



Associate partners:

| | |
|-------------------------------------|------------------------|
| ETUI, Brussels | Solidarność, Katowice |
| UNI Europa, Brussels | Solidarność, Gdansk |
| WageIndicator Foundation, Amsterdam | CCOO Servicios, Madrid |

COLBAR-EUROPE

EUROPE-wide analyses of COLlective BARgaining agreements

Supported by the European Commission - Industrial Relations and Social Dialogue Program (nr project grant number of COLBAR is VS/2019/00777)¹ April 2019 - March 2021

REPORT 10: Feasibility of a Europe-wide data collection of Collective Agreements

Kea Tijdens, WageIndicator Foundation

4 March 2021



¹ Disclaimer: The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

TABLE OF CONTENT

| | |
|---|-----------|
| Management summary | 1 |
| 1 Introduction | 3 |
| 2 Literature review | 5 |
| 2.1 Four dimensions of collective bargaining | 5 |
| 2.2 The impact of collective bargaining coverage on wages and working conditions..... | 6 |
| 2.3 The impact of CBA clauses on wages and working conditions | 7 |
| 3 Overview of CBA registries in Europe | 11 |
| 3.1 Registries of CBAs..... | 11 |
| 3.2 Coding of CBAs..... | 13 |
| 3.3 How many CBAs are there in Europe?..... | 15 |
| 3.4 Feasibility of EU-wide CBA Database..... | 17 |
| 3.5 Feasibility of representative sampling of CBAs..... | 18 |
| 4 Technical feasibility of a Europe-wide CBA Database..... | 20 |
| 4.1 The WageIndicator CBA Database | 20 |
| 4.2 Gathering CBA texts | 21 |
| 4.3 Coding CBA texts..... | 22 |
| 4.4 Machine-reading technologies | 23 |
| 5 Conclusions | 25 |
| 6 References | 27 |

TABLE OF FIGURES

| | | |
|----------|--|----|
| Figure 1 | Coding of a collective agreement in Estonia, by the Ministry of Social Affairs (CBA of the Tallinn University of Technology) | 14 |
| Figure 2 | Coding of a collective agreement in Austria, by the Chamber of Commerce (CBA of the Landes-Hypothekbanken) | 14 |

TABLE OF TABLES

| | | |
|---------|--|----|
| Table 1 | CBA registration required, registry agency, archive maintained, full text downloadable, coded CBA by country | 12 |
| Table 2 | Estimated number of collective agreements, if possible broken down by Multi- and Single-employer agreements | 16 |
| Table 3 | Feasibility of including the country's CBAs in a EU-wide CBA Database, January 2021 | 18 |
| Table 4 | Ten topics in the coding scheme of the WageIndicator CBA Database | 23 |

Acknowledgements

The COLBAR-EUROPE project and its data-collection, reports and online webinars were made possible by the great efforts of the teams at WageIndicator Foundation, CELSI, CNEL and coordinator Amsterdam Institute for Advanced Labour Studies at the University of Amsterdam. We are grateful to all Collective Bargaining actors and all Archives of Collective Bargaining Agreements who helped acquiring full text collective agreements from European countries.

Management summary

In their renowned study “What do unions do?”, Freeman and Medoff (1984) argue that trade unions bargain for higher wages, equal pay and fair working conditions, but little is known about the bargaining outcomes agreed in collective bargaining agreements (CBAs) in the EU. In view of the re-launched dialogue with social partners at European level, such data is critical for monitoring progress in wage-setting and working conditions. Yet knowledge about what exactly is concluded in collective bargaining remains a blind spot. This article aimed to explore the feasibility of collecting and coding CBA texts in Europe to tackle this blind spot in the body of knowledge in industrial relations research, knowing that Internet may accelerate data collection. This article is based on desk research, experience with three EU-funded projects to explore the content of CBAs and the WageIndicator CBA Database.

The review of research showed that most knowledge about the impact of collective bargaining on wages and working conditions is based on survey data where a binary variable for coverage versus non coverage is used. Research using coded information about the content of CBAs revealed a fragmented picture with studies that coded the content of CBAs for the purpose of the study, not for tracing changes in CBAs over time.

The overview of CBA registries in the European Union revealed that in almost all EU countries CBAs need to be registered with the government or with an institution acting on its behalf to become binding. More than half of the EU countries maintain a registry of extended CBAs and most of these countries post the registry online. Various countries identify the metadata, the topics addressed or the headings used in the CBAs and a few countries code the CBA texts. Our study estimated the number of sector and company CBAs in EU28 in 2019 at 185,382 CBAs, of which Germany and France jointly make up 85%. Another six countries contribute between 1,000 and 8,000 CBAs each (Czech Republic, Hungary, Latvia, Romania, Spain, UK), and the remaining 19 countries less than 1,000 CBAs each. In exploring the feasibility of a EU-wide CBA Database it became clear that such a database could include either all or sector CBAs only from 11 countries. For another 10 countries some but not all CBAs could be included and for seven countries hardly any CBAs could be included.

The building blocks of an EU-wide CBA registry-in-the-making should specify the requirements for gathering, coding and annotating CBAs. The WageIndicator CBA Database can be taken as an example, as are the options for machine-reading of CBA texts.

Three technical components are essential for a Europe-wide data collection of collective agreements. The CBA texts need to be gathered and prepared for a coding tool, the possibilities of drafting representative samples for countries with many CBAs need to be explored, and coding of CBA texts is essential for understanding of CBA clauses. The WageIndicator CBA Database was populated with annotated and coded CBA texts in consecutive EU projects. Its coding scheme addresses ten topics to capture the core of each CBA. WageIndicator has explored machine reading options and concludes that this is a promising way forward to reduce the time needed for coding and thus allowing for the annotation and coding of more CBAs. WageIndicator is currently trying to understand the minimum number of CBAs and clauses needed for a satisfactory reliability score and which number of keywords allows for increasing reliability.

1 Introduction

In their renowned study “What do unions do?”, Freeman and Medoff (1984) argue that trade unions bargain for higher wages, equal pay and fair working conditions, implying that collective bargaining is central to wage setting processes and that wage outcomes will vary according to the wage levels agreed in collective bargaining. Almost forty years later, little is known about the wage outcomes agreed in collective bargaining agreements (CBAs) in the European Union (EU). The CAWIE report (Van Gyes and Schulten, 2015) showed that some knowledge of agreed wages is available from National Statistical Offices or Central Banks for ten EU countries only. Even less is known about the working conditions agreed in CBAs in the EU whereas, in view of the European Commission’s (EC) re-launched dialogue with social partners at European level (European Commission, 2016), such data is critical for monitoring progress in wage-setting and in setting standards for working conditions. The knowledge gap remains evident in the ESDE 2020 report (European Commission, 2020), in the OECD ‘Negotiating Our Way Up’ report (2019), and in EUROFOUND’s flagship report ‘Industrial relations: Developments 2015–2019’ (Eurofound 2020). Knowledge about the impact of collective bargaining is based on survey data of individuals or companies, on inventories of national bargaining systems or on legal regulations, but not on details about what exactly has been agreed in the large number of CBAs concluded throughout Europe. A main reason for this knowledge gap is that a European-level registry that collects and codes CBAs is lacking. In the EU knowledge about what exactly is concluded in collective bargaining remains a blind spot; therefore, neither cross-sectional nor longitudinal knowledge can be accumulated. This article aims to explore the feasibility of building a database of CBAs in Europe to tackle this blind spot in the body of knowledge in industrial relations research, taking as a starting point that the use of the Internet and advances in natural language processing may accelerate data collection and coding (Askitas and Zimmermann, 2015).

