, 2013, 5-6), which include the definition of the concept by Cukier and Mayer-Schoenberger, critically analysing its implications.



In the following sections we will study in detail the following three data sources: the PSCP, the RPC and the Open Data Portal of the Generalitat de Catalunya.

For each of these data sources, we will briefly review the main characteristics and the regulatory framework, the available data and their quality, and its integration and interoperability with other data sources or systems.

Our analysis of the available data of these sources will focus on those data useful and relevant for the purposes of constructing corruption risk indicators, taking as reference the document *Red Flags for Integrity* prepared and published by Open Contracting Partnership (OCP)⁸⁸, which is an essential reference in this field and, specially, its annexes which, in its latest version, updated in May 2021, offers a detailed list of 73 possible alerts (red flags) defined according to the public contracting data standard OCDS designed by OCP.

4.3 Public procurement data sources in Catalonia

As it has just been mentioned, there are three main public procurement data sources in the Catalan public sector:

- a. the PSCP, a data source in real time —since public procurement is reflected in this platform as it is being developed: prior information notices, contract notices, contract award notices, etc.—, of special interest in so far as it provides the most up-to-date data;
- b. the RPC, a source of data that we could qualify as reference, since it constitutes the public record of all contractual activity in the public sector; and
- c. the Open Data Portal of the Generalitat de Catalunya, a data source specifically data reuse-oriented which, therefore, offers maximum ease of access and downloading of data.

A. The Public Procurement Services Platform (PSCP)

a. Background. Regulatory framework

It has already been pointed out that an essential element in the process of datafication of public procurement was the introduction, by the third generation of European Directives on public procurement, of the **buyer profile**, which is nothing other than a virtual space on the Internet —a web page — in which all contracting authorities must publish the data and information related to their contractual activity, thus promoting transparency and public access to said information.

⁸⁸ Both the *Red Flags for Integrity* document and the complementary document *Red Flags to OCDS Mapping* can be found in the official OCP website, accessible at: <https://www.open-contracting.org/es/resources/red-flags-integrity-giving-green-light-open-data-solutions/> (last consulted: March 2022).

More information about the document, its objectives and the experts who participated in its elaboration, among others, can be found in the 30/11/2016 post of the OCP blog, accessible at: <https://www.open-contracting.org/es/2016/11/30/red-flags-integrity-giving-green-light-open-data-solutions/> (last consulted: March 2022).



In the document *Green Paper. Public procurement in the European Union: exploring the way forward*, approved on 11/27/1996 (COM/96/583/FINAL), which at the time constituted the starting point for the preparation and promulgation of the classic Directive 2004/18/EC, the Commission formulated, among other recommendations to achieve an effective public procurement policy at European level, the publication of tender notices and procurement information on the Internet, as well as the implementation and development of fully electronic procurement systems.

This is the origin of the concept of the *buyer profile*, provided for in article 35.1 of the classic Directive 2004/18/EC and which is defined in section 2 of annex VIII, which specifies it is an Internet advertising tool through which the contracting authorities must publish the notices and the documentation of the procurement procedures. When transposing Directive 2004/18/EC, the Spanish Act of 2007 used the expression *contracting profile*, defined in article 42 which stipulates that «the contracting authorities will disseminate, via the Internet, their contracting profile» —section 1— which «may include any data and information referring to the contractual activity of the contracting authority» —section 2—.

Obviously the buyer profile —in the EU terminology— or the contracting profile —in the Spanish national regulation— has remained being provided for in the next generation of procurement Directives: in article 48.1 and annex VIII of the current Directive 2014/24/EU, and in article 63 of the current LCSP, which defines the contracting profile as the «element that groups the information and documents related to [...] [the] contractual activity» of the contracting authority, located on the Internet, «in order to ensure transparency and public access to them».

With the enactment and the entry into force of the transparency laws —Spanish Act 19/2013, of 9 December, and Catalan Act 19/2014, of 29 December—, the contracting profiles have also become the instrument of compliance by the contracting authorities of their obligations of active publicity in the public procurement field. In this respect, article 13.2 of the Catalan Transparency Act specifies that «Information on public procurement must be included in a separate space on the Transparency Portal, configured as an electronic platform for specific advertising in this field».

Focusing on Catalonia, we have already seen how the Generalitat was a pioneer in approving, even a few months ahead of the classic Directive 2004/18/EC, the abovementioned Decree 96/2004, of 20 January, which established the regulatory framework in which the Department of Economics and Finance of the Generalitat, through the Public Procurement DG —the unit with transversal competence in this matter— has been developing and implementing the Catalan public corporate e-procurement strategy. This strategy was structured around five main lines of action: the improvement of the regulatory framework; the removal of obstacles to the effective implementation of e-procurement; the development and implementation of the electronic procurement file manager; the development and implementation of the PSCP; and the development of a strategy for obtaining aggregated information on public procurement in Catalonia.

Once technologically developed, the PSCP —hereinafter also referred as *the platform*— became operational in 2008, and was initially regulated by the Department Decree ECF/313/2008, of 23 June, on approving the application of the PSCP. Article 1.1 of this regulation set the aim of the platform of «offering a unique and complete point of reference to the tenderers and contractors of the administration [...] and to the public managers of the procurement».

However, the platform was initially intended to host only the contracting profiles of the departments of the Generalitat de Catalunya and the entities of its linked or dependent public sector, according to article 2.1 of said Department Decree.

Another interesting aspects of this initial regulation of the PSCP are the following:

- a. pursuant to article 2.2 of said regulation, «in order to ensure the interoperability of the platform, the data model, as well as the exchange of data with other procurement systems, the CODICE content and format specifications are implemented, according to the definition of the [Spanish] Ministry of Economy and Finance»; the same provision establishes that the integration of the platform with the electronic procurement file manager of the Generalitat de Catalunya will use this standard;
- b. article 3 of the Department Decree defines the content of the platform, detailing its different modules, both of the backend and the frontend, and article 4.1 provides the integration of the PSCP with other corporate applications of the Generalitat —identity manager, electronic signature, interoperability platform, etc.—;
- c. finally, article 4.2 stipulates that «the platform also contemplates the possibility of its use by other administrations and public entities in Catalonia».

Since then, the platform has been modified and technically improved: in 2010 a specific application was incorporated providing the electronic auction service, regulated by Department Decree ECF/457/2010, of 29 September; later, in 2012, Department Decree ECO/58/2012, of 22 February, modified again the platform's regulation by integrating the Digital Envelope tool, which allows the electronic presentation of tenders.

The current regulation of the platform is established in Department Decree VEH/172/2017, of 25 July, approving the applications of the PSCP and of the Digital Envelope. This new regulation includes, among other regulatory changes, those derived from the transparency laws of 2013 and 2014, and also incorporates the modifications derived from the new version 2.0 of the Digital Envelope tool.

The regulation in force now clearly establishes that the platform is the «unique and complete point of reference of tenders and their results» not only of the Generalitat de Catalunya administration and its linked or dependent public sector, but also of the entities of the local administration and of the public universities of Catalonia, as well as the entities linked or dependent on them —article 1.1—.

Article 2 of the current regulation, however, establishes a difference:

- a. The contracting profiles of all the entities of the administration of the Generalitat, of the Catalan public universities and of the public sector linked to or dependent on the above must be hosted on the platform according to article 2.1.
- b. With regard to the entities of the local administration, they can opt to host their contracting profile on the platform or to have their own independent profile, although in this case they must



communicate the information and data to the PSCP «by interconnection with electronic devices of aggregation of information», as provided in article 2.2.

In short, the use of the platform to publish the public procurement notices, documentation and data is **mandatory for all the entities of all the levels of the Catalan public sector**, either directly, by hosting their contracting profile in the platform, or indirectly, by communicating the information, documents and data.

Article 3 of the current regulation still defines the content of the platform, and its different modules including new specific mentions to the integration with the Spanish Public Sector Procurement Platform. Indeed, since 11 December 2013, when the Act 20/2013, of 9 December, on the guarantee of market unity, came into force, it is mandatory for any regional or local administration to use the Public Sector Procurement Platform for the management of public contracting⁸⁹. Accordingly, the current regulation of the Catalan platform provides its «syndication with the state Public Sector Procurement Platform» through a «bidirectional integration», allowing both sending the publications made in the PSCP and receiving the publications made in the state Public Sector Procurement Platform, as well as searching in both platforms simultaneously —articles 3 and 4.3—.

The same article 3 of current regulation also provides the possibility of generating XML files with the «most relevant data of the actives publications», allowing its consultation in the Open Data Portal of the Generalitat de Catalunya.

As in the former regulation, arts. 3 and 4.1 of current one provide the integration of the PSCP, through web services with other corporate applications of the Generalitat, and in particular with the electronic procurement file managers, GEEC and TEEC. And arts. 3 and 4.2 provide also the exchange of data and contracting documents and the integration with other applications of the local entities and of the public universities through a specific interface —arts. 3 and 4.2—.

b. Available data. Accessibility. Interoperability

The first thing to note is that the PSCP is primarily focused on tender and award phases of procurement procedure. This constitutes a limitation, since the data that can be found in the platform does not cover the entire procurement cycle; basically this is data related to the preparation, tender and award stages; with regard to the execution stage, only the modifications of the contract are published.

The platform's web address can be accessed at the following link: <https://contractaciopublica.gencat.cat>. The header of the front page is the following:

⁸⁹ The 3rd additional provision of Act 20/2013 have renamed the until then known as State Contracting Platform, that became the Public Sector Procurement Platform, and establishes that any national, regional or local administration or any public entity linked to or dependent on the above shall publish the contracting notices and the outcomes of public procurement procedures in that Platform, either directly by hosting there their contracting profile, or by interconnection with electronic devices for aggregating the information of the different administrations and public entities.

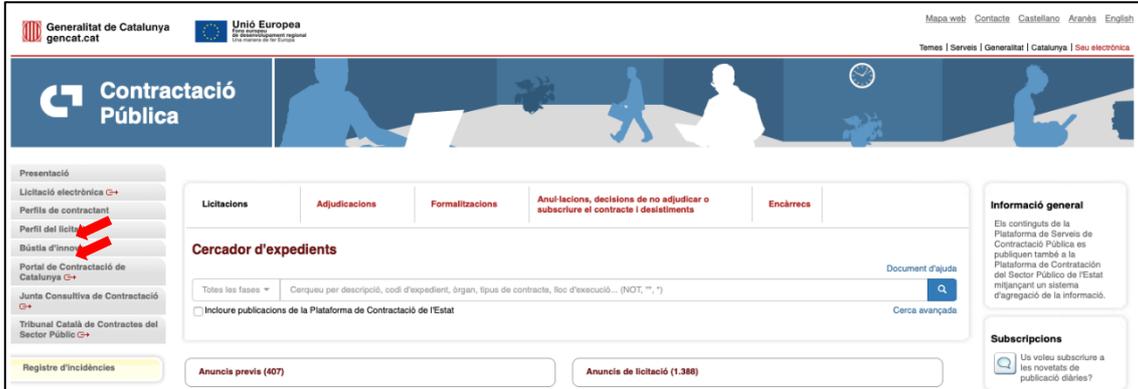


Figure 8: Screen capture of the header of the front page of the PSCP website

The buttons on the left allow the access, among others, to the finders of contracting profiles (*Perfils de contractant* → contracting authorities) and contractor profiles (*Perfil de licitador* → bidders or tenderers) —red arrows in Figure 2 above—.