In section 2 the body of knowledge regarding the impact of CBA clauses on wages and working conditions is reviewed. Section 3 explores which EU countries maintain a CBA registry, how large the stock of CBAs is in Europe, whether an EU-wide registry is feasible, and what the possibilities are for representative sampling. Section 4 details the building blocks of an EU-wide CBA registry-in-the-making, including the requirements for gathering, coding and annotating CBAs with the WageIndicator CBA Database as an example and the options for machine-reading of CBA texts. Section 5 draws conclusions on how to establish a

continuous, Europe-wide data collection of coded CBAs and thus fulfilling the main condition for contributing to the body of knowledge about the outcomes of collective bargaining throughout the EU.

This article is based on desk research and on three experiences with CBA Databases. A first one regards the coded database of collective agreements in the Netherlands maintained by the trade union confederation FNV and the employers' association AWWN (Tijdens and Van Klaveren, 2003; Schreuder and Tijdens, 2004; Yerkes and Tijdens, 2010). A second one concerns the coding of collective agreements by means of survey questions in Europe in an EC-funded Social Dialogue project (WIBAR-3 VS/2014/0533). A third one concerns experiences with the WageIndicator CBA Database in two consecutive EC-funded Social Dialogue projects (BARCOM VS/2016/0106, COLBAR-EUROPE VS/2019/0077), BARCOM VS/2016/0106, COLBAR-EUROPE VS/2019/0077, Ceccon et al., 2016).²

² See for details about the projects <https://wageindicator.org/Wageindicatorfoundation/projects>

2 Literature review

2.1 Four dimensions of collective bargaining

Four dimensions can be distinguished when reviewing the literature about collective bargaining in a cross-country comparative perspective. The first dimension refers to the bargaining systems, the hierarchy in these systems, and the collective bargaining coverage. Most studies addressing collective bargaining have discussed country-level trends in bargaining coverage, extension regimes, wage coordination, vertical structure of collective bargaining, and issues related to single- versus multi-employer bargaining, while often using the ICTWSS database (OECD and AIAS, 2021). The second dimension refers to the actors involved in collective bargaining processes, including employers, employers' organisations, trade unions, and other workers' representatives. In its overview of collective bargaining in Europe in the 21st century, Eurofound (2015) extensively elaborated on the actors involved in bargaining. The third dimension covers the power relations between the actors and the incidence of industrial action, as captured in Eurofound (2019) and in the ETUI/AIAS Collective Bargaining Newsletter.³ The fourth dimension covers the outcomes of bargaining processes, thus the content of CBAs. These outcomes refer to topics such as wages or wage increases; working hours; working schedules and holidays; sickness and disability arrangements; social security; training; work-family arrangements; job security; internal mobility; work organisation, or workforce numbers. Across Europe, the body of knowledge on bargaining systems, processes and actors is better developed than the one on bargaining outcomes. Currently, it is not exaggerated to characterize the level of knowledge about what exactly is concluded in CBAs as a blind spot. In the remainder of this section, we focus on this fourth dimension.

Following the argumentation of Freeman and Medoff (1984), we take the position that knowledge about the outcomes of collective bargaining is relevant to explore the impact of collective bargaining. The absence of knowledge about bargaining outcomes is most likely related to the fact that at the European level no person or institution is systematically

³ See <https://www.etui.org/newsletters/collective-bargaining-newsletter>

collecting full texts of CBAs and coding their content. Some EU member states maintain national databases with coded content of their collective agreements, but a majority does not. However, due to different coding schemes and coding practices even for the latter countries, the coded content is not comparable across countries. Moreover, national databases are available for trade union members only in some countries.

2.2 The impact of collective bargaining coverage on wages and working conditions

The impact of collective bargaining coverage on an individual worker's wages and working conditions is primarily studied by means of survey data. The most studied impact of collective bargaining coverage addresses wages: does a wage premium exist for collective bargaining coverage? Empirical studies regarding the impact of collective bargaining typically explore wage differentials for a binary variable, a worker is either covered or not covered, occasionally taking industry dummies as a proxy for the variation in collective agreements. For example, using a large, matched employer-employee dataset Magda et al. (2012) explored the impact of company and industry bargaining agreements on wages in the Czech Republic, Hungary and Poland, finding that industry agreements increase wages for low-skilled workers and that company agreements increase medium- and high-skilled wages. Blien et al. (2011) did so over time for Western Germany, whereas Heinbach and Schropfer (2007) concluded that the impact of opening clauses in German CBAs on employment levels in collective negotiations affected wage bargaining as well as non-wage issues. OECD (2018) shows that in OECD countries workers had higher wages with firm-level bargaining, while no effect was found for workers covered by sector bargaining. OECD used micro-level data from the Structure of Earnings Survey and other data sources, and they controlled for gender, age, educational attainment, industry, occupation, firm size, type of contract and job tenure, but not for the prevalence of wage scales or the structure of the wage scales. Using Belgian linked employer-employee panel data, Garnero et al. (2020) concluded that firm-level CBAs benefit both employers and employees through higher productivity and wages, without being very detrimental to firms' performance. However, except for a few single-country studies the width and depth of these issues has remained relatively unexplored.

Many studies have used cross-sectional survey data, but this does not allow to control for companies' self-selection into collective bargaining. The question remains whether

companies with, for example, a predominantly low-skilled or with a predominantly high-skilled workforce engage in firm-level collective bargaining, thus having an effect on the wage outcomes of collective bargaining. This was partly solved in Addison et al. (2014) by using establishment-level wage data from Germany. The authors explored the CBA wage premium as to find that average wages increased by 3 to 3.5 percent for establishments entering into collective bargaining whereas they did not had a CBA before, and decreased by 3 to 4 percent after abandoning collective bargaining.

Survey data has also been used to explore the impact of bargaining coverage on other outcomes than wages. Taiji and Mills (2020) used the European Social Survey to conclude that at country level the degree to which workers are covered by CBAs showed up as the strongest factor shaping the social consequences of non-standard schedules. Paolucci (2017) concluded that the emphasis on flexibility versus security clauses in CBAs in the chemical and pharmaceutical sector in Italy and Denmark were strongly related to firm-specific characteristics. Using web-survey data about the impact of the economic crisis during the early 2010s on companies' wage or workforce adjustments in Germany and the Netherlands, Tjijdens et al. (2014) found that employees in crisis-hit organisations reported wage adjustments less often and workforce adjustments more often when they were covered by a collective agreement, compared to those who were not covered.

2.3 The impact of CBA clauses on wages and working conditions

Collective bargaining is presumably pivotal for wage setting and working conditions, but – as said – little is known about what exactly is agreed about these topics. Two methods are available to explore the content of CBAs. The first one is interviewing bargaining actors about the issues concluded in the CBA they are responsible for. Alternatively, national correspondents could be asked about the content of the CBAs in their country, which requires them to interview the bargaining actors. The second method is collecting and coding the CBA texts. Only few EU countries code agreements. This section will review the research findings.

More than 20 years ago, Dunn and Wright (1994) examined the contents of approximately 100 CBAs in the United Kingdom and found that clauses on flexible working practices demonstrated greater formalization in some agreements while shifting towards more general managerial prerogative clauses in others. Using country reports provided to the ILO, Hayter and Stoevska (2010) reviewed the issues covered by collective agreements. They

observed that the bargaining agenda has expanded in many parts of the world, now including issues such as work organization, vocational training, the regularization of employment, parental leave and family responsibilities. Using a qualitative approach based on reports from national correspondents, Olgiati and Shapiro (2002) showed that gender equality issues are poorly represented in collective agreements in seven EU countries, while Jacob (2013) found a weakening of employment protection clauses in the Chicago Public Schools' bargaining agreement on teacher absenteeism. Based on a Swiss and a French case study, Bonvin et al. (2013) found that the clauses in the CBAs in the firms under study did not enhance workers' voice in restructuring processes. Based on the Australian Workplace Agreements Database, Baird and Murray (2014) concluded that between 2005 and 2010 increasing media debates about parental leave were positively related to an increase in CBA clauses about paid parental leave. For Greece, Nicolitsas (2020) used detailed CBA information from a sectoral database for the metal sector and enriched it with business statistics. The author concluded that during the recent economic crisis in Greece negotiated wages appeared binding for subsectors facing less product market competition, whereas the ability to opt out of the sectoral agreement and thereafter signing firm-level agreements had mainly been used by firms suffering accounting losses. For three states in the United States, namely California, Michigan, and Washington, Strunk et al. (2018) investigated a set of 43 key provisions, comparing more than 1,000 teacher CBAs. They concluded that CBAs vary substantially within and across states and that this variation is more associated with district size than with the proportion of low-income students within the districts. Using coded data from the FNV Database of CBAs in the Netherlands, Yerkes and Tjeldens (2010) concluded that between 1995 and 2008 CBA clauses largely compensated for the declining welfare state coverage with respect to disability and work-life arrangements. Using the WageIndicator CBA Database, a comparison of the content of 249 collective agreements from 11 developing countries revealed that although the reference to wage setting was an integral part of almost all collective agreements, CBAs vary largely with respect to the detail and they did not commonly include pay scales, leaving the determination of exact wages up to individual contracts (Besamusca and Tjeldens, 2015). Working hours, paid annual leave days and paid maternity leave are core issues and if agreements contain one of these three clauses they also more likely to include the other two. In summary, the studies discussed here typically focus on a small number of clauses, which were coded manually from the relevant CBAs.

This brings us to the next question: which topics are bargained? For claims that about the bargaining agenda in collective agreements one would expect to find overviews regarding

the topics included in CBAs. However, we found only few of such overviews. The government in Estonia applies a unique present/absent coding scheme for all CBAs concluded, though the list of topics varies across CBAs.⁴ Schulten (2018, 78) provided a table based on information from the Ministry of Labour in Germany regarding 13 broad topics negotiated in extended collective agreements, namely: general framework agreements; wages and salaries; additional pensions; capital-forming benefits; holidays and holiday bonuses; minimum wages; wage structure; apprenticeship pay; training; annual bonuses; dismissal/job protection; working time, and a category 'other'. According to this study, around 80 per cent of all extended collective agreements in Germany covered issues other than basic wages and more than one fifth concerned general framework agreements (*Manteltarifverträge*), which typically included aspects of working and employment conditions such as pay structure, working time, holidays, and social contributions. In addition, a relatively large number of extended agreements covered additional pensions, capital-forming benefits, holidays and holiday bonuses. A few extended agreements regulated annual bonuses, wage structures, training, employment protection and working time. We did not come across empirical studies that traced changes over time in bargaining topics.

An inventory of industrial relations of the five largest companies in five major industries, namely: metal and electronics manufacturing, wholesale, retail, ICT, and transport and telecommunications, in 23 EU countries provided information about the topics covered in the CBAs of these 575 companies (Van Klaveren and Gregory, 2019). If they were covered by a collective agreement these companies were asked about the topics covered, resulting in information about 181 CBAs. Most CBAs were found in metal and electronics manufacturing and the lowest number in the ICT industry. Ten CBA topics were coded according to a simple present/absent coding scheme, indicating that almost all agreements had clauses on wages (97%). Yet, only two-thirds dealt with wage *increases* and not wage *levels* (65%). Almost nine in ten agreements contained clauses on working hours, schedules and holidays (88%). About half to three quarters of the CBAs included clauses on sickness and disability (76%); social security (72%); training (69%); work - family arrangements (68%); medical assistance (65%); job security (63%), and internal mobility (50%). Fewer agreements

⁴ See <https://klak.sm.ee/index.html>

had clauses on work organisation (38%) and relatively few contained clauses on agreed workforce numbers (8%). Multi-employer CBAs more often included any of the ten topics compared to single-employer CBAs, apart from those on work organisation. For a European-wide analysis of the agreed wages and working conditions in CBA clauses, CBA texts should be coded, as will be discussed in the next section.

3 Overview of CBA registries in Europe

3.1 Registries of CBAs

Our inventory of online CBA registries in Table 1 is based on desk research in early 2021 and complemented with information from country experts. To be declared legally binding, either for the enterprises covered or for all enterprises in the industry, in most EU countries CBAs need to be registered with the government or with an institution acting on its behalf.

Registration typically aims to verify whether the bargaining actors are eligible and whether general and specific validity criteria are met. In some countries such a registries are maintained by Ministries of Labour, Employment, Social Affairs, Economics or similar. In other countries an agency has been assigned this task, such as the Chamber of Labour in Austria or the National Council for Economics and Labour (CNEL) in Italy. In a third group of countries the social partners maintain a registry of CBAs, like in Germany the *WSI Tarifarchiv* on behalf of the main trade union movement. More than one archiving method may be identified per country, such as in the Netherlands, where the organisation servicing the employers' association and the largest trade union confederation both maintain databases, as does the Ministry of Social Affairs and Employment.

In our inventory, we found no evidence that CBAs need to be registered to become binding in four of 28 EU member states (still including the UK), namely in Croatia, Denmark, Latvia, United Kingdom (Column 2 in Table 1). In Czech Republic and Slovakia, only the higher-level (sectoral or multi-employer) CBAs have to be registered. In almost all countries where registration is required, the government or the institution acting on its behalf maintains a registry in which names, signatories, duration and features of extended CBAs are registered. This is the case in 22 countries; in another two countries the social partners or a research institute maintain a registry. This last group includes Denmark and the UK (Column 4 in Table 1).

In previous decades governments mostly announced the agreements declared binding in the Government Gazette, but increasingly they do so online. Currently, sixteen EU countries publish the full texts of all CBAs or the sector CBAs online, mostly in PDF format, namely, Austria, Belgium, Cyprus, Czech Republic, Estonia, Finland, France, Germany, Greece, Italy, Lithuania, the Netherlands, Slovakia, Spain, Sweden, and the UK. Two countries, Germany

and the UK publish CBA texts online accessible for members only (Column 5 in Table 1). Registries in six countries do not publish CBAs online whereas for six countries we could not find this data.

We found that CBAs can be archived in different ways. Some registries archive the metadata only, others archive full texts of CBAs, and a third group registers the changes agreed in each subsequent CBA. The latter is the case in the Registry of the Directorate-General for Collective Labour Relations in Belgium and in the archive of the Dutch organisation servicing the country's employers' association.