Once a contracting profile hosted in the platform has been selected, the structure of information within it is the following:

<p>Tender</p> <ul style="list-style-type: none"> • Preliminary market consultations • Prior information notices • Contract notices 	<p>Award / Contract</p> <ul style="list-style-type: none"> • Award notices • Aggregated notices • Contract signature notices
<p>Cancellations, non award decisions and withdrawal</p>	<p>Commissions</p>

This structure appears in the PSCP website as follows:

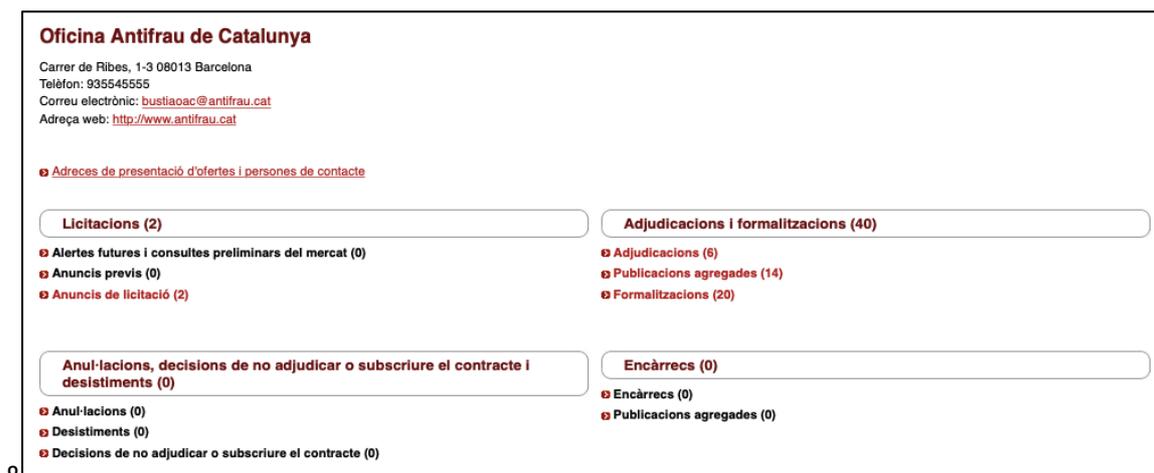


Figure 9: Screen capture of the contracting profile of the Anti-Fraud Office of Catalonia in the PSCP

Considering a specific procurement process the relevant information is shown as follows:



Figure 10: Screen capture of the main information of the current maintenance contract of the premises of the Anti-Fraud Office of Catalonia in the Offices' contracting profile

As can be observed, the top box contains the main data of the contract: contracting authority, procedure reference, procurement procedure and contract type, etc. The bottom bar of this figure shows the different possible stages of the procurement cycle, even though only until the signature stage:

- future alerts and preliminary market consultations (first box of the bar);
- prior information notices (second box);
- contract notice (third box);
- cancellation, non award decision or withdrawal (fourth box);



- e. contract award (fifth box); and
- f. contract signature (sixth box).

Looking for a specific contract, it will be reached through the section of the contracting profile —cf. structure of the contracting profile and Figure 2 ut supra— corresponding to the stage of procurement it is at —tender, award, signature, etc.—

Obviously, all the stages do not exist in all the cases: if the procedure is cancelled or a non-award decision is adopted, there will not be any award nor signature.

As we have already noted, the PSCP is a data source of special interest in so far as it provides the most up-to-date information: any new activity of the procurement procedure is published in the first place in the contracting profile of the contracting authority. This is the greater strength of the PSCP; its most important weakness lies in the availability of the data, since their gathering must be done by scraping techniques, which implies limitations and problems in the accuracy and, generally, the quality of the data thus obtained.

Using scraping techniques, we were able to extract from the PSCP the following data:

- a. 64 data fields from the *Contract notices* section;
- b. 65 data fields from the *Award notices* section; and
- c. 62 data fields from the *Contract signature* section.

The detail of all these data fields is listed in **appendix 1**. As can be observed, there is not a complete coincidence of the data available in each section; nor is there an incremental pattern from one section to the next one, following the natural chronological order —tender (*Contract notices* section) → award (*Award notices* section) → signature (*Contract signature* section)—. In fact, some data of the earlier stages of the procedure —tender and award— no longer appear in the later section —award and signature— while other data, generated in the subsequent stages, do appear. Moreover, certain data fields that appear in all the sections could be empty in some of them, if the respective platform’s user has not filled them from the beginning.

This creates, obviously, additional difficulties in the exploitation of the PSCP data.

However, some of these difficulties can be overcome by obtaining part of the PSCP data by an alternative mean.

Let us remember, that the PSCP is syndicated with the state Public Sector Procurement Platform, which use is mandatory for any regional or local administration since 2013, and which web address is: <https://contrataciondelestado.es/wps/portal/plataforma>. The header of its front page is the following:



Figure 11: Screen capture of the header of the front page of the state Public Sector Procurement Platform website

As can be observed, the buttons and the structure of information are quite similar to the Catalan PSCP's, with finders of contracting profiles (*Perfil contratante* → contracting authorities) and of contractor profiles (*Empresas* → bidders or tenderers). However, there is here a very interesting feature that did not exist directly in the PSCP: the open data tab, accessible from the button *Datos abiertos* —red arrow in Figure above—.

From this tab, different sets of open data could be downloaded, with data coming both from the contracting profiles hosted on the State's platform itself and from the contracting profiles hosted in regional platforms —like the PSCP—, received in the State's platform through the corresponding aggregation mechanisms. These data sets are provided using open and machine-readable formats.

It should be noted, however, that the same possibility also exists in the PSCP, indirectly, through the Open Data Portal of the Catalan public sector: in this regard, it shall be reminded that article 3 of current regulation of the PSCP specifically provides the possibility of generating XML files with the «most relevant data of the actives publications», that could be consulted in the Open Data Portal of the Generalitat de Catalunya.

The point is what does the concept of «most relevant data» means, in one and another platform. We will analyse deeply the Catalan Open Data Portal in a later section of this paper, but in the case of the State's platform, the 25 data fields that could be obtained from the open data tab are the following:

Colum name / data field	Description
Id	Contract unique id. (given by the state platform)
Link	Link to the whole contract data in the PSCP
Summary	Summary of the main contract elements
Title	Title of the contract
Updated	Date of the last available data
Contract_folder_id	Contract ref. (from the PSCP)
Contract_folder_status_code	Procurement stage code (tender, award, signature)
Contracting_party_name	Contracting authority identification (name)
Contracting_party_father_name	Contracting authority type
Estimated_value	Estimated value of the contract
Tax_excluded_amount	Tender amount (tax excluded)

Period_unit	Contract duration
Period_value	
CPV_code	Classification of the subject of the contract, through CPV code ⁹⁰ .
Tender_quantity	Number of tenders received
Tender_result_code	Result of tender process (code)
Winning_party_nif	Contractor id. number (tax identification number: NIF, by its Spanish acronym)
Winning_party_name	Contractor name
Tax_excluded_amount_tendered_project	Awarded contract amount (tax excluded)
Num_lot	Number of the lot
Tender_deadline_date	Tender deadline
Tender_deadline_time	
Legal_document_url	URL of legal documentation (usually, terms of reference)
Technical_document_url	URL of technical documentation

This is scarce data, which only describe the main elements of the contract. The remaining data and information are accessible through the reference to the PSCP —URLs of the whole contract information and, specifically, to the legal and technical documentation—.

The most interesting aspect of these two views of the contracts data, the state platform one and the PSCP one, is that we can obtain in the state platform, in a very easy manner —open data in machine-readable formats— an extremely important datum: the contractor id. number —tax id. number (NIF, by its Spanish acronym)— which is, ultimately, the unique and unambiguous identifier of any contractor. In the PSCP the only available datum about the contractor is its name, which could be ambiguous as far as written in several different ways —exact full name, acronym, commercial name, etc.—. That would be especially problematic in the case of companies belonging to the same group with very similar names.

To summarize, the following main aspects of the PSCP can be highlighted:

- a. The platform is the unique and complete point of reference where all the entities of all the levels of the Catalan public sector have to publish their procurement notices, documentation, data and information.
- b. The main strength of the PSCP is that it provides the most up-to-date information.
- c. Its main limitation is that data are not available in open, machine-readable and reusable formats. The collection of data must be carried out using scraping techniques.
- d. The PSCP is interoperable with other corporate applications of the Generalitat de Catalunya — especially the electronic procurement file managers, GEEC and TEEC—, as well as with

⁹⁰ The common procurement vocabulary (CPV) establishes a single classification system for public procurement at EU level aimed at standardising the references used by contracting authorities and entities to describe procurement contracts. Information on CPV —including the lists of CPV codes— is available at: https://ec.europa.eu/growth/single-market/public-procurement/digital-procurement/common-procurement-vocabulary_es (last consulted: April 2022).

applications from other administrations —universities and local administrations— through specific interfaces.

- e. The PSCP is interoperable and syndicated with the Spanish state Public Sector Procurement Platform, from which part of the data —including some datum no published in the PSCP— can be directly downloaded.

c. Data quality analysis (1): methodology

With regard to data quality, the ISO/IEC 25012⁹¹ —part of a series of International Standards under the general title of Software product Quality Requirements and Evaluation (SQuaRE)— is the absolute reference.

However, there has been many attempts to adapt this generic standard to the data of open data portals. One of the most interesting documents in this field is (Vetrò et al., 2016), which is based on a previous work (Moraga, Moraga, Calero, & Caro, 2009).

For the purposes of our analysis, we have applied a methodology strongly inspired on the ISO 25012 standard, adapted to the limitations of the data under consideration and the time and means inherent to this study. The applied methodology is the following:

- a. We have concentrated our data quality analysis on the most relevant data fields for the purpose of constructing corruption risk indicators and red flags; accordingly, the analysis is limited to 27 of the 62 data fields of the *Contract signature* section of the PSCP.
- b. We have used the three following metrics in our analysis:
 - i. **completeness**, defined for each data field as the ratio of non-null data field (value between 0 and 1);
 - ii. **validity**, defined for some of the analysed data fields according to certain rules exclusively referred to the considered data field —e.g. names of more than one letter; dates within a reasonable range; etc.— (value between 0 and 1); and
 - iii. **consistency**, defined for some of the analysed data fields according to certain rules referred to the relations between different data fields —e.g. date of award necessarily following the date of contract notice and date of signature necessarily following the award; award amount less than tender amount; etc.— (value between 0 and 1).

It should be noted that the validity and consistency metrics are not applicable to all the analysed data fields: in certain cases there is not any applicable rule to determine the value of these metrics. In the following table summarizing the results of our analysis, these cases are denoted with the value -1.

⁹¹ ISO/IEC 25012. *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model*, ISO/IEC, International Standard, 2008. A very detailed explanation on how the ISO 25012 methodology is applied in different specific cases could be found in (Gualo, Rodríguez, Verdugo, Caballero, & Piattini, 2021).

c. When possible, for some of the analysed data fields, we have defined an **overall score**, according to the following rules:

- i. the overall score is calculated only if the value of completeness is greater than 0,50; otherwise it means that there is more than 50% of missing data in this field and the score would not be significant; these cases are denoted in the following table with the value 0 in the overall score column;
- ii. the overall score has also a value between 0 and 1;
- iii. if neither validity nor consistency are applicable, the overall score is the value of completeness;
- iv. if only validity or only consistency are applicable, the overall score is the result of the applicable of these metrics by completeness; and
- v. if both validity and consistency are applicable, the overall score is given by the formula: $\text{completeness} \times (\text{validity} + \text{consistency})/2$.

d. Data quality analysis (2): analysis and findings

Applying this methodology to the selected data fields, we have obtained the following results:

Data field	completeness	validity	consistency	overall score
CPV code	0,042474	-1	-1	0
Procurement file code	1	0,999877	-1	0,999877
Harmonized ⁹² contract	0,736793	1	-1	0,736793
Name	1	0,998595	-1	0,998595
Date of award	0,999886	0,999962	0,991391	0,995563
Date of signature	0,999953	0,999972	0,991391	0,995634
Date of contract notice	1	1	0,998500	0,999250
Description of contract subject	0,997912	-1	-1	0,997912
Description of lot	0,374438	-1	-1	0

⁹² According to article 14 of LCSP, harmonized contracts are those which estimated value is equal or greater than the thresholds of article 4 of the current Directive 2014/24/EU.



Contract duration	0,901915	-1	-1	0,901915
Modifications provided in procurement documents	0,739991	1	-1	0,739991
Tenderers id.	0,245363	0,987389	-1	0
Contract amount	0,946753	0,999409	0,906509	0,902217
Technical reports	0,009624	-1	-1	0
Nationality	1	-1	-1	1
Number of tenders received	0,997770	1	-1	0,997770
Tenders opening	0,474648	-1	-1	0
Contracting authority	1	0,999934	-1	0,999934
Tender budget	0,981530	0,994943	0,906509	0,933166
Procurement procedure	0,994989	0,998579	-1	0,993574
Contract extension provided	0,432839	1	-1	0
Tenders comparative table	0,009700	-1	-1	0
Submission deadline	0,863048	-1	-1	0,863048
Implementation period	0,096784	-1	-1	0
Contract type	1	0,997656	-1	0,997656
Procedure type	1	1	-1	1
Contract estimated value	0,989977	0,988159	-1	0,978255

Generally speaking, the quality of part of these relevant fields is acceptable, although some very significant weaknesses are noted:

- a. Regarding the contract duration field, the completeness is about 0,90, which means there is 10% of missing data in this very relevant field.
- b. Regarding the modifications field —another very important field, since the contract modifications are a well-known risk area—, the completeness is of 0,74, which means there is 26% of missing data.
- c. With regard to the contract extension field —another important field, closely related to the previous one—, the completeness is under 0,50.