Table 1 CBA registration required, registry agency, archive maintained, full text downloadable, coded CBA by country

| 1 Cntr | 2 Registration required | 3 Registration by | 4 Archive | 5 Full text online | 6 Coded |
|--------|-------------------------|---|-----------|--------------------|--------------|
| AT | Yes | Chamber of Labour | Yes | Yes | Yes |
| BE | Yes | Directorate-General for Collective Labour Relations | Yes | Yes | Yes |
| BG | Yes | Ministry of Labour and Social Policy | Yes | No | No |
| CRO | DK | No evidence | No | No | No |
| CY | Yes | Department of Labour Relations of the Government | Yes | Yes | No |
| CZ | Yes, for sector CBAs | Ministry of Labour and Social Affairs | Yes | Yes | No |
| DE | Yes | BMAS-Tarifregister of WSI-Hans Boeckler Foundation | Yes | Yes (only members) | Yes |
| DK | No | Confederation of Danish Employers DA | Yes | No | Yes |
| EE | Yes | Ministry of Labour | Yes | Yes | Yes |
| ES | Yes | Ministry of Employment and Social Economy | Yes | Yes | No |
| FI | Yes | Ministry of Justice / Trade Unions | Yes | Yes | Yes |
| FR | Yes | National Collective Bargaining Commission of the Ministry of Labour | Yes | Yes | Yes |
| GR | Yes | Ministry of Labour and Social Security | Yes | Yes | No |
| HU | Yes | Ministry of Innovation and Technology | Yes | No | No |
| IE | Yes | Labour Court | Yes | DK | No |
| IT | Yes | National Council of Economy and Labour (CNEL) | Yes | Yes | Periodically |
| LT | Yes | Ministry of Social Security and Labour | Yes | Yes | No |
| LU | Yes | Ministry of Labour | Yes | DK | DK |
| LV | DK | No evidence | DK | DK | DK |
| MT | Yes | Department for Industrial and Employment Relations | No | No | No |
| NL | Yes | Ministry of Social Affairs and Employment | Yes (3) | Yes | Yes |
| PL | Yes | Ministry of Development and Labour | Yes | No | No |
| PT | Yes | Ministry of Labour | Yes | DK | DK |

| 1 Cntr | 2 Registration required | 3 Registration by | 4 Archive | 5 Full text online | 6 Coded |
|--------|-------------------------|--|--------------------------|--------------------|---------|
| RO | Yes | Appropriate authorities | DK | DK | DK |
| SE | Yes | National Mediation Office | Yes (trade union) | Yes | No |
| SI | Yes | Ministry of Labour, Family, Social Affairs and Equal Opportunities | Yes | DK | DK |
| SK | Yes, for sector CBAs | Ministry of Labour | Yes | Yes | No |
| UK | No | Labour Research Department | Yes (research institute) | Yes (only members) | Yes |

Source: Websites accessed 12/13 JAN 2021, DK = Don't know

3.2 Coding of CBAs

This section explores whether the CBA's metadata, full texts and number of workers covered are included in the registries.

Metadata includes the name(s) of the company or the sector covered, the names of the employers' and employees' signatories, the operative date and the duration or the end date if agreed. Typically, this is registered in all countries with a registry.

Full text coding can be done by identifying whether keywords are present, by identifying the headings in the full text, by the annotation of relevant pieces in the full texts, by coding the annotated texts, and by registering the wage increases. Estonia provides a good example of keyword identification and comparing the clause to the law (Figure 1). The country also registers the number of employees covered by the CBA. Austria provides a fine example of identification of the headings in the CBA text, with clicks to the content (Figure 2). In the Netherlands, the FNV union confederation applies full-text coding for more than 500 variables; the organisation servicing the employers' association AWWN codes the changes in CBAs while the Ministry of Social Affairs is coding CBAs for specific research objectives, which vary from year to year. Column 6 in Table 1 shows that in nine EU countries some form of coding has been applied, though coding schemes vary across registries.

Figure 1 Coding of a collective agreement in Estonia, by the Ministry of Social Affairs (CBA of the Tallinn University of Technology)

| Additional information form | | Regulated in the contract | More favorable than the law |
|--|--|---|---|
| 1. Agreement on the terms of the individual employment relationship? | | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No |
| 1.1. Agreement on salary conditions? | | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No |
| 1.1.1. employee remuneration agreement; | | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No |
| 1.1.2. an agreement on the periodic increase of the employee's remuneration; | | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No |
| 1.1.3. overtime pay agreement (§ 44 subsections 6-7 of the TLS); | | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No |
| 1.1.4. on-call time supplement agreement (§ 48 (1) TLS); | | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No |
| 1.1.5. night work bonus agreement (§ 45 (1) TLS); | | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No |
| 1.1.6. public holiday bonus agreement (§ 45 (2) TLS); | | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No |
| 1.1.7. an agreement on the payment of a performance bonus (more efficient work, good economic performance of the company, etc.); | | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No |
| 1.1.8. an agreement on the payment of non-work-related remuneration (Christmas bonus, marriage and child benefits, funeral allowance, etc.); | | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No |
| 1.1.9. agreement on other terms of remuneration. | | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No |
| 1.2. Agreement on holiday conditions? | | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No |
| 1.2.1. annual leave agreement (§ 55 TLS); | | <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No | <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No |

Source: Screenshot from <https://klak.sm.ee/leping/1777.html>, accessed 20210106

Figure 2 Coding of a collective agreement in Austria, by the Chamber of Commerce (CBA of the Landes-Hypothekbanken)

| Kollektivvertrag | |
|--|--|
| für die Angestellten der österreichischen Landes-Hypothekbanken vom 18.10.2010 (in Kraft getreten am 1.1.2011) in der Fassung vom 1.4.2020 | |
| abgeschlossen zwischen dem VERBAND DER ÖSTERREICHISCHEN LANDES-HYPOTHEKENBANKEN 1040 Wien, Brucknerstraße 8 | ÖSTERREICHISCHEN GEWERKSCHAFTSBUND Gewerkschaft der Privatangestellten, Druck, Journalismus, Papier Wirtschaftsbereich Volks-, Hypobanken / Raiffeisen 1034 Wien, Alfred Dallinger Platz 1 |
| und dem | |
| Inhaltsverzeichnis | |
| Abschnitt 1 Allgemeine Bestimmungen | |
| § 1 Personenbezogene Bezeichnungen | |
| § 2 In-Kraft-Treten und Aufkündigung des Kollektivvertrages | |
| § 3 Räumlicher, fachlicher und persönlicher Geltungsbereich | |
| Abschnitt II Gehaltsregelung | |
| § 4 Einreihung, Einstufung und Anrechnung | |
| § 5 Beschäftigungsgruppen | |
| § 6 Gehaltsstufen | |

Source Screenshot from https://www.wko.at/service/kollektivvertrag/kollektivvertrag-landes-hypothekbanken-2020.html#heading_1, accessed 20210116

The coding of agreed wages requires special attention. In spite of the assumption that collective bargaining aims at setting wages and working hours, the BARCOM project found that only one-third of 108 commerce CBAs in EU28 included details about wage levels; agreements about wage increases as such were reported more often (Besamusca et al. 2018). For the 602 CBAs in the COLBAR-EUROPE project, Besamusca (2021) found that wages are to be set in individual contracts (28% of these CBAs), at the company level (24%), at the

sector level (40%), or they wage-setting is not specified (9%). For cross-country comparisons of agreed wages, the coding of wage levels and wage increases requires a well-defined system to convert wage tables in CBAs into hourly wages.

Only few registries provide information about the number of workers covered by CBAs. The registry in Estonia does, but the registry in the Netherlands does not.⁵ Based on experience with the WageIndicator CBA Database, the number of workers covered is one of the variables most difficult to be traced.

Outside the EU, two countries are worth mentioning here. In Australia, the Attorney-General's Department maintains the Workplace Agreements Database. Since 1991 this database provides information about developments in coverage, wage increases and conditions of employment included in CBAs and currently includes 160,000 agreements.⁶ In New Zealand, the Centre for Labour, Employment, and Work of the Victoria University of Wellington (CLEW) maintains a full text and coded database of the collective agreements. Since 1992, the Centre has collected employment agreements from the country's employers and trade unions.⁷

3.3 How many CBAs are there in Europe?

The question how many CBAs are concluded in EU countries is difficult to answer. First, the stock of CBAs is volatile, because of entries and exits due to mergers, bankruptcies, or removals, or because the uncertain status of non-renewed CBAs. If not renewed, CBAs may lose their force immediately or they continue to have force unchanged. Of course, the latter practice hampers CBA counting. Second, the nested nature of collective bargaining challenges CBA counting. Third, if single-employer agreements do not need to be registered and the signatories do not distribute the CBA in question beyond the company, CBA counting will underestimate the real number of CBAs.

For the CBA counting in EU28 we explored relevant national websites, using two sources of information, namely the ETUI website about National Industrial Relations, based on Fulton

⁵ See <https://www.uitvoeringarbeidsvoorwaardenwetgeving.nl/mozard/!suite16.scherm1168?mGmr=66>

⁶ See <https://www.ag.gov.au/industrial-relations/enterprise-agreements-data/workplace-agreements-database>

⁷ See <https://www.wgtn.ac.nz/clew/research/services>

(2020), and the country profiles on the Eurofound website.⁸ Based on this information the number of multi-employer and single-employer agreements is estimated at 185,382 CBAs, of which Germany and France jointly make up 85%. Another six countries contribute between 1,000 and 8,000 CBAs each (Czech Republic, Hungary, Latvia, Romania, Spain, UK), and the remaining 19 countries less than 1,000 CBAs each (Table 2 Column C).

Table 2 Estimated number of collective agreements, if possible broken down by Multi- and Single-employer agreements

| Cntr | # CBAs | Estimate # | Of which #MEB | Of which #SEB | Year |
|------|---|------------|--|-----------------------------------|------|
| AT | 800+ (yearly 450+ collective agreements are negotiated) | 800 | large majority | few | 2020 |
| BE | NA | 100 | 100 joint committees and 64 joint sub-committees | increasing | 2020 |
| BG | 914 | 914 | 3 | 911 | 2019 |
| CRO | 570 | 570 | 16 | 554 | 2014 |
| CY | 237 renewed in 2018 | 237 | NA | NA | 2018 |
| CZ | NA | 3770 | NA | 3,770, signed by ČMKOS affiliates | 2018 |
| DE | 77,316 | 77316 | 29,136 | 48,180 | 2018 |
| DK | NA | | NA | NA | NA |
| EE | 677 | 677 | NA | NA | 2020 |
| ES | 2,677 | 2677 | 526 | 2,151 | 2020 |
| FI | 290 | 290 | 198 | NA | 2014 |
| FR | NA | 82780 | 200 | 80,780 | 2019 |
| GR | 234 | 234 | | | 2020 |
| HU | approx 2,800, most unlimited | 2800 | NA | 2,800 | 2020 |
| IE | 3 | 3 | 0 | 3 | 2021 |
| IT | 856 | 856 | 856 | | 2020 |
| LT | 419 | 419 | NA | NA | 2020 |
| LU | 75 | 75 | 6 | 69 | 2011 |
| LV | 1,152 | 1152 | almost none | almost all | 2016 |
| MT | 34 | 34 | NA | 34 | 2019 |
| NL | 635 approx | 635 | 500 approx | 135 | 2020 |
| PL | 220 new agreements + 386 additional protocols | 220 | 46 | 174 | 2020 |
| PT | 323 | 323 | NA | NA | 2019 |
| RO | 8,233 | 8233 | 0 | 8,233 | 2019 |
| SE | 194 | 194 | NA | NA | 2020 |
| SI | 45 | 45 | 45 | NA | 2020 |
| SK | NA | 28 | 28 | NA | 2012 |
| UK | 2000 | 2000 | 0 | 2000 | 2020 |
| TOT | | 185382 | | | |

Source: Own deskresearch, based on national websites, NA = Not Available

⁸ See <http://www.worker-participation.eu/National-Industrial-Relations/Countries> respectively <https://www.eurofound.europa.eu/country>

3.4 Feasibility of EU-wide CBA Database

This paper aims to explore the feasibility of an EU-wide CBA Database. From the previous sections it has become clear that such a database certainly cannot include all EU countries, because countries do not have an online CBA registry or because their registry is accessible for members only. Table 3 shows that for eight countries all CBAs and for three countries all sectoral CBAs could be included in such a database. For ten countries some but not all CBAs could be included because these CBAs have to be collected from different websites, via national correspondents, or by directly contacting CB actors. For seven countries hardly any CBAs could be included because a registry is lacking or only accessible for members only, or because very few CBAs have been concluded in the country. Table 3 shows that for eight countries all CBAs and for three countries all sectoral CBAs could be included in such a database. For ten countries some but not all CBAs could be included because these CBAs have to be collected from different websites, via national correspondents, or by directly contacting CB actors. For seven countries hardly any CBAs could be included because a registry is lacking or only accessible for members only, or because very few CBAs have been concluded in the country. Table 3 shows that for eight countries all CBAs and for three countries all sectoral CBAs could be included in such a database. For ten countries some but not all CBAs could be included because these CBAs have to be collected from different websites, via national correspondents, or by directly contacting CB actors. For seven countries hardly any CBAs could be included because a registry is lacking or only accessible for members only, or because very few CBAs have been concluded in the country. Table 3 shows that for eight countries all CBAs and for three countries all sectoral CBAs could be included in such a database. For ten countries some but not all CBAs could be included because these CBAs have to be collected from different websites, via national correspondents, or by directly contacting CB actors. For seven countries hardly any CBAs could be included because a registry is lacking or only accessible for members only, or because very few CBAs have been concluded in the country.

Table 3 Feasibility of including the country's CBAs in a EU-wide CBA Database, January 2021

| Country | CBAs included | Country | CBAs included |
|----------------|-------------------|----------------|-------------------|
| Austria | all | Ireland | few |
| Belgium | all | Italy | all sectoral CBAs |
| Bulgaria | few | Lithuania | few |
| Croatia | some | Luxembourg | few |
| Cyprus | few | Latvia | few |
| Czech Republic | all sectoral CBAs | Malta | some |
| germany | few | Netherlands | all |
| Denmark | some | Poland | some |
| Estonia | all | Portugal | some |
| Spain | all | Romania | some |
| Finland | some | Sweden | all |
| France | all | Slovenia | some |
| Greece | all | Slovakia | all sectoral CBAs |
| Hungary | some | United Kingdom | some |

3.5 Feasibility of representative sampling of CBAs

If an exhaustive EU-wide CBA Database is not within reach, sampling CBAs might be a possibility. If so a sampling frame needs to be identified. Here we will discuss the possibilities for four sampling strategies, including their drawbacks.

A representative random or stratified sample from the stock of CBAs assumes lists of all CBAs and their number of covered employees. As shown in the previous section, this option is available for a limited number of EU countries. Additionally, a random or stratified sampling of the CBA universe requires well-developed views about the sampling of multi-employer versus single-employer CBAs, nested agreements, separate agreements for manual and non-manual workers, and annexes to agreements. Finally, this strategy needs to take into account that the stock of CBAs is volatile in view of the many entries and exits due to mergers, bankruptcies, or removals. This includes expired and not renewed CBAs that possibly continue to be valid. Having actively participated in CBA name identification for quite some years in the Netherlands, the first author estimates that on an annual basis almost 10% of all agreements 'enter' or 'exit' the universe.

A second sampling strategy proposes employees as the unit of analyses by sampling the dependent labour force, as in the Labour Force Surveys. These surveys need to include a question 'are you covered by a collective agreement'. For merging the survey data with CBA data, the name of the employer should be known for countries with predominantly single-employer bargaining; the sector should be known for countries with predominantly multi-employer or sectoral bargaining. The former might be hampered by privacy concerns of the survey holders and the latter by the fact that the sector demarcations in the NACE industry

classification and those in CBAs in many cases do not overlap, requiring a domain identification connected with the NACE coding. Alternatively, the survey could ask covered respondents to provide an identifiable CBA name assuming that they are willing to disclose the name of their employer.