- c. With regard to the contract amount field, the overall score is 0,90; however, since the completeness is 0,95, that indicates the existence of some consistency flaws.
- d. And regarding the submission deadline field, the completeness is of 0,86, which means there is 14% of missing data.

The tenderer identity field is a very important field with a very low level of completeness —far under 0,50—, but this issue will be discussed further. At this point, let us just note that the web configuration of the platform makes extremely difficult to obtain the identity of the tenders by scraping, which is, as has been seen, the main mean to collect data from the PSCP.

As regards the state platform, the completeness analysis of the 25 data fields that could be obtained from the open data tab, give the following generally acceptable values:

Data field	completeness
Id	1
Link	1
Summary	1
Title	1
Updated	1
Contract_folder_id	1
Contract_folder_status_code	0,999479
Contracting_party_name	1
Contracting_party_father_name	1
Estimated_value	0,990069
Tax_excluded_amount	1
Period_unit	1
Period_value	0,999014
CPV_code	0,999014
Tender_quantity	0,990494
Tender_result_code	1
Winning_party_nif	1
Winning_party_name	0,959398
Tax_excluded_amount_tendered_project	0,959398
Num_lot	0,567247
Tender_deadline_date	0,999989
Tender_deadline_time	0,870065
Legal_document_url	0,870065
Technical_document_url	0,383462

Finally, to conclude this section of our study, focused on the PSCP, two general flaws should be noted:

- a. The lack in the platform of a unique contract identifier —the reference of each contract is given by each platform’s user, according to its own numbering and classification—, which is an important limitation from the traceability point of view.



- b. And the lack in the PSCP available data, of a key data: the contractor id. number, unique and unambiguous identifier of any contractor, although this flaw may be partially overcome by obtaining this datum from the state platform, as already mentioned.

B. The Public Registry of Contracts (RPC)

a. Background. Regulatory framework

The RPC is the oldest of the 3 data sources we are examining. It was initially created in 1986, by Decree 214/1986, of 26 June, as the Registry of Contracts of the Generalitat, conceived as an internal control and monitoring tool of the contracts of the Generalitat de Catalunya. Ten years later Decree 376/1996, of 2 December, renamed it as the Public Registry of Contracts. The new RPC became fully operational in 1997, when Department Decree of 26/02/1997 approved the application for the communication of contract data by contracting authorities.

Initially, only the contracts awarded by the Administration of the Generalitat and its linked or dependent public sector were entered into the Registry. But in the same way already seen with respect to the PSCP, a progressive extension of the use of the RPC by all the different Catalan administrations has occurred.

Thus, according to the public procurement legal framework of 2011, the new regulation of the Registry, by Department Decree ECO/47/2013, of 15 March, extended to the public Catalan universities the obligation of registration of their contracts. For the entities of the local administration the registration of their contracts remained optional until 2015, when new amendments of the legal framework —among others, by the Catalan Transparency Act of 2014— made the registration of their contracts also mandatory for the local administrations and their linked or dependent public sector.

The current regulation of the RPC is the Department Decree ECO/47/2013, of 15 March, on the regulation of the Public Registry of Contracts and approving its application, modified by Department Decree ECO/294/2015, of 18 September.

Article 1 of said regulation provides:

- a. the mandatory communication to the Registry, by the regional administration of the Generalitat, the Catalan public universities and the entities of their linked or dependent public sector, of «the basic data of all the contracts awarded [...], as well as, if applicable, their modifications, extensions, variations in terms or prices, final amount and termination» —article 1.1—; and
- b. the mandatory communication to the Registry, by the entities and organisms of the local administration, of «the data and documents of the public contracts that they award, including minor contracts» —article 1.2—.



The communication of their contracts is then currently optional only for the statutory bodies of the parliamentary sphere and similar entities according to article 1.3.

Article 3 of the current regulation provides the communication of the contract data must be done within one month from the signature of the contract of its modification, or from the decision approving the extension, termination, settlement or any other incidence of the contract.

In 2013, a specialized ‘Studies and opinions’ section was created within the RPC, in the wake of the 2009 scandal over the huge volume of spending of the Generalitat on such contracts in the crisis context of these years. This specialized section is provided in article 4 which does not apply to the local administration.

b. Available data. Accessibility. Interoperability

According to article 1.4 of the Registry regulation, covered entities must communicate to the Registry, through its computer application, all the data contained in the annex of the regulation, if applicable. The detail of all these data fields is listed in **appendix 2.A**, under the following headings:

- 1.1 A. Data of contract award.
- 1.1 B. Contract documents
- 1.2 Data of the lot
 - 2.1 Data of contract modifications
 - 2.2 Data of contract transmission
 - 2.3 Data of subcontracting
- 3. Data of contract extension
- 4. Data of contract termination

However, this provision does not mean all these data are available to the public. Depending on the type of access to the Registry, more or less data are accessible. In this respect, we can distinguish the following three possible access ways.

b.1. Public open access. One of the 2015 amendments of Department Decree ECO/47/2013 introduced a new chapter II in the Registry regulation entitled «Public citizen access to the Public Registry of Contracts of the Generalitat de Catalunya» —arts. 10 to 13—.



This open access operates via a search engine with two very important limitations:

- a. on the one hand, it is not possible through this way to download all the Registry available data, only a very small part of them;
- b. on the other hand —and this is a major limitation— only 8 data fields are available as any search result:
 - i. the contracting authority;
 - ii. the file or contract reference;
 - iii. the award date;
 - iv. the type of procurement procedure;
 - v. the type of contract;
 - vi. the contract subject description —that does not include the CPV code—;
 - vii. the contract amount; and
 - viii. the contractor name

A little more information, but in a very limited manner, could be obtained: the search engine allows to search by each one of abovementioned data fields and also by CPV code, contract extensions and cancellations, and contractor tax id. number.

This access is regulated by article 12 of the Registry regulation, that specifies that the results of any search will only contain data of the last 5 years contracts, which is an additional limitation.

b.2. Restricted access. The second way of access to the Registry data is far broader and powerful. Through a specific agreement, a restricted access is available, which requires the identification of any user and allows it to make much more comprehensive searches.

Again, this access operates via a search engine but the result of any search includes 154 data fields. The detail of all these data fields is listed in **appendix 2.B**.

The main limitation of this access is that it still does not allow downloading all the Registry available data since the application collapses if a query result exceed of about 5.000 lines (contracts).

Furthermore, another important limitation is that a very important data, identity of the tenders, neither is available by this access.

b.3. Full access. Finally, the complete and direct access to the RPC management application allows handling up 226 data fields, or even more, without, obviously, any of the abovementioned limitations.

Although identity of the tenders is available, obviously when these data have been communicated by contracting authorities.

Finally, with regard to the interoperability of the Registry, it should be noted that article 7 of the Registry regulation provides its «integration with the electronic procurement file managers for the contract data transfer, both of the contract award, the contract status during the implementation and the final settlement, and the contract documents, if applicable».

In our opinion, there is here a relevant flaw, which is the absence of a direct integration and interoperability of the RPC with the PSCP. As a matter of fact, and within the sphere of the regional administration of the Generalitat—in the local administration the situation is even worse—, both the RPC and the PSCP interact with the electronic procurement file managers, GEEC and TEEC, but not directly between them. Considering that the PSCP is the point of reference for the publication of procurement notices, data and information, widely used by almost all the entities of the Catalan public sector, a direct and automatized interaction with the RPC would be suitable for the sake of completeness of the Registry.

c. Data quality analysis and findings

We have applied the data quality analysis methodology described above to a selection of relevant data fields of the RPC, and the following results were obtained:

Data field	completeness	validity	consistency	overall score
Year of contract implementation	1	1	0,999883	0,999941
Date of award	1	0,999777	0,999883	0,999830
Contracting authority	1	0,982751	-1	0,982751
Contractor tax id. number	0,999985	0,998557	0,908199	0,953363
Contractor name	0,999985	0,994346	0,988118	0,991217
Type of contracting authority	1	1	0,999318	0,999659
Unique contract id. reference	1	-1	1	1
Contracting authority code	1	-1	1	1
Contract register number	1	-1	-1	1



Contract group code	1	1	-1	1
Contract code	1	1	-1	1
Multi-year contract	1	1	-1	1
Description of contract subject	0,999747	-1	-1	0,999747
Procurement procedure code	0,994571	1	1	0,994571
Procurement procedure	0,994571	0,996463	1	0,992812
Legal ground of non competitive procedure	0,034923	-1	-1	0
Number of invited bidders / tenderers	0,062078	0,998024	-1	0
Procurement file code	0,999467	0,998742	1	0,998838
Procedure type	0,551177	1	-1	0,551177
Contract amount	0,999345	0,999976	-1	0,999321
Number of lots	1	1	0,992645	0,996322
Stage of lot procurement	1	1	-1	1
Stage of contract procurement	1	1	-1	1
Date of RPC internal review of data	0,999621	1	0,998083	0,998663
Date of communication of contract data	1	1	-1	1
Contract based on framework agreement	0,016409	1	-1	0
Centralised purchasing technique code	0,015750	-1	-1	0
Awarding criteria code	0,077818	-1	-1	0
Non-awarded lot	1	1	0	0,500000
CPV code	0,503165	0,999876	-1	0,503103
Number of the lot	1	0,999994	0,999776	0,999885
Temporary grouping	0,999985	1	-1	0,999985



Country	0,985866	-1	0,996521	0,982437
Region (province)	0,927750	0,979393	0,937956	0,889410
Municipality	0,927572	0,378714	0,672784	0,487670
Contracting authority type	0,998897	1	-1	0,998897
Documentation included	1	1	-1	1

As can be seen, the quality of the most relevant of the analysed fields is acceptable, and globally greater than the data quality of the PSCP analysed data fields —which is logical given the legal nature of the RPC as public sector contacts data repository—. However, some weaknesses should be noted:

- a. Regarding the Contractor tax id. number field —a key data, as we have seen, since it is its unique and unambiguous identifier—, the overall score is 0,95; however, since the completeness almost 1, that indicates the existence of some consistency issue.
- b. Regarding the Number of invitations field —another important field, since related to the existence and extent of competition in the procurement process—, the completeness is of 0,06, which means there is 94% of missing data.
- c. And with regard to the CPV code field, the completeness is just up to 0,50.

Worthy of special mention is the data on the identity of the tenders as a key element to analyse different risk factors —collusion between tenders, existence of actual competition, etc.—. We have already seen that this data field have a very low level of completeness in the PSCP and is very difficult to obtain by scraping; and we have noted that this data was available only through full access to the RPC.

A specific analysis of this data was conducted with the following results:

First of all, the overall level of completeness has been growing since 2014, but is still under 0,50, as shown in the following figure:

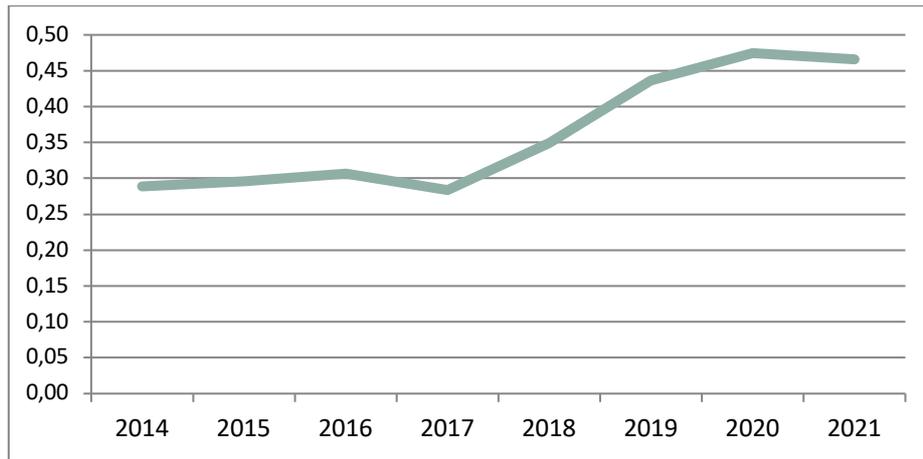


Figure 12: Tenderers identity completeness evolution

If we differentiate now between the different types of contracts —work, service, supply and others— the following results were obtained:

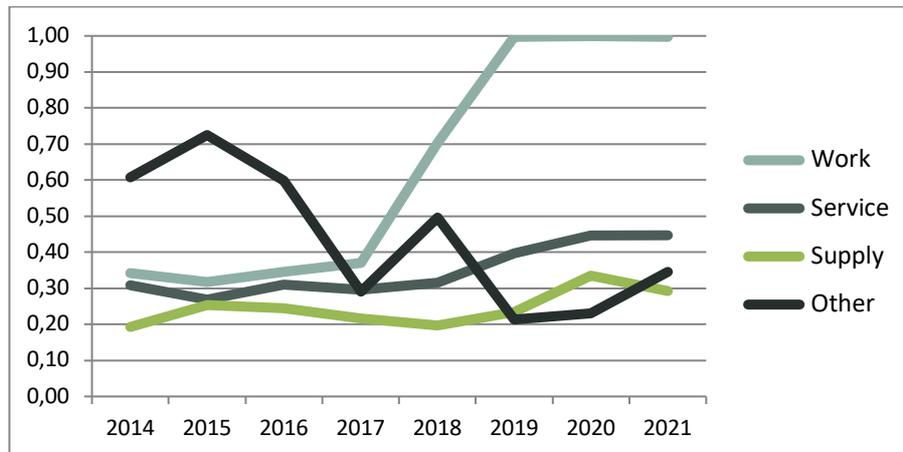


Figure 13: Tenderers identity completeness evolution by types of contracts

As can be seen, the identity of the tenders became a mandatory data to be communicated to the Registry since 2019, while it has been optional for the other types of contracts. Taking into account the importance of this data, this distinction does not seem reasonable.

C. The Open Data Portal of the Generalitat de Catalunya

a. Background. Regulatory framework



The commitment of Catalan Government with open government and, as an integral part of this wider concept, with open data, is not new⁹³.

One of the first achievements in this field in Catalonia was the adoption by the Government of the Generalitat of the Decision of 09/11/2010, launching the Project for opening the Government of Catalonia's public data. The main aim of the Agreement was to improve public information, promote its reuse, ensure access to it for everyone, and create the Open Data Portal.

The first version of the Transparency Portal of the Generalitat was launched three years later, providing access to all relevant public information.

Obviously, the enactment of Transparency laws —Spanish Act 19/2013, of 9 December, and Catalan Act 19/2014, of 29 December— had a key influence on the open government and open data policies, insofar as this legal framework determines the information that the Catalan authorities and the parties receiving public funds must actively publish, and, at the same time, it establishes the right of access by citizen to public data and information.

The current regulatory and operational framework is the Government Decision GOV/154/2018, of 20 December, on the Open Data Strategy of the Generalitat de Catalunya and on the adherence to the International Open Data Charter Principles.

The most interesting aspects of this decision and, specially, of its annex 1, which contain the Open Data Strategy of the Generalitat, are the following:

- a. According to section 3 of the Strategy: «The administration of the Generalitat de Catalunya and its public sector promotes the openness by default of all its data, both those produced internally and those it possesses as a consequence of its activity or the exercise of its functions, included those supplied by third parties according to Act 19/2014, of 29 December, on transparency»; the only limitations and exceptions to this openness are those established in said Transparency Act.
- b. Section 4 of the Strategy provides the Generalitat «will promote and facilitate the integration of data originated by municipalities and the rest of local entities».
- c. Section 5 establish the access to data is free of charge.
- d. Section 7 contains the commitment of the Generalitat with the open data principles, by proclaiming that «the Government [of the Generalitat] subscribes the principles of the International Open Data Charter», which are listed below:
 1. Open by Default
 2. Timely and Comprehensive

⁹³ A lot of information about Catalan open government and open data strategies and policies can be found at: <https://governobert.gencat.cat/en/que-es/Que-es-el-govern-obert/> (consulted on 29/04/2022). Interesting historical background can be found in the first section of the Open Government Plan 2017-2018. Accessible at: <https://governobert.gencat.cat/en/details/article/Pla-de-Govern-Obert-00002> (last consulted: March 2022).

3. Accessible and Usable
 4. Comparable and Interoperable
 5. For Improved Governance and Citizen Engagement
 6. For Inclusive Development and Innovation
- e. Section 14 contains an explicit reference to public procurement data by providing that «The data obtained through the management of public contracts awarded by the administration of the Generalitat [...] and by its public sector will be processed in accordance with the open data strategy of the Generalitat de Catalunya».
 - f. Finally, sections 17 to 24 refer to what could be called the *data ecosystem* and its main elements: the Open Data Platform of Catalonia —section 17—, the Open Data Portal of the Generalitat —section 18—, the catalogue of open data and the inventory of data —sections 19 and 24—, the service quality levels —section 20—, the data standards and the data use licences —sections 21 and 22—, and the integration with the Transparency Portal of Catalonia —section 23—.

b. Available datasets. Interoperability. Accessibility and reusability

The Open Data Portal contains 1077 datasets⁹⁴ and offers different ways to search for specific data: a finder to search by keywords and filters by categories, types of data —archives and documents, calendars, datasets, maps, etc.—, geographic information, or tags.

A query with the words «contractació pública» —public procurement in Catalan⁹⁵— yields 14 results, the most relevant of which are the following:

- a. Dataset **Public Procurement in Catalonia: ongoing tenders and awards**. This dataset includes tenders and awarded contracts of the whole public sector of Catalonia (regional administration of the Generalitat, public universities, local entities and other public bodies, and their linked or dependent public sectors). There are 43 data fields (columns) and about 185 thousand rows, each row being a contract notice. The existing data fields are the following:

Colum name / data field	Description
CODI_AMBIT	Code of administrative level
NOM_AMBIT	Administrative level (Departments and public sector of the Generalitat de Catalunya / entities of the local administration / universities / statutory bodies of the parliamentary sphere and similar entities / other entities)
CODI_DEPARTAMENT/ENS	Department code
NOM_DEPARTAMENT/ENS	Department description

⁹⁴ Accessible at: <https://analisi.transparenciacatalunya.cat/> (last consulted: March 2022). All the following references should be understood as results of the query on that date.

⁹⁵ The web has an English version, and the finder also works with the English words «public procurement», but the results thus obtained are less: only 8 instead of 14.

CODI_ORGAN	Contracting authority id.
NOM_ORGAN	Contracting authority description
CODI_UNITAT	Contracting unit id.
NOM_UNITAT	Contracting unit description
CODI_INE10	INE10 code (INE is the Spanish National Statistics Institute)
CODI_DIR3	Public administrations units and registries catalogue id.
CODI_EXPEDIENT	Procurement file code
TIPUS_CONTRACTE	Type of contract (work, service, supply,...)
SUBTIPUS_CONTRACTE	Subtype of contract
PROCEDIMENT	Procurement procedure
FASE_PUBLICACIO	Last published procurement procedure stage
DENOMINACIO	Contract name
OBJECTE_CONTRACTE	Contract description
PRESSUPOST_LICITACIO	Tender amount
VALOR_ESTIMAT_CONTRACTE	Estimated value of the contract
CODI_NUTS	NUTS statistical code
LLOC_EXECUCIO	Place of implementation
DURACIO_CONTRACTE	Duration of the contract (years, months, days or date range)
TERMINI_PRESENTACIO_OFERTES	Submission deadline
DATA_PUBLICACIO_FUTURA	Date of future alert or preliminary market consultation notice
DATA_PUBLICACIO_PREVI	Date of prior information notice
DATA_PUBLICACIO_ANUNCI	Date of contract notice
DATA_PUBLICACIO_ADJUDICACIO	Date of award notice
DATA_PUBLICACIO_FORMALITZACIO	Date of contract signature notice
DATA_PUBLICACIO_ANUL	Date cancellation notice
NUMERO_LOT	Lot number
DESCRIPCIO_LOT	Lot description
CODI_CPV	CPV code
TIPUS_IDENTIFICACIO_ADJUDICATARI	Type of contractor id. (NIF, UTE, DUNS, VIES, others)
IDENTIFICACIO_ADJUDICATARI	Contractor id.
DENOMINACIO_ADJUDICATARI	Contractor name
IMPORT_ADJUDICACIO_SENSE_IVA	Award amount (tax excluded)
IMPORT_ADJUDICACIO_AMB_IVA	Award amount (tax included)
OFERTES_REBUDES	Number of tenders received
RESULTAT	Result of tender process for the lot (unsuccessful, cancelled, non-award decision, withdrawal, award, contract signature)
ENLLAC_PUBLICACIO	Link to the whole contract data in the PSCP
ES_AGREGADA	Indicator of aggregated publication (Yes/No)
DATA_ADJUDICACIO_CONTRACTE	Date of award
DATA_FORMALITZACIO_CONTRACTE	Data of signature

- b. Dataset **Emergency Procurement during the COVID-19 pandemic**. This dataset includes contracts of the Generalitat and its public sector and a few of public universities. There are 15 data fields (columns) and 5.567 rows, each row being a contract. The existing data fields are the following:

Colum name / data field	Description
Àmbit subjectiu	Administrative level
Òrgan de contractació	Contracting authority
Codi d'expedient	Procurement file code
Objecte del contracte	Contract subject
Descripció de la prestació	Description of product or service to be delivered
Tipus d'expedient	Type of procedure
Tipus de contracte	Type of contract
Pressupost de licitació (sense IVA)	Tender amount
Valor estimat del contracte	Estimated value of the contract
Data de publicació	Date of notice
Empresa adjudicatària	Contractor name
Import d'adjudicació (sense IVA)	Award amount (tax excluded)
Enllaç a la publicació	Link to the contract award notice in the PSCP
Codi CPV	CPV code
Data d'adjudicació	Date of award

- c. Dataset **Contracts of the Generalitat de Catalunya: small amount contracts**⁹⁶. This dataset includes data of the small amount contracts of the Generalitat and its public sector of the last 5 years. There are 10 data fields (columns) and about 661 thousand rows, each row being a contract. The existing data fields are the following:

Colum name / data field	Description
Any	Year
Codi d'expedient	Procurement file code
Departament d'adscripció	Contracting department
Òrgan de contractació	Contracting authority
Àmbit subjectiu	Administrative level
Tipus de contracte	Type of contract
Procediment d'adjudicació	Procurement procedure
Objecte del contracte	Contract subject
Import adjudicate sense IVA	Award amount (tax excluded)
Empresa adjudicatària	Contractor
Tipus d'expedient	Type of procedure

- d. Dataset **Public Procurement of Catalonia**. This dataset is the most comprehensive one, with data of the public procurement of the whole public sector of Catalonia (regional administration of the Generalitat, public universities, local entities and other public bodies, and their linked or dependent public sectors) for the period 1998-2022 —although the relevant data are as of 2017—. There are

⁹⁶ According to article 118.1 of LCSP, small amount or *minor* contracts are those with an estimated value below 40.000 €, in the case of work contracts, or 15.000 €, in the case of service or supply contracts. This kind of contracts could be directly awarded without any competitive procedure.

24 data fields (columns) and about 1,72 million rows, each row being a contract. The existing data fields are the following:

Column name / data field	Description
SITUACIÓ CONTRACTUAL	Contract status (award / signature / administrative contract / small amount contract / contract extension)
EXERCICI	Budgetary year
SUBJECTE/ÀMBIT	Administrative level (Departments and public sector of the Generalitat de Catalunya / entities of the local administration / universities / statutory bodies of the parliamentary sphere and similar entities / other entities)
ID AGRUPACIÓ ORGANISME	Department code
AGRUPACIÓ ORGANISME	Department description
ID ORGANISME CONTRACTANT	Contracting authority id.
ORGANISME CONTRACTANT	Contracting authority description
CODI EXPEDIENT	Procurement file code
PROCEDIMENT ADJUDICACIÓ	Procurement procedure
TIPUS CONTRACTE	Type of contract (work, service, supply,...)
DESCRIPCIÓ EXPEDIENT	Contract description
NUMERO LOT	Lot number
CODI CPV	CPV code
ADJUDICATARI	Contractor name
IMPORT ADJUDICACIÓ	Award amount or extension amount
DATA ADJUDICACIÓ	Date of award
CONTRACTE	Lot description
LOT DESERT	Unsuccessful lot tender indicator (Yes/No)
DIES DURADA	Duration of the contract (days)
MESOS DURADA	Duration of the contract (months)
ANYS DURADA	Duration of the contract (years)
NUMERO PRORROGA	Number of extension, if applicable
DATA INICI PRORROGA	Date of extension starting, if applicable
DATA FI PRORROGA	Data of extension ending, if applicable

- e. Dataset **Programmed Procurement of the Generalitat de Catalunya and its public sector**. This dataset includes the planned annual procurement, excluding small amount contracts, of the administration of the Generalitat and its public sector, with 11 data fields (columns) and about 17,5 thousand rows, each row being a planned procurement procedure.
- f. Dataset **Contracts awards of the Generalitat de Catalunya**. This group of datasets includes the data of all the contracts awards of the departments of the Generalitat and its public sector by fortnight, for the period 2018-2022, in Excel files with 16 data fields (columns).
- g. Decisions of the Catalan Court of Public Sector Contracts.