A third sampling strategy proposes employers as the unit of analysis by sampling companies through company surveys.⁹ The sampling of such surveys is typically stratified by company size. The employer could be questioned about collective bargaining coverage as for example in Eurofound's European Company Survey (ECS). If covered, the employer could be asked to hand in a copy of the CBA or to answer a limited set of questions about the content of the CBA. In case of a multi-employer or sectoral CBA the respondent must identify the domain of the CBA following the NACE classification. This method has been applied in the WIBAR3 survey of companies and respondents find it relatively easy to answer (Van Klaveren and Gregory, 2019).

A fourth sampling strategy proposes bargaining actors as the unit of analysis by sampling these actors and asking details about the CBAs they have concluded. This requires a list of actors to be collected through their employers' organisations respectively trade unions or professional organisations. Such lists do not exist and would therefore need to be built.

⁹ Where we use 'company', we also refer to public sector organisations.

4 Technical feasibility of a Europe-wide CBA Database

Three technical components are essential for initiating a Europe-wide data collection of collective agreements. First, CBA texts need to be gathered and prepared for a coding tool, which might be challenging with an estimated 183,841 CBAs. Second, the possibilities of drafting representative samples of these CBAs should be explored. Third, coding CBA texts is essential for understanding of CBA clauses, but coding is manual work. The coding work for a Europe-wide CBA data collection should therefore explore the possibilities of machine reading. In this section we will use our experiences with the global WageIndicator CBA Database for assessing the feasibility of a Europe-wide CBA data collection.

4.1 The WageIndicator CBA Database

WageIndicator Foundation is a non-profit NGO, that develops, operates and owns national WageIndicator websites with labour-related content. The first WageIndicator website was launched in the Netherlands in 2001. Today, it is operational in over 190 countries and yearly receives millions of visitors (40 million in 2020). WageIndicator's mission is to promote labour market transparency for the benefit of employers, employees and workers worldwide by sharing and comparing information on wages, labour law and careers. WageIndicator does so by making this information freely available on easy-to-reach-and-read national websites in the national language(s), applying sophisticated search engine optimization. The contents of the websites are derived from data collected through a Collective Agreement Database, a Labour Law Database and a related DecentWorkCheck survey, a Salary and Working Conditions Survey and a related Salary Check, a Minimum Wages Database, and a Cost-of-Living Survey with a related Living Wages Database.¹⁰

For its CBA Database, WageIndicator has developed a central web-based platform to upload, annotate and code CBA texts, using a predefined coding scheme. The CBA Database aims to enrich the content of the national websites and allows to browse agreements online, though it also provides a unique opportunity to closely examine the variation within and across agreements. The CBA texts are published online on the national websites, unless the signatories refuse so. The coding scheme allows to compare agreements for the coded topics

¹⁰ See www.wageindicator.org

within and across countries on WageIndicator's home page.¹¹ The first CBA was entered in December 2013 and 1,594 agreements from 62 countries were included by December 2020. The database allows for statistical analysis and has been used in EU projects and for analyses of CBAs in the garment industry in Indonesia and in middle- and low-income countries (Besamusca and Tijdens, 2015).

In conclusion, the WageIndicator CBA Database shows that a Europe-wide CBA Database is technically feasible, but that the gathering, uploading and coding of more than hundred thousand CBAs will require substantial resources. Hence, a reduction of the number of countries or the number of CBAs in the Database should be considered.

4.2 Gathering CBA texts

WageIndicator employs three approaches to collect CBAs: downloading from national CB registries, through Internet search with smart Googling, and asking bargaining actors for full-text agreements. Table 3 shows that online gathering of all CBAs is possible for eight EU-countries and for all sector CBAs in Italy and France. Online gathering of some but not all CBAs is feasible for 13 countries, including the UK, as some bargaining actors post their CBAs online in these countries. For 5 countries gathering CBAs seems too challenging because they are not available online and thus bargaining actors need to be contacted, because CBAs are available for members only. CBA collection by asking bargaining actors is problematic when these actors are not willing to share their CBAs because of competitive reasons, as we've noticed for Poland and Hungary, or because they are available for members only as in the UK and Germany. A final problem arises for countries with very few CBAs, such as Ireland, or for countries where the CBA registration consists of addendums only, such as Belgium. With CBAs more often posted online, gathering CBAs has become more convenient. Countries with the lowest coverage rates happen to be those countries posting their agreements least online. We assume that the more relevant CBAs are as a wage setting instrument, the more likely we are able to collect the texts online.

Gathered CBAs can be in any format, be it in Word or PDF format, but also in JPEG format or even as a printed booklet. These formats need to be converted into text format so that

¹¹ See <https://wageindicator.org/labour-laws/collective-bargaining-agreements/collective-agreement-database-visual>

they can be uploaded in a database. Word files can be copied to text format, using Notepad; PDF files can be converted to text format using optical character recognition software. If the CBA includes tables - and many do - additional controls are needed, because the software cannot convert tables while maintaining sufficient quality. JPEG files need to be retyped. Booklets can be copied into a PDF file. Once a CBA is formatted as text, WageIndicator uses Amaya software to assign headings. An overall heading is typically defined as H1, chapters as H2, and articles within chapters as H3. Once this is done, the text can be uploaded in HTML format in the WageIndicator CBA Database. Depending on the size and complexity of the CBA text, uploading may take between 30 minutes up to a couple of hours per CBA. The complexity of CBAs relates to their size and whether tables are included. Once texts are uploaded in the platform, they can be annotated by selecting the part of the text or the clause where the answer to the question in the predefined coding scheme can be found. The great advantage of a web-based coding tool is that annotators can be located anywhere in the world and still enter their CBAs in one coding platform.

4.3 Coding CBA texts

Coding CBA texts is a condition for any statistical analysis of CBAs. To do so, a coding scheme is needed for the coding of the CBA's metadata and content. The metadata refers to the affiliations of the signatories: employers or their associations, trade unions, and if relevant works councils or professional associations. Metadata refers to the CBA's status as a single-employer or a multi-employer CBA, a framework agreement, an appendix or a transnational agreement. It requires coding whether a ratification process is applicable, whether the CBA is extended to employers not concluding the agreement. In case of a multi-employer CBA it requires the identification of the sector boundaries. Operative date as well as the duration, if agreed, need to be coded, and so is the inclusion or exclusion of certain groups of workers. Preferably, metadata should also include the number of employees covered, although the WageIndicator CBA Database shows that these numbers are typically not mentioned in the CBA text itself. Some registries require CB actors report the number of employees covered. Otherwise, this information has to be collected from negotiators, which sometimes might be hard.

To register the signatories the WageIndicator CBA Database developed a pick list for signatories from the employers' side and one for signatories from the employees' side. In the annotators platform, these lists take the form of an Application Programming Interface

(API), stored on a server. In the coding form an autosuggest box is used for identification of the signatories. After typing in a few characters, the API shows a list of matches and the annotator can select the right name or enter a new name. If the signatory is not in the list, it has to be added.

For the CBA content WageIndicator has developed a coding scheme covering the topics assumed to be most common for collective agreements around the world. The coding scheme addresses ten topics (Table 4). Each topic starts with a Yes/No question: Does the CBA include any clauses on this topic? If so, the coding scheme follows with detailed questions. If not, the coding scheme goes to the next topic. The coding scheme of the WageIndicator CBA Database generates 740 variables (Ceccon et al., 2017).

Table 4 Ten topics in the coding scheme of the WageIndicator CBA Database

| Nr | Topic |
|----|---|
| 1 | Job titles |
| 2 | Wages |
| 3 | Working Hours, Schedules, Paid Leaves And Paid Holidays |
| 4 | Employment Contracts |
| 5 | Work and Family Arrangements |
| 6 | Health and Safety and Medical Assistance |
| 7 | Sickness and Disability |
| 8 | Social Security and Pensions |
| 9 | Training |
| 10 | Gender Equality Issues |

WageIndicator started with annotators from many countries to code the CBAs in their language. However, it turned out that CBA coding requires skilled and experienced annotators rather than native annotators not familiar with the coding tool. WageIndicator now has skilled annotators managing many languages. In case CBAs in another language need to be coded, the coding team uses Google translate, that has reached sufficient quality for the purpose of coding the texts. In case a substantial number of CBAs in another language needs to be coded, a native annotator is trained to do so. For this purpose WageIndicator has developed a training kit with videos and instructions available.