For each dataset of the Portal complete information is provided. For example, in the case of the first of the abovementioned datasets, the information is the following:

<p>Updated May 10, 2022</p>		<p>Actors</p>							
<p>Data Last Updated May 10, 2022</p>	<p>Metadata Last Updated May 10, 2022</p>	<table border="1"> <tr> <td>Tipus actor</td> <td>Generalitat de Catalunya</td> </tr> <tr> <td>Departament</td> <td>Economia i Hisenda</td> </tr> </table>		Tipus actor	Generalitat de Catalunya	Departament	Economia i Hisenda		
Tipus actor	Generalitat de Catalunya								
Departament	Economia i Hisenda								
<p>Date Created January 10, 2019</p>	<p>Freqüència d'actualització</p> <table border="1"> <tr> <td>Freqüència d'actualització</td> <td>Diària</td> </tr> </table>			Freqüència d'actualització	Diària				
Freqüència d'actualització	Diària								
<p>Views 4,521</p>	<p>Downloads 4,350</p>	<p>Idioma</p> <table border="1"> <tr> <td>Idioma</td> <td>Català</td> </tr> <tr> <td>Nombre del conjunto de datos (ES)</td> <td>Contratación pública en Catalunya: licitaciones y adjudicaciones en curso</td> </tr> <tr> <td>Dataset name (EN)</td> <td>Public procurement in Catalonia: bidding and awards in progress</td> </tr> </table>		Idioma	Català	Nombre del conjunto de datos (ES)	Contratación pública en Catalunya: licitaciones y adjudicaciones en curso	Dataset name (EN)	Public procurement in Catalonia: bidding and awards in progress
Idioma	Català								
Nombre del conjunto de datos (ES)	Contratación pública en Catalunya: licitaciones y adjudicaciones en curso								
Dataset name (EN)	Public procurement in Catalonia: bidding and awards in progress								
<p>Data Provided by Direcció General de Contractació Pública</p>	<p>Dataset Owner Dades Obertes Catalunya</p>	<p>Àmbit geogràfic</p> <table border="1"> <tr> <td>Àmbit geogràfic</td> <td>Catalunya</td> </tr> <tr> <td>Informació geogràfica</td> <td>Sense informació geogràfica</td> </tr> </table>		Àmbit geogràfic	Catalunya	Informació geogràfica	Sense informació geogràfica		
Àmbit geogràfic	Catalunya								
Informació geogràfica	Sense informació geogràfica								
<p>Contact Dataset Owner</p>		<p>Topics</p> <table border="1"> <tr> <td>Category</td> <td>Public Sector</td> </tr> <tr> <td>Tags</td> <td>contractació pública, contractes, empreses, licitacions, adjudicacions</td> </tr> </table>		Category	Public Sector	Tags	contractació pública, contractes, empreses, licitacions, adjudicacions		
Category	Public Sector								
Tags	contractació pública, contractes, empreses, licitacions, adjudicacions								
<p>Licensing and Attribution</p> <table border="1"> <tr> <td>License</td> <td>See Terms of Use</td> </tr> <tr> <td>Source Link</td> <td>http://governobert.gencat.cat/ca/dades_obertes/licencia-oberta-i-nformacio-catalunya/</td> </tr> </table>				License	See Terms of Use	Source Link	http://governobert.gencat.cat/ca/dades_obertes/licencia-oberta-i-nformacio-catalunya/		
License	See Terms of Use								
Source Link	http://governobert.gencat.cat/ca/dades_obertes/licencia-oberta-i-nformacio-catalunya/								

Figure 14: General information about the dataset in the Open Data Portal



What's in this Dataset?

Rows	Columns	Each row is a
185K	43	Anunci

Columns in this Dataset

Column Name	Description	Type
CODI_AMBIT	Identificador de l'àmbit organitzatiu	Number #
NOM_AMBIT	Àmbit organitzatiu (Departaments i sector públic de la gen...	Plain Text T
CODI_DEPARTAMENT/ENS	Identificador del departament o entitat d'adscripció	Number #
NOM_DEPARTAMENT/ENS	Descripció del departament o entitat d'adscripció	Plain Text T
CODI_ORGAN	Identificador de l'organisme contractant	Number #
NOM_ORGAN	Descripció de l'organisme contractant	Plain Text T
CODI_UNITAT	Identificador de la unitat de contractació	Number #

[Show All \(43\)](#)

Figure 15: Information about the content and the structure (data fields and rows) of the dataset in the Open Data Portal

As can be observed in Figure 14., the information about the origin of the data is not very precise: it is stated that data are provided by the Public Procurement DG of the Department of Economics and Finance of the Generalitat, which is, as we have already seen, the unit with transversal competence in this matter responsible of the management and development of the whole Corporate Electronic Public Procurement System of Catalonia. We can thus infer that the data sources are mainly the PSCP and the RPC.

However, details on data transmission from the sources to the Open Data Portal —id est on the extent of the interoperability of Portal with PSCP and RPC— are not available. This is a data traceability flaw of the Portal at more granular level.

Regarding the accessibility and reusability of the data, the Open Data Portal is clearly the best of the examined source: data are downloadable for offline use in CSV, CSV for Excel, TSV for Excel, RDF, XML and RSS formats, and the Socrata Open Data API (SODA) provides programmatic access to all the datasets. There is also a specific tab in the Portal dedicated to data reuse,⁹⁷ with a guide of metadata implementation, information for developers and a list of data reuse cases.

The Open Data Portal includes a link to the Transparency Portal of Catalonia,⁹⁸ where a *Public Procurement* button opens a new web site⁹⁹ with a lot of different data and information on this matter.

⁹⁷ at: https://governobert.gencat.cat/en/dades_obertes/reutilitzacio-dades/index.html (last consulted: May 2022).

⁹⁸ Accessible at: <http://www.transparenciacatalunya.cat/ca/inici>. (last consulted: May 2022).

⁹⁹ Accessible at: <https://governobert.gencat.cat/en/transparencia/Contractacio/index.html> (last consulted: May 2022).

c. Data quality analysis of most relevant datasets

Again, we have applied the data quality analysis methodology described above to the most relevant datasets of the Portal, obtaining the following results:

- a. Regarding the most comprehensive dataset, **Public Procurement of Catalonia**, the quality analysis gives the following generally high values:

Data field	completeness	validity	consistency	overall score
SITUACI_CONTRACTUAL	1	1	-1	1
EXERCICI	1	0,999975	1	0,999988
SUBJECTE_AMBIT	1	1	-1	1
ID_AGRUPACIO_ORGANISME	1	-1	1	1
AGRUPACIO_ORGANISME	1	-1	1	1
ID_ORGANISME_CONTRACTANT	1	-1	0,972714	0,972714
ORGANISME_CONTRACTANT	1	0,988345	1	0,994172
CODI_EXPEDIENT	0,999992	0,999221	-1	0,999213
PROCEDIMENT_ADJUDICACIO	0,999979	0,993491	-1	0,993470
TIPUS_CONTRACTE	1	0,999965	-1	0,999965
DESCRIPCIO_EXPEDIENT	1	-1	-1	1
NUMERO_LOT	1	0,999998	-1	0,999998
CODI_CPV	0,477033	-1	-1	0
ADJUDICATARI	0,999999	0,994936	-1	0,994935
IMPORT_ADJUDICACIO	1	0,999986	-1	0,999986
DATA_ADJUDICACIO	1	0,999762	1	0,999881
CONTRACTE	1	-1	-1	1
LOT_DESERT	1	1	1	1
DIES_DURADA	1	1	-1	1



MESOS_DURADA	1	1	-1	1
ANYS_DURADA	1	1	-1	1
NUMERO_PRORROGA	0,021657	1	-1	0
DATA_INICI_PRORROGA	0,021657	0,999920	0,999920	0
DATA_FI_PRORROGA	0,021657	0,999734	0,999920	0

We can observe a high data quality level, with only one field, the CVP code, with a low value for completeness: 0,48, which means there is 52% of missing data.

- b. Regarding the dataset *Public Procurement in Catalonia: ongoing tenders and awards* the completeness analysis gives the following values:

Data field	completeness
CODI_AMBIT	1
NOM_AMBIT	1
CODI_DEPARTAMENT_ENS	1
NOM_DEPARTAMENT_ENS	1
CODI_ORGAN	1
NOM_ORGAN	1
CODI_INE10	1
CODI_DIR3	0,977221
CODI_EXPEDIENT	0,998320
TIPUS_CONTRACTE	0,940610
PROCEDIMENT	0,974087
FASE_PUBLICACIO	1
DENOMINACIO	1
OBJECTE_CONTRACTE	0,997951
PRESSUPOST_LICITACIO	0,934270
VALOR_ESTIMAT_CONTRACTE	0,934391
CODI_NUTS	0,934887
LLOC_EXECUCIO	0,935405
DURACIO_CONTRACTE	0,937041
TERMINI_PRESENTACIO_OFERTES	0,736049
DATA_PUBLICACIO_ADJUDICACIO	0,904258
CODI_CPV	0,921594
TIPUS_IDENTIFICACIO	0,788641
IDENTIFICACIO_ADJUDICATARI	0,737123
DENOMINACIO_ADJUDICATARI	0,788641
IMPORT_ADJUDICACIO_SENSE	0,788641
IMPORT_ADJUDICACIO_AMB_IVA	0,788641
OFERTES_REBUDES	0,843431
RESULTAT	0,847535
ENLLAC_PUBLICACIO	1
ES_AGREGADA	1
DATA_ADJUDICACIO_CONTRACTE	0,784520
CODI_UNITAT	0,165609



NOM_UNITAT	0,164920
SUBTIPUS_CONTRACTE	0,872732
DATA_PUBLICACIO_ANUNCI	0,759158
DATA_FORMALITZACIO_CONTRACTE	0,589164
NUMERO_LOT	0,315391
DESCRIPCIO_LOT	0,315391
DATA_PUBLICACIO_ANUL	0,028838
DATA_PUBLICACIO_FUTURA	0,024150
DATA_PUBLICACIO_PREVI	0,024145
DATA_PUBLICACIO_FORMALITZACIO	0,005206

We can observe in this dataset some significant flaws:

- i. With regard to the type of contract field, the completeness is 0,94, which means there is 6% of missing data.
 - ii. With regard to the contractor id. field —as noted, an important one— the completeness is 0,74, which means there is 26% of missing data.
 - iii. Regarding the contract amount field, the completeness is 0,79, which means there is 21% of missing data.
 - iv. Regarding the number of tenderers field, the completeness is 0,84, which means there is 16% missing data.
 - v. And regarding the tender submission period field, the completeness is 0,74, which means there is 26% missing data.
- c. Finally, the completeness analysis of the dataset ***Emergency Procurement during the COVID-19 pandemic*** gives the following values:

Data field	completeness
Àmbit subjectiu	1
Òrgan de contractació	1
Codi d'expedient	0,999820
Objecte del contracte	1
Descripció de la prestació	0,999820
Tipus d'expedient	1
Tipus de contracte	1
Pressupost de licitació (sense IVA)	1
Valor estimat del contracte	1
Data de publicació	1
Empresa adjudicatària	1
Import d'adjudicació (sense IVA)	1
Enllaç a la publicació	1
Codi CPV	1
Data d'adjudicació	1

As we can observe, there is a high level of completeness.

The detail of all these data fields is listed in **appendix 1**.

Data fields PSCP

Data RPC **appendix 2.A**, under the following headings:

- 1.1 A. Data of contract award.
- 1.1 B. Contract documents
- 1.2 Data of the lot
- 2.1 Data of contract modifications
- 2.2 Data of contract transmission
- 2.3 Data of subcontracting
- 3. Data of contract extension
- 4. Data of contract termination

Again, this access operates via a search engine, but the result of any search includes 154 data fields. The detail of all these data fields is listed in **appendix 2.B**.

- acceso mediante id. (GICAR): 154 campos:

Codi organisme contractant
Nom organisme contractant
Codi unitat de contractació
Nom unitat de contractació
Número registre
Any execució
Expedient
Tipus contracte
Procediment adjudicació
Descripció del contracte



Descripció del lot
Lot
Lot desert
NIF/CIF contractista
Nom contractista
Import adjudicat lot/pròrroga
Import licitat lot
Import licitat lot amb IVA
Import VEC
Pluriennal
Codi liquidació
Data inici execució
Termini durada anys
Termini durada mesos lot
Termini durada dies lot
Despesa anticipada
Codi CPV
Grup contracte
Import adjudicat lot/pròrroga amb IVA
Data adjudicació
Data aprovació
Data formalització
Anualitats
Tipus IVA
Procedència de producte
Concessió



Data recepció
Data liquidació
Import liquidació compliment
Import liquidació compliment amb Iva
Termini prorrogat
Anys prorrogat
Mes prorrogat
Causa resolució
Data resolució
Revisió de preus
Incautació garantia
Codi CPA
Obra pública
Contracte mixt
Contracte marc
IVA múltiple
Import adjudicació IVA múltiple
Cofinanciació
Import cofinançat
Import cofinançat amb IVA
Import cànon
Import cànon amb IVA
Import subvenció administració
Import subvenció administració amb IVA
Percentatge subvenció administració



Import liquidació resolució
Import liquidació resolució amb Iva
Període cànon
Tipus pagament
Tipus de producte
Estudis i dictàmens
Data tramesa lot
Data alta
Data modificació
Data primera tramitació
Data tramitació contracte
Compte Comptable
Contracte objecte
Estat
Estat de la liquidació
Classificació
Tipus recurs
Sentit de la resolució
País origen
Partida pressupostaria
Numero lot anterior
Motiu utilitat contracte
Lloc execució
Foment social
Garantia complementària
Garantia complementària percentatge



Garantia complementària import
Garantia definitiva
Garantia definitiva percentatge
Garantia definitiva import
Garantia global
Garantia global percentatge
Garantia global import
Garantia provisional
Garantia provisional percentatge
Garantia provisional import
Revisió de preus
Entitat social
Número d'invitacions
Observacions
Reserva social article 35
Subhasta electrònica
Segon Adj.
Homologat
Agregat
Data BOE
Data DOGC
Data DOUE
Data BOP
Data premsa
Data altres
Data anunci



Data alta contracte
Data modificació contracte
Estat tramitació contracte
Data tramitació contracte
Data primera tramitació contracte
Modalitat
Complement d'obres
Forma adjudicació
Tipus subministraments
Classe expedient
Import licitació contracte
Import licitació contracte amb IVA
Número de lots
Imports determinar
Causa legal
Acord Marc
Sistema adjudicació
Criteri valoració
Llei TRLCSP
Sistema dinàmic
Reservat
Pertany cat2 llei 302007
Sense despesa associada
Modalitat determinació preus
Responsable contracte
Clàusules socials



Clàusules ambientals
Clàusules ètiques
Preu unitari
Contracte complementari
Classificació empresarial
Aportacions públiques construcció
Clàusules lingüístiques
Es contracte menor
Codi Organisme anterior
Número registre anterior
Any execució anterior
Compra pública d'innovació
S'ha comprat innovació?
Tramitació simplificada
Contractació conjunta



5. Bibliography

- Abdou, A., Czibik, Á., Tóth, B., & Fazekas, M. (2021). *COVID-19 emergency public procurement in Romania: Corruption risks and market behavior*.
- Anita, N., & Tünde, T. (2015). *New warning systems for the identification of red flags in public procurements*.
- Arosa Otero, D., Arvelo Flores, J. C., Cano Rodríguez, M., Colomer Pedrosa, A., & García Rodríguez, M. J. (2021). La Contratación Pública en España: Fuentes de Datos, Normativa y Aplicaciones Tecnológicas. *Revista de la Escuela Jacobea de Posgrado*(21), 87-112.
- Auriol, E., Straub, S., & Flochel, T. (2016). Public procurement and rent-seeking: the case of Paraguay. *World Development*, 77, 395-407.
- Aziani, A., Ferwerda, J., & Riccardi, M. (2020). Who are our owners? Exploring the ownership links of businesses to identify illicit financial flows. *European Journal of Criminology*, 1477370820980368.
- Beke, M. B., & Blomeyer&Sanz. (2015). *Towards a European Strategy to Reduce Corruption by Enhancing the Use of Open Data*. *National Research Spain*.
- Bernstein, D. S., Recanatini, F., & Georgieva-Andonovska, E. G. (2018). *Public sector governance indicators for EU regions*.
- Brandeis, L. (1914). *Other People's Money—and How Bankers Use It (cap. V)*.
- Calabuig Rodríguez, J. M., Falciani, H., Ferrer Sapena, A., García Raffi, L. M., Raso, E., Sánchez del Toro, I., & Sánchez Pérez, E. A. (2018). Detección y prevención de las malas prácticas y la corrupción desde la perspectiva de las matemáticas. *Revista internacional de transparencia e integridad*, 8.
- Caneppele, S., Calderoni, F., & Martocchia, S. (2009). Not only banks: Criminological models on the infiltration of public contracts by Italian organized crime. *Journal of Money Laundering Control*.
- Carlóni, E. (2017). Misurare la corruzione? indicatori di corruzione e politiche di prevenzione. *Politica del Diritto*, 3, 445-466.
- Cerrillo i Martínez, A. (2010). Cooperación entre Administraciones públicas para el impulso de la Administración electrónica. In E. Gamero Casado & J. Valero Torrijos (Eds.), *La Ley de Administración electrónica. Comentario sistemático a la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos* (3a ed.). Cizur Menor: Thomson-Aranzadi.
- Cerrillo i Martínez, A. (2012a). La contribución de las TIC a la mejora de la transparencia administrativa. *Arbor*, 188(756), 707-724.
- Cerrillo i Martínez, A. (2012b). The re-use of public sector information in Europe and its impact on transparency. *European Law Journal*, 18(6), 770–792.
- Cerrillo i Martínez, A. (2017). *Contratación Abierta*. Barcelona: Generalitat de Catalunya.
- Charron, N., Dahlström, C., Fazekas, M., & Lapuente, V. (2017). Careers, Connections, and Corruption Risks: Investigating the impact of bureaucratic meritocracy on public procurement processes. *The Journal of Politics*, 79(1), 89-104.
- Cichy, C., & Rass, S. (2019). An overview of data quality frameworks. *IEEE Access*, 7, 24634-24648.
- Cingolani, L., Fazekas, M., Martínez Barranco Kukutschka, R., & Tóth, B. (2015). Towards a comprehensive mapping of information on public procurement tendering and its actors across Europe.
- Clare, A., Sangokoya, D., Verhulst, S., & Young, A. (2016). *Open contracting and procurement in Slovakia*.
- Comisión Europea. (2017). *Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones de 3 de octubre de 2017 Conseguir que la contratación pública funcione en Europa y para Europa [COM(2017) 572 final]*.
- Comisión Nacional de los Mercados y la Competencia. (2015). *PRO/CNMC/001/15: Análisis de la contratación pública en España: Oportunidades de mejora desde el punto de vista de la competencia*.
- Comisión Nacional de los Mercados y la Competencia. (2019). *Overview of Public Procurement Procedures in Spain*. Madrid: Comisión Nacional de los Mercados y la Competencia
- Czibik, Á., Tóth, B., & Fazekas, M. (2015). How to Construct a Public Procurement Database from Administrative Records. *With examples from the Hungarian public procurement system of 2009–2012*.



- Decarolis, F., & Giorgiantonio, C. (2022). Corruption red flags in public procurement: new evidence from Italian calls for tenders. *EPJ Data Science*, 11(1), 16.
- Di Nicola, A., & McCallister, A. (2006). Existing experiences of risk assessment. *European Journal on Criminal Policy and Research*, 12(3), 179-187.
- Dragos, D. C., Kovač, P., & Marseille, A. T. (2018). *The laws of transparency in action: A European perspective*: Springer.
- Duguay, R., Rauter, T., & Samuels, D. (2020). The impact of open data on public procurement.
- Fazekas, M. (2017). *Assessing the quality of government at the regional level using public procurement data*. Brussels.
- Fazekas, M., Cingolani, L., & Tóth, B. (2018). Innovations in objectively measuring corruption in public procurement. *Governance indicators. Approaches, progress, promise*, 154-185.
- Fazekas, M., & Kocsis, G. (2020). Uncovering high-level corruption: cross-national objective corruption risk indicators using public procurement data. *British Journal of Political Science*, 50(1), 155-164.
- Fazekas, M., Nishchal, S., & Søreide, T. (2021). Public procurement under and after emergencies. In O. Bandiera, E. Bosio, & G. Spagnolo (Eds.), *Procurement in Focus Rules, Discretion, and Emergencies*. London: Centre for Economic Policy Research.
- Ferwerda, J., Deleanu, I., & Unger, B. (2017). Corruption in public procurement: finding the right indicators. *European Journal on Criminal Policy and Research*, 23(2), 245-267.
- G20, & OECD. (2017). *Compendium of good practices on the use of open data for Anti-corruption: Towards data-driven public sector integrity and civic auditing*.
- Gallego, J., Prem, M., & Vargas, J. F. (2020). Corruption in the Times of Pandemia. *Documentos de Trabajo*.
- Gottschalk, P., & Solli-Sæther, H. (2009). *E-Government Interoperability and Information Resource Integration: Frameworks for Aligned Development: Frameworks for Aligned Development*. Hershey-New York: IGI Global.
- Granickas, K. (2014). Open Data as a Tool to Fight Corruption. *European Public Sector Information Platform Topic Report, 2014/04*.
- Group of States against Corruption (GRECO). (2020). *Corruption Risks and Useful Legal References in the context of COVID-19*.
- Gualo, F., Rodríguez, M., Verdugo, J., Caballero, I., & Piattini, M. (2021). Data quality certification using ISO/IEC 25012: Industrial experiences. *Journal of Systems and Software*, 176, 110938.
- Halsbenning, S., Scholta, H., & Niemann, M. (2020). *The Data Doesn't Lie-An Approach for Form Usability Evaluation Based on Data Quality*. Paper presented at the ECIS.
- Hardman, J. (2021). The butterfly effect: Theoretical implications of an apparently minor corporate transparency proposal. *Common Law World Review*, 14737795211037701.
- Haute Autorité pour la transparence de la vie publique. (2016). *Open data & intégrité publique. Les technologies numériques au service d'une démocratie exemplaire*. Retrieved from <http://www.hatvp.fr/wordpress/wp-content/uploads/2016/12/Open-data-integrite-publique.pdf>
- Heitmüller, F., Harari, M., & Meinzer, M. (2019). Regulatory Capacity of Corporate Registries in the European Union. Survey Results. doi:dx.doi.org/10.2139/ssrn.3481438
- Huber, M., & Imhof, D. (2019). Machine learning with screens for detecting bid-rigging cartels. *International Journal of Industrial Organization*, 65, 277-301.
- Hyytinen, A., Lundberg, S., & Toivanen, O. (2007). Politics and procurement: Evidence from cleaning contracts.
- Independent Commission Against Corruption. (2020). *Managing corrupt conduct during the COVID-19 outbreak*. New South Wales.
- Integrity Vice Presidency. (2009). *Fraud and Corruption Awareness Handbook: How It Works and What to Look For*. Washington, D.C.
- Jiménez, C. E., Criado, I., & Gascó, M. (2011). Technological e-government interoperability. An analysis of Ibero American countries. *IEEE Latin American Transactions*, 9(7), 1112-1117.