4.4 Machine-reading technologies

Coding CBAs requires an annotator has to read the full CBA text to find where a topic is addressed, and to do so again for a next topic. When texts are very long annotation is really time consuming Thanks to the SSHOC project (EU-H2020 nr 823782) keyword extraction for CBAs could be explored (Ceccon and Kaandorp 2019; Ceccon and Ceccon 2020). This allows

annotators to look for information in the text area where keywords are identified and exclude the rest of the text. To identify the most common used keywords, the annotated clauses in the CBA Database were collected for each language with at least 30 agreements.

To identify the keywords most typical for the annotated clause, a Python script has been created using a Natural Language Toolkit library, available for many languages. The Toolkit allows for lemmatization to accurately identify the lemma for each word and it applies stopword removal. Per topic the lemmatized clauses from the CBA Database are randomly split into a training set (60% of the clauses) and a testing set (40%). A script is run to extract the most corresponding words across the clauses, producing a list of keywords. These keywords are tested as to produce a reliability score. With a satisfactory score, the keyword list can be used for a newly added CBAs to identify if this text includes clauses relevant to the coding variable. This approach allows to generate an absent/present table of clauses. When greater coding detail is needed, such as the number of weeks of maternity leave, the keyword system identifies instantly in which part of the text the relevant clause can be found.

WageIndicator is currently trying to understand the minimum number of CBAs and clauses needed for a satisfactory reliability score and which number of keywords allows for increasing reliability (Ceccon and Ceccon 2020). The findings so far indicate that at least 206 clauses (6-7 CBAs) are needed to produce some results. The number of words can be customized in the script each time for a different language. Five words are enough for languages with 67-203 CBAs (1751-6116 clauses), but for English with 401 CBAs (10135 clauses) a minimum of 15 words is needed. At least 10 words are needed for languages with 7 to 11 CBAs (206 to 928 clauses).

5 Conclusions

In their renowned study “What do unions do?”, Freeman and Medoff (1984) argue that trade unions bargain for higher wages, equal pay and fair working conditions, but little is known about the bargaining outcomes agreed in collective bargaining agreements (CBAs) in the EU. In view of the re-launched dialogue with social partners at European level, such data is critical for monitoring progress in wage-setting and working conditions. Yet knowledge about what exactly is concluded in collective bargaining remains a blind spot. This article aimed to explore the feasibility of collecting and coding CBA texts in Europe to tackle this blind spot in the body of knowledge in industrial relations research, knowing that Internet may accelerate data collection. This article is based on desk research, experience with three EU-funded projects to explore the content of CBAs and the WageIndicator CBA Database.

The review of research showed that most knowledge about the impact of collective bargaining on wages and working conditions is based on survey data where a binary variable for coverage versus non coverage is used. Research using coded information about the content of CBAs revealed a fragmented picture with studies that coded the content of CBAs for the purpose of the study, not for tracing changes in CBAs over time.

The overview of CBA registries in the European Union showed that in almost all Member States CBAs need to be registered with the government or with an institution acting on its behalf to become binding. More than half of the EU countries maintain a registry of extended CBAs and most of these countries post the registry online. Our study estimated the number of sector and company CBAs in EU28 in 2019/20 at 185,382 CBAs. In exploring the feasibility of an EU-wide CBA Database it became clear that such a database could include all or sectoral CBAs only from 11 countries and that for another 17 countries some but not all CBAs could be included.

The building blocks of an EU-wide CBA registry-in-the-making should specify the requirements for gathering, coding and annotating CBAs. The WageIndicator CBA Database can be taken as an example, as are the options for machine-reading of CBA texts.

Three technical components are essential for a Europe-wide data collection of collective agreements. The CBA texts need to be gathered and prepared for a coding tool, the possibilities of drafting representative samples for countries with many CBAs need to be explored, and coding of CBA texts is essential for understanding of CBA clauses. The WageIndicator CBA Database was populated with annotated and coded CBA texts in

consecutive EU projects. Its coding scheme addresses ten topics to capture the core of each CBA. WageIndicator has explored machine reading options and concludes that this is a promising way forward to reduce the time needed for coding and thus allowing for the annotation and coding of more CBAs. WageIndicator is currently trying to understand the minimum number of CBAs and clauses needed for a satisfactory reliability score and which number of keywords allows for increasing reliability.

6 References

- Addison J, Teixeira P, Evers K and Bellmann L (2014) Indicative and Updated Estimates of the Collective Bargaining Premium in Germany. *Industrial Relations: A Journal of Economy and Society* 53(1): 125–156. <https://doi.org/10.1111/irel.12049>
- Askitas N and Zimmermann, KF (2015) The internet as a data source for advancement in social sciences. *International Journal of Manpower* 36(1): 2-12. <https://doi.org/10.1108/IJM-02-2015-0029>
- Baird M and Murray J (2014) Collective bargaining for paid parental leave in Australia 2005–2010: A complex context effect. *The Economic and Labour Relations Review* 25(1): 47-62. <https://doi.org/10.1177/1035304614522566>
- Besamusca J, Kahancová M, Van Klaveren M, Kostolný J, Szüdi G and Tijdens K (2018) *BARCOM REPORT 3: Bargaining Systems and Collective Bargaining Agreements in the Commerce Sector*. Amsterdam etc.: University of Amsterdam-AIAS/CELSI/EUBA/EuroCommerce/Uni Global Union/WageIndicator Foundation
- Besamusca J and Tijdens KG (2015) Comparing collective bargaining agreements for developing countries. *International Journal of Manpower* 36(1): 86 – 102. <https://doi.org/10.1108/IJM-12-2014-0262>
- Besamusca J (2021) *COLBAR-EUROPE REPORT 7: Collectively agreed wages in Europe*. Amsterdam/Bratislava/Rome: University of Amsterdam-AIAS, CELSI, CNEL
- Blien U, Dauth W, Schank T and Schnabel C (2011) The Institutional Context of an ‘Empirical Law’: The Wage Curve under Different Regimes of Collective Bargaining. *British Journal of Industrial Relations* 51(1): 59-79. <https://doi.org/10.1111/j.1467-8543.2011.00883.x>
- Bonvin JM, Dif-Pradalier M and Moachon E (2013) A capability approach to restructuring processes: Lessons from a Swiss and a French case study. *International Journal of Manpower* 34(4): 382-396. <https://doi.org/10.1108/IJM-05-2013-0096>
- Ceccon D, Tijdens KG, Osse P, Pralitasari N, Ndoricimpa A, Ngeh Tingum E (2016). Manual and codebook of the WageIndicator Collective Agreements Database – Version 2. Amsterdam, WageIndicator Foundation
- Ceccon D and Kaandorp C (2019). Facilitating Annotation of Collective Agreements with Keywords Extraction (Update 14/10/2019). WageIndicator Foundation, Amsterdam and University of Amsterdam <https://wageindicator.org/documents/publicationslist/wageindicator-org-publications-2019/ceccon-kaandorp-sshoc-update-14102019.pdf>
- Ceccon D and Ceccon S (2020). Facilitating Annotation of Collective Agreements with Keywords Extraction. (Update 18/9/2020). WageIndicator Foundation, Amsterdam – CLARIN, Utrecht <https://wageindicator.org/documents/publicationslist/publications-2020/ceccon-ceccon-sshoc-update-18092020.pdf>
- Dunn S and Wright M (1994) Maintaining the ‘Status Quo’? An Analysis of the Contents of British Collective Agreements, 1979–1990. *British Journal of Industrial Relations* 32(1): 23–46. <https://doi.org/10.1111/j.1467-8543.1994.tb01128.x>
- Eurofound (2015) *Collective bargaining in Europe in the 21st century*. Luxembourg: Publications Office of the European Union.