- Juddoo, S., George, C., Duquenoy, P., & Windridge, D. (2018). Data Governance in the Health Industry: Investigating Data Quality Dimensions within a Big Data Context. *Applied System Innovation*, 1(4), 43. Retrieved from <https://www.mdpi.com/2571-5577/1/4/43>
- Kane, R. J., & White, M. D. (2009). Bad cops: A study of career-ending misconduct among New York City police officers. *Criminology & Public Policy*, 8(4), 737-769.
- Kaufmann, D. (2002). *Transparency, Incentives and Prevention (TIP) for Corruption Control and Good Governance Empirical Findings, Practical Lessons, and Strategies for Action based on International Experience*. Paper presented at the Qinghua University-Carnegie Conference on Economic Reform and Good Governance: Fighting Corruption in Transition Economies, Beijing.
- Kaufmann, D. (2005). Myths and Realities of Governance and Corruption. *Working Paper Series. The Brookings Institution*.
- Kaufmann, D., & Bellver, A. (2005). *Transparenting Transparency: Initial Empirics and Policy Applications*. Retrieved from <http://mpra.ub.uni-muenchen.de/8188/>
- Kayte, S., & Schneider-Kamp, P. (2019). *A Mixed Neural Network and Support Vector Machine Model for Tender Creation in the European Union TED Database*. Paper presented at the KMIS.
- Kenny, C., & Musatova, M. (2010). 'Red Flags of Corruption' in World Bank Projects. An Analysis of Infrastructure Contracts. *Policy Research Working Paper*, 5243.
- Knobel, A., Meinzer, M., & Harari, M. (2017). What Should Be Included in Corporate Registries? A Data Checklist-Part 1: Beneficial Ownership Information. *A Data Checklist-Part, 1*.
- Köbis, N., Starke, C., & Rahwan, I. (2021). Artificial Intelligence as an Anti-Corruption Tool (AI-ACT)--Potentials and Pitfalls for Top-down and Bottom-up Approaches. *arXiv preprint arXiv:2102.11567*.
- Kotlyar, D., & Pop, L. (2021). *Automated Risk Analysis of Asset and Interest Declarations of Public Officials. A Technical Guide. The Stolen Asset Recovery Initiative (StAR)*. Washington, D.C.
- Kutlina-Dimitrova, Z. (2018). Government procurement: data, trends and protectionist tendencies. *Chief Economist Note*(3), 1-27.
- Lindstedt, C., & Naurin, D. (2010). Transparency is not Enough: Making Transparency Effective in Reducing Corruption. *International Political Science Review*, 31(3), 301-322. doi:10.1177/0192512110377602
- Mantini, P. (2015). *Nel cantiere dei nuovi appalti pubblici. Semplificazione, efficienza, concorrenza, anticorruzione*: Giuffrè editore.
- Mendes, M., & Fazekas, M. (2017). Recommendations for the Implementation of Open Public Procurement Data. An Implementer's Guide.
- Mendes, M., & Fazekas, M. (2018). *DIGIWHIST Recommendations for the Implementation of Open Public Procurement Data An Implementer's Guide*.
- Merloni, F., & Ponti, B. (2010). La trasparenza. In F. Merloni & L. Vandelli (Eds.), *La corruzione amministrativa. Cause, prevenzione e rimedi*. Firenze-Antella: Passigli Editori.
- Miranzo Díaz, J. (2020). *Hacia la configuración de una estrategia eficiente de integridad en la contratación pública*. Cizur Menor: Thomson-Reuters-Aranzadi.
- Miranzo Díaz, J., Cerrillo i Martínez, A., Castro, J., & Galindo Caldés, R. (2022). Corruption Risk Analysis in Local Public Procurement: A Look at the Àrea Metropolitana de Barcelona. *International Review of Administrative Sciences, forthcoming*.
- Moraga, C., Moraga, M. Á., Calero, C., & Caro, A. (2009). *SQuaRE-aligned data quality model for web portals*. Paper presented at the 2009 Ninth International Conference on Quality Software.
- Németh, A., & Tátrai, T. (2015). Red Flags Project: New Warning System for the Identification of Red Flags in Public Procurements.
- OCDE. (2020). *Integridad pública para una respuesta y recuperación efectivas ante el COVID-19*. Paris: OCDE
- OECD. (2003). *Managing Conflict of Interest in the Public Sector. OECD Guidelines and Country Experiences*. Paris: OCDE
- OECD. (2005). *Managing Conflict of Interest in the Public Sector*. Retrieved from Paris:
- OECD. (2007). *Integrity in Public Procurement. Good practice from A to Z*. Paris: OCDE
- OECD. (2009). *OECD Principles for Integrity in Public Procurement*. Paris: OCDE
- OECD. (2017). *Public integrity. A strategy against corruption*. Paris: OCDE



- OECD. (2020). *Preventing and Managing Conflicts of Interest in the Public Sector. Good practices guide. Prepared at the request of the G20 Anticorruption Working Group by the World Bank, OECD and UNODC*. Paris: OCDE
- Oficina Antifrau de Catalunya. (2021). *Riscos per a la integritat en la contractació pública: informe final*.
- Oficina Independiente de Regulación y Supervisión de la Contratación. (2020). *Informe especial de supervisión. Principio de publicidad en los contratos tramitados por emergencia durante la vigencia de la declaración del estado de alarma como consecuencia del COVID-19*. Madrid: Oficina Independiente de Regulación y Supervisión de la Contratación.
- Oficina Independiente de Regulación y Supervisión de la Contratación. (2021). *Informe de Supervisión de la Contratación Pública de España*. Madrid: Oficina Independiente de Regulación y Supervisión de la Contratación.
- O'Neil, C., & Schutt, R. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc.
- Open Contracting Partnership. (2016). *Open contracting. A guide for practitioners by practitioners*.
- Page, D., McClelland, D., Oyenubi, A., Bridgman, G., & Kollamparambil, U. (2021). COVID-19 emergency procurement trends, deviations, expansions, and excessive pricing. *SA-TIED Technical Note, 5*.
- Petheram, A., Pasquarelli, W., & Stirling, R. (2019). *The Next Generation of Anti-Corruption Tools: Big Data, Open Data and Artificial Intelligence*.
- Poerting, P., & Vahlenkamp, W. (1998). Internal strategies against corruption: Guidelines for preventing and combating corruption in police authorities. *Crime, Law and Social Change, 29(2)*, 225-249.
- Prier, E., Prysmakova, P., & McCue, C. P. (2018). Analysing the European Union's tenders electronic daily: Possibilities and pitfalls. *International Journal of Procurement Management, 11(6)*, 722-747.
- Rabuzin, K., & Modrusan, N. (2019). *Prediction of Public Procurement Corruption Indices using Machine Learning Methods*. Paper presented at the KMIS.
- Rose-Ackerman, S. (2004). Governance and Corruption. In B. Lomborg (Ed.), *Global Crises, Global Solutions*. Cambridge: Cambridge University Press.
- Rose-Ackerman, S., & Palifka, B. J. (2019). *La corrupción y los gobiernos. Causas, consecuencias y reforma* (2a ed.). Madrid: Marcial Pons.
- Sanchez-Graells, A. (2020). Procurement in the Time of Covid-19 Notes and Commentaries. *N. Ir. Legal Q., 71*.
- Schultz, J., & Søreide, T. (2008). Corruption in emergency procurement. *Disasters, 32(4)*, 516-536. doi:10.1111/j.1467-7717.2008.01053.x
- Soylu, A., Corcho, O., Elvesæter, B., Badenes-Olmedo, C., Blount, T., Yedro Martínez, F., . . . Taggart, C. (2021). TheyBuyForYou platform and knowledge graph: Expanding horizons in public procurement with open linked data. *Semantic Web, 13*, 265–291. doi:10.3233/SW-210442
- Soylu, A., Corcho, Ó., Elvesæter, B., Badenes-Olmedo, C., Yedro-Martínez, F., Kovacic, M., . . . Taggart, C. (2022). Data Quality Barriers for Transparency in Public Procurement. *Information, 13(2)*, 99.
- TACOD. (2015). *Revolution delayed. The impact of open data on the fight against corruption*. Retrieved from <https://www.transparency.it/wp-content/uploads/2015/09/2015-TACOD-REPORT.pdf>
- Tátrai, T., & Németh, A. (2018). *Improving red flag instruments for public procurement*. Paper presented at the ERA Forum.
- Thibous, J.-F. (2019). La transformation numérique de la commande publique: la dématérialisation complète pour 2022. *Gestion Finances Publiques(3)*, 82-88.
- Tirado, J. M., Serban, O., Guo, Q., & Yoneki, E. (2016). Web data knowledge extraction. *University of Cambridge, Computer Laboratory, technical report, 881*.
- Transparencia Internacional España. (2020). *Transparencia y publicidad activa: COVID-19 y estado de alarma en España. Recomendaciones para la transparencia y prevención de la corrupción en el sector público y privado*.
- Tribunal de Contas. (2020). Processo No. 1/2020, Riscos na Utilização de Recursos Públicos na Gestão de Emergência (COVID-19). Retrieved from <https://www.tcontas.pt/pt-pt/ProdutosTC/Relatorios/relatorios-oac/Documents/2020/relatorio-oac-2020-01.pdf>



- Troncoso Reigada, A. (2009). Reutilización de información pública y protección de datos personales. *Revista general de información y documentación*, 19(1).
- U4. Anticorruption Research Centre. (2015). Ebola and corruption: Overcoming critical governance challenges in a crisis situation. Retrieved from <https://www.u4.no/publications/ebola-and-corruption-overcoming-critical-governance-challenges-in-a-crisis-situation.pdf>
- United Nations. (2014). *A world that counts. Mobilizing the data revolution for sustainable development*.
- Valcárcel Fernández, P. (2019). Tres dimensiones de la transparencia en la contratación pública. Rendición de cuentas, respeto de los derechos de operadores económicos y mejora global de la gestión de este sector a través del big data. In J. M. Gimeno Feliu (Ed.), *Observatorio de contratos públicos 2018* (pp. 93-129). Cizur Menor: Thomson Reuters-Aranzadi.
- Vasconcelos, M. O., & Cavique, L. (2022). Dataset for corruption risk assessment in a public administration. *Data in brief*, 40. doi:10.1016/j.dib.2021.107768
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325-337.
- Villamil, I., Kertész, J., & Wachs, J. (2022). Computational Approaches to the Study of Corruption. *arXiv preprint arXiv:2201.11880*.
- Web Foundation. (2018). *Barómetro de los Datos Abiertos - Edición de los Líderes*. Washington DC:
- Wensink, W., & de Vet, J. M. (2013). Identifying and Reducing Corruption in Public Procurement in the EU. Study prepared for the European Commission by PwC and Ecorys.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. doi:10.1038/sdata.2016.18
- World Bank. (2007). *Improving development outcomes – annual integrity report*. Washington, D.C.: The World Bank
- World Bank Group. (2020a). Ensuring Integrity in Governments' Response to COVID-19. *Governance & Institutions COVID-19 Response Resources*, April 28. Retrieved from <https://openknowledge.worldbank.org/handle/10986/33705>
- World Bank Group. (2020b). *Governance and Institutions Emergency Measures for State Continuity during COVID-19 Pandemic*.
- Worthy, B. (2020). Freedom of information in Europe: Creation, context, and conflict. In *Transparency and Secrecy in European Democracies* (pp. 36-59): Routledge.



Appendix

Appendix 1 PSCP data fields list

PSCP - CONTRACT SIGNATURE

Column
Acords:
Àmbit geogràfic:
BOE:
BOP:
Cànon d'exploració:
Codi CPA:
Codi CPV:

PSCP - AWARD NOTICES

Column
Acords:
Àmbit geogràfic:
BOE:
BOP:
Cànon d'exploració:
Codi CPA:
Codi CPV:

PSCP - CONTRACT NOTICES

Column
Àmbit geogràfic:
BOP:
Cànon d'exploració:
Centres especials d'ocupació:
Classificació empresarial:
Codi CPA:
Codi CPV:



Codi d'expedient:

Compra pública d'innovació:

Condicions d'execució:

Condicions i obligacions assumides:

Contracte harmonitzat:

Dades de l'empresa adjudicatària:

Dades de les empreses adjudicatàries:

Data d'adjudicació del contracte:

Data de formalització del contracte:

Data de publicació:

Denominació:

Descripció de la prestació:

Descripció de l'esmena:

Codi d'expedient:

Compra pública d'innovació:

Condicions d'execució:

Condicions i obligacions assumides:

Contracte desert

Contracte harmonitzat:

Dades de l'empresa adjudicatària:

Dades de les empreses adjudicatàries:

Data d'adjudicació del contracte:

Data de publicació:

Denominació:

Descripció de la prestació:

Descripció de l'esmena:

Codi d'expedient:

Compra pública d'innovació:

Condicions d'execució:

Contracte harmonitzat:

Criteris d'adjudicació:

Criteris objectius per a la selecció del
nombre limitat de candidats:

Data de publicació:

Descripció de la prestació:

Descripció de l'esmena:



<p>Descripció del lot:</p> <p>DOGC:</p> <p>DOUE:</p> <p>Drets i deures dels usuaris:</p> <p>Durada del contracte:</p> <p>Es compra innovació:</p> <p>Es preveuen modificacions als plecs:</p> <p>Facultats d'inspecció:</p> <p>Fase Compra Pública Precomercial</p> <p>Identitat d'empreses licitadores:</p>	<p>Descripció del lot:</p> <p>DOGC:</p> <p>DOUE:</p> <p>Drets i deures dels usuaris:</p> <p>Durada del contracte:</p> <p>Es compra innovació:</p> <p>Es preveuen modificacions als plecs:</p> <p>Facultats d'inspecció:</p> <p>Identitat d'empreses licitadores:</p>	<p>Descripció:</p> <p>Divisió en lots:</p> <p>DOGC:</p> <p>DOUE:</p> <p>Durada del contracte:</p> <p>Elements i condicions de les variants:</p> <p>Es preveuen modificacions als plecs:</p> <p>Expressió:</p> <p>Garantia provisional:</p> <p>Identificació d'àmbits de negociació:</p>
--	--	---



COrruption
Risk Indicators in
REmergency

Co-funded by the
European Union



Import sense IVA:

Import:

Informació complementària sobre la
pròrroga:

Informes Tècnics:

Nacionalitat:

Import sense IVA:

Import:

Informació complementària sobre la
pròrroga:

Informes Tècnics:

Lot desert

Motiu d'adjudicació:

Nacionalitat:

Informació complementària sobre la
pròrroga:

Limitació d'empreses candidates:

Mínim d'acreditació:

Mitjà de solvència:

Nombre de lots:

Nombre màxim de lots al que un mateix
licitador pot presentar oferta:

Nombre màxim de lots que es pot
adjudicar a un mateix licitador:

Nombre màxim:

Nombre mínim:



Número de lot:	Número de lot:	
Número d'ofertes rebudes:	Número d'ofertes rebudes:	
Obertura de pliques:	Obertura de pliques:	Obertura de pliques:
Observacions:	Observacions:	Observacions:
		Oferta integradora:
Òrgan de contractació:	Òrgan de contractació:	Òrgan de contractació:
		Període:
		Plec de clàusules administratives
		Plec de prescripcions tècniques
		Pressupost de licitació:
Pressupost de licitació:	Pressupost de licitació:	Procediment d'adjudicació:
Procediment d'adjudicació:	Procediment d'adjudicació:	
Procediments per a formular queixes	Procediments per a formular queixes	
Projecte finançat amb fons del Mecanisme de Recuperació i Resiliència:	Projecte finançat amb fons del Mecanisme de Recuperació i Resiliència:	Projecte finançat amb fons del Mecanisme de Recuperació i Resiliència:
Projecte finançat amb fons REACT-EU:	Projecte finançat amb fons REACT-EU:	Projecte finançat amb fons REACT-EU:
Pròrroga:	Pròrroga:	Pròrroga:
Quadre comparatiu d'ofertes i puntuacions:	Quadre comparatiu d'ofertes i puntuacions:	



Requisits de prestació del servei:

Reserva social:

Simplificat de tramitació sumària:

Sistema de racionalització:

Subhasta electrònica:

Subtipus de contracte:

Termini de presentació d'ofertes:

Termini d'execució:

Tipus de contracte:

Tipus de tramitació:

Requisits de prestació del servei:

Reserva social:

Resolució d'adjudicació :

Simplificat de tramitació sumària:

Sistema de racionalització:

Subhasta electrònica:

Subtipus de contracte:

Termini de presentació d'ofertes:

Termini d'execució:

Termini per a la formalització del contracte:

Tipus de contracte:

Tipus de tramitació:

Reserva social:

S'accepten variants:

Simplificat de tramitació sumària:

Sistema de racionalització:

Solvència econòmica i financera

Solvència tècnica i professional

Subhasta electrònica:

Subtipus de contracte:

Termini de presentació d'ofertes:

Termini d'execució:

Tipus de contracte:

Tipus de tramitació:



Tramitació amb mesures de gestió
eficient:

Tramitació simplificada:

Unitat de contractació:

Url

Valor estimat del contracte:

Tramitació amb mesures de gestió
eficient:

Tramitació simplificada:

Unitat de contractació:

Url

Valor estimat del contracte:

Tramitació amb mesures de gestió
eficient:

Tramitació simplificada:

Unitat de contractació:

Url

Valor estimat del contracte:



Appendix 2.A RPC data list

Contingut del Registre públic de contractes

1.1 A. Dades d'adjudicació del contracte

Tipus de contracte.

Entitat contractant.

Any d'execució.

Nombre de lots.

Número de l'expedient de contractació.

Tramitació ordinària, urgent, d'emergència, anticipada ordinària, anticipada urgent.

Caràcter pluriennal.

Descripció de l'objecte del contracte.

Procediment d'adjudicació (obert, restringit, negociat amb publicitat, negociat sense publicitat, derivat d'acord marc, diàleg competitiu, específic de sistema dinàmic, contracte menor, altres adjudicacions segons instruccions internes).

Sistema de racionalització de la contractació (acord marc, sistema dinàmic de contractació).

Criteri de determinació de l'oferta més avantatjosa (un criteri, diversos criteris).

Indicador de subhasta electrònica.

Pressupost de licitació.

Valor estimat del contracte.

Indicador d'import licitació a determinar.

Preus unitaris.

Modalitat de la contractació, en contractes de gestió de serveis públics.

Durada de la contractació, en contractes de gestió de serveis públics.

Tipus de subministrament.

Indicador de contracte complementari.



Modalitat de determinació del preu, en contractes de serveis.

Aportacions públiques a la construcció, en contractes de concessió d'obra pública.

Causa legal, en procediments negociats.

Identificació de l'acord marc o del sistema dinàmic de contractació.

Nombre d'invitacions, en procediments negociats amb publicitat i sense publicitat i en procediments restringits.

Indicador de contracte amb clàusula d'arbitratge.

Indicador de designació de responsable del contracte.

Indicador de contracte amb clàusules socials. Tipus de clàusules socials.

Indicador de contracte amb clàusules ambientals.

Indicador de contracte amb clàusules d'R+D+I.

Data de publicació al DOGC.

Data de publicació al BOE.

Data de publicació al DOUE.

Data de publicació al BOP.

Data de publicació en altres diaris oficials.

Data de publicació en la premsa diària

Data de publicació en perfil de contractant.

1.1 B. Documents contractuals

Contracte.

Plec de clàusules administratives particulars.

Plec de prescripcions tècniques particulars, si escau.

Projecte, si escau.

Qualsevol altre document que conformi el contracte, si escau.

1.2 Dades del lot



Número del lot.

Import licitació.

Termini d'execució.

Lloc d'execució (província).

Classificació exigida: en contractes d'obres i serveis.

Garanties exigides (provisional, definitiva, complementària i global).

Durada de la concessió, en contractes de concessió d'obra pública.

Import d'adjudicació.

Tipus d'IVA.

Partida pressupostària, en contractes d'estudis i dictàmens.

Data d'adjudicació.

Data de formalització.

Data d'inici d'execució.

Data fi prevista.

Indicador contracte mixt.

Indicador lot desert.

Data d'aprovació del Govern o òrgan competent, en contractes pluriennals.

Descripció de l'objecte del lot.

País de procedència del producte, en contractes de subministrament.

Codi del vocabulari comú de contractes (CPV).

Import cànon, si escau.

Periodicitat del cànon, si .escau.

Modalitats que determinen l'import del contracte, si escau.

Import subvenció de l'Administració.

Revisió de preus (no prevista, IPC, altres índexs oficials, fórmules tipus generals, altres).

Indicador de cofinançament.



Import de cofinançament.

Detall de les ofertes presentades.

NIF/CIF de l'empresa contractista.

Nom de l'empresa contractista.

Domicili social de l'empresa contractista: país, província i municipi.

Indicador de si l'empresa contractista és o no una UTE.

Llista d'empreses que formen la UTE.

Tipus d'entitat (indicador d'entitat pertanyent al tercer sector, en què s'identifica si es tracta d'un centre especial de treball, una empresa d'inserció social, una fundació, una associació o d'altres).

2.1 Dades sobre modificacions del contracte

Tipus.

Import de la modificació o modificacions.

Data d'aprovació.

Codi d'expedient.

Indicador modificació superior al 10%.

Causa de la modificació.

Anys termini.

Mesos termini.

Dies termini.

Variació del termini d'execució o durada.

2.2 Dades sobre cessió del contracte

Data d'inici.

Data aprovació.



Import restant d'execució.

2.3 Dades sobre subcontractació del contracte

Data d'inici.

Data de notificació a l'òrgan contractant.

Import subcontractat.

Percentatge subcontractació.

Nom i NIF de l'empresa subcontractada.

Indicador de si l'empresa subcontractada és o no una UTE.

Domicili social de l'empresa subcontractada: país, província i municipi.

Tipus d'entitat (indicador d'entitat pertanyent al tercer sector, en què s'identifica si es tracta d'un centre especial de treball, una empresa d'inserció social, una fundació, una associació o d'altres).

3. Dades sobre pròrroga del contracte

Data d'inici.

Data fi.

Data acord.

Data formalització.

Import prorrogat.

Codi expedient.

Indicador de pluriennal, si escau.

4. Dades sobre extinció del contracte

Tipus d'extinció.



Indicador revisió de preus.

Data de recepció, en extincions per compliment.

Data de resolució, en extincions per resolució.

Causa de resolució, en extincions per resolució.

Indicador d'incautació de garantia definitiva, en extincions per resolució.

Import de liquidació o import final que s'ha de pagar.

Data de liquidació, en extincions per compliment.

Indicador de termini d'execució prorrogat, en extincions per compliment.

Anys de termini.

Mesos de termini.

Indicador de contracte objecte de recurs, tipus de recurs i sentit de la resolució.

Appendix 2.B RPC data field list (search engine)

Nom organisme contractant
Codi unitat de contractació
Nom unitat de contractació
Número registre
Any execució
Expedient
Tipus contracte
Procediment adjudicació
Descripció del contracte
Descripció del lot
Lot
Lot desert
NIF/CIF contractista
Nom contractista
Import adjudicat lot/pròrroga
Import licitat lot
Import licitat lot amb IVA
Import VEC
Pluriennal
Codi liquidació
Data inici execució
Termini durada anys
Termini durada mesos lot
Termini durada dies lot
Despesa anticipada



Codi CPV
Grup contracte
Import adjudicat lot/pròrroga amb IVA
Data adjudicació
Data aprovació
Data formalització
Anualitats
Tipus IVA
Procedència de producte
Concessió
Data recepció
Data liquidació
Import liquidació compliment
Import liquidació compliment amb Iva
Termini prorrogat
Anys prorrogat
Mes prorrogat
Causa resolució
Data resolució
Revisió de preus
Incautació garantia
Codi CPA
Obra pública
Contracte mixt
Contracte marc
IVA múltiple



Import adjudicació IVA múltiple
Cofinanciació
Import cofinançat
Import cofinançat amb IVA
Import cànon
Import cànon amb IVA
Import subvenció administració
Import subvenció administració amb IVA
Percentatge subvenció administració
Import liquidació resolució
Import liquidació resolució amb Iva
Període cànon
Tipus pagament
Tipus de producte
Estudis i dictàmens
Data tramesa lot
Data alta
Data modificació
Data primera tramitació
Data tramitació contracte
Compte Comptable
Contracte objecte
Estat
Estat de la liquidació
Classificació



Tipus recurs
Sentit de la resolució
País origen
Partida pressupostaria
Numero lot anterior
Motiu utilitat contracte
Lloc execució
Foment social
Garantia complementària
Garantia complementària percentatge
Garantia complementària import
Garantia definitiva
Garantia definitiva percentatge
Garantia definitiva import
Garantia global
Garantia global percentatge
Garantia global import
Garantia provisional
Garantia provisional percentatge
Garantia provisional import
Revisió de preus
Entitat social
Número d'invitacions
Observacions
Reserva social art.35
Subhasta electrònica



Segon Adj.
Homologat
Agregat
Data BOE
Data DOGC
Data DOUE
Data BOP
Data premsa
Data altres
Data anunci
Data alta contracte
Data modificació contracte
Estat tramitació contracte
Data tramitació contracte
Data primera tramitació contracte
Modalitat
Complement d'obres
Forma adjudicació
Tipus subministraments
Classe expedient
Import licitació contracte
Import licitació contracte amb IVA
Número de lots
Imports determinar
Causa legal
Acord Marc



Sistema adjudicació
Criteri valoració
Llei TRLCSP
Sistema dinàmic
Reservat
Pertany cat2 llei 302007
Sense despesa associada
Modalitat determinació preus
Responsable contracte
Clàusules socials
Clàusules ambientals
Clàusules ètiques
Preu unitari
Contracte complementari
Classificació empresarial
Aportacions públiques construcció
Clàusules lingüístiques
Es contracte menor
Codi Organisme anterior
Número registre anterior
Any execució anterior
Compra pública d'innovació
S'ha comprat innovació?
Tramitació simplificada
Contractació conjunta