<https://www.eurofound.europa.eu/publications/report/2015/industrial-relations/collective-bargaining-in-europe-in-the-21st-century>

Eurofound (2019) *Annual Review of Working Life 2018*. Luxembourg: Publications Office of the European Union.

https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef19029en.pdf

Eurofound (2020) *Industrial relations: Developments 2015–2019, Challenges and prospects in the EU*. Luxembourg: Publications Office of the European Union.

<https://www.eurofound.europa.eu/publications/flagship-report/2020/industrial-relations-developments-2015-2019>

European Commission (2016) *A new start for social dialogue*. Luxembourg: Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/2d1df4a6-66ae-11e7-b2f2-01aa75ed71a1>

European Commission (2017) *Establishing a European Pillar of Social Rights*. Brussels: Commission Communication (COM(2017)0250). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017DC0250>

European Commission (2020) *Employment and Social Developments in Europe 2020. Leaving no one behind and striving for more: Fairness and solidarity in the European social market economy*. Luxembourg: Publications Office of the European Union. [KE-BD-20-001-EN-N.pdf](#)

Freeman RB and Medoff JL (1985) What Do Unions Do? *ILR Review* 38(2): 244-263. <https://doi.org/10.1177/001979398503800207>

Fulton L (2020) *National Industrial Relations, an update*. London/Brussels: Labour Research Department / ETUI (online publication). <http://www.worker-participation.eu/National-Industrial-Relations>

Garnero A, Rycx F and Terraz I (2020) Productivity and Wage Effects of Firm-Level Collective Agreements: Evidence from Belgian Linked Panel Data. *British Journal of Industrial Relations* 58(4): 936-972. <https://onlinelibrary.wiley.com/doi/10.1111/bjir.12525>

Hayter S and Stoevska V (2010) *Social dialogue indicators. Trade union density and collective coverage. International Statistical Inquiry 2008–09. Technical Brief*. Geneva: International Labour Office, Industrial and Employment Relations Department https://www.ilo.org/wcmsp5/groups/public/---ed_dialogue/---dialogue/documents/genericdocument/wcms_165194.pdf

Heinbach WD and Schröpfer S (2007) *What a Difference Trade Makes - Export Activity and the Flexibility of Collective Bargaining Agreements*. Hohenheim: Diskussionspapiere aus dem Institut für Volkswirtschaftslehre der Universität Hohenheim 293/2007.

Jacob BA (2013) The effect of employment protection on teacher effort. *Journal of Labor Economics* 31(4): 727-761. <https://doi.org/10.1086/669942>

Magda I, Marsden D and Moriconi S (2012) Collective Agreements, Wages, and Firms' Cohorts: Evidence from Central Europe. *ILR Review* 65(3): 607-629. <https://doi.org/10.1177/001979391206500306>

Nicolitsas D (2020) Wage-setting practices in the metalworking sector in Greece. *Employee Relations* 42(6): 1441-1463. <https://doi.org/10.1108/ER-09-2019-0356>

- OECD (2018) The role of collective bargaining systems for good labour market performance. In *OECD Employment Outlook* (pp. 73–123). Paris: OECD
- OECD (2019) *Negotiating Our Way Up: Collective Bargaining in a Changing World of Work*. Paris: OECD Publishing. <https://doi.org/10.1787/1fd2da34-en>
- OECD and AIAS (2021), Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts, OECD Publishing, Paris
www.oecd.org/employment/ictwss-database.htm
- Olgati E and Shapiro G (2002) *Promoting gender equality in the workplace, report for Eurofound*. Luxembourg: Publications Office of the European Union.
https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef0161en.pdf
- Paolucci V (2017) The role of collective bargaining in addressing flexibility and security. *European Journal of Industrial Relations* 23(4): 329-346.
<https://doi.org/10.1177/0959680116687091>
- Schreuder K and Tijdens KG (2004) *Codeboek DUCADAM Dataset. Een statistisch bestand van de FNV cao-databank*. Amsterdam: University of Amsterdam, AIAS Working Paper 24.
[https:// DOI.org/10.13140/RG.2.2.31214.1312](https://DOI.org/10.13140/RG.2.2.31214.1312)
- Schulten T (2018) The role of extension in German collective bargaining, in Hayter S and Visser J (eds) *Collective Agreements: Extending Labour Protection*. Geneva: International Labour Organization, pp. 65-92. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/publication/wcms_633672.pdf
- Schulten T (2019) *Statistisches Taschenbuch Tarifpolitik*. WSI-Tarifarchiv 2019 Düsseldorf: WSI-Tarifarchiv. https://www.boeckler.de/pdf/p_ta_tarifaschenbuch_2019.pdf
- Strunk KO, Cowen J, Goldhaber D, Marianno BD, Kilbride T and Theobald R (2018) *Collective Bargaining and State-Level Reforms: Assessing Changes to the Restrictiveness of Collective Bargaining Agreements across Three States*. Washington, DC: CALDER, Working Paper No. 210-1218-1.
- Taiji R and Mills MC (2020) Non-standard Schedules, Work-Family Conflict, and the Moderating Role of National Labour Context: Evidence from 32 European Countries. *European Sociological Review* 36(2): 170-197. <https://doi.org/10.1093/esr/jcz048>
- Tijdens KG, Van Klaveren M, Bispinck R, Dribbusch H and Öz F (2014) Wage and workforce adjustments in the economic crisis in Germany and the Netherlands. *European Journal of Industrial Relations* 20(1): 1-19. <https://doi.org/10.1177/0959680113516181>
- Tijdens KG, Cecon D, Osse P, Pralitasari N, Ndoricimpa A and Ngeh Tingum E (2016) *Manual and codebook of the WageIndicator Collective Agreements Database – Version 1*. Amsterdam: WageIndicator Foundation.
<https://wageindicator.org/Wageindicatorfoundation/publications/2016/tijdens-k-g-cecon-d-osse-p-pralitasari-n-ndoricimpa-a-and-ngeh-tingum-e-2016-manual-and-codebook-of-the-wageindicator-collective-agreements-database-2013-version-1-2013-february-2016>
- Tijdens KG and Van Klaveren M (2003) *Een onderzoek naar CAO-afspraken op basis van de FNV CAO-databank en de AWWN-database*. Amsterdam: University of Amsterdam, AIAS Working Paper 03/19.
https://www.researchgate.net/publication/349003637_Een_onderzoek_naar_CAO-afspraken_op_basis_van_de_FNV_CAO-databank_en_de_AWWN-

[database A study of the Collective Bargaining Agreements archived in the FNV CAO Database and in the AAVN Database](#)

Van Gyes G and Schulten T (eds) (2015) Wage bargaining under the new European Economic Governance. Brussels: ETUI.

<https://www.etui.org/publications/books/wage-bargaining-under-the-new-european-economic-governance>

Van Klaveren M and Gregory D (2019) *Restoring multi-employer bargaining in Europe: prospects and challenges*. Brussels: ETUI. <https://www.etui.org/publications/books/restoring-multi-employer-bargaining-in-europe-prospects-and-challenges>

Yerkes M and Tijdens KG (2010) Social Risk Coverage and Decentralisation in Collective Labour Agreements: Empirical Evidence from the Netherlands. *European Journal of Industrial Relations* 16(4): 369-383. <https://doi.org/10.1177/0959680110384608>