Knowledge Management at Multiple Decision Levels: a Use Case about COVID-19 Pandemic

Franck Favetta and Fabien Duchateau (firstname.lastname@liris.cnrs.fr)
LIRIS, UMR5205, Université de Lyon, France

Geographical information systems (GIS) have become essential tools in our daily lives. Many application domains such as transportation, points of interest search or emergency response systems, include spatial concepts as part of their basic knowledge. For example, events (e.g., concert, moon landing or nuclear disaster) are typically represented using time and location concepts. Rubenstein *et al.* defines knowledge as *"information that has been organized and transformed into something understandable and applicable to problem solving and decision making"* [13]. Besides, such knowledge has to be built on recent, accurate, consistent and complete information – when possible – to allow decision-makers to take relevant actions and is mainly supported by domain rules.

In this paper, we focus on knowledge management at different levels of decision, as in the context of urban planning or emergency response support [8]. Although the latter implies quick decisions with possible significant impact, both domain applications share similar characteristics and goals, namely the detection of conflicting situations and the recommendation of actions. As an example, many urban rules are defined and applied to lead and constrain the development of territories such as the Right to Build [3]. These rules are produced at different levels, for instance national and local (city). Besides they are mainly scattered into numerous textual documents (e.g., *PLU-H* in France, which stands for Urban Local Plan), thus making them difficult to use for non-experts.

Gathering and formalizing domain knowledge from multiple decision levels could have many benefits such as an enhanced planning with global standardization and simplification of rules, automatic checking and recommendations, reuse and sharing of rules, better transparency for citizens, an improved understanding of the past or a comparison between territories through spatial analysis. In this paper, we therefore explore the challenges related to knowledge management at different decision levels. We illustrate these challenges by presenting a use case about a recent worldwide disaster, the COVID-19 pandemic.

1 Challenges of level-based knowledge management

To enable the transition towards knowledge management at different levels, several challenges need to be tackled. First, a generic model has to be established for representing rules [11, 14], which takes into account the hierarchy (e.g., rule inheritance from the top-level) and exemptions at local level. Exchanges between domain experts (urbanists, computer scientists, decision-makers, etc.) and agreements on the concepts definition are crucial for a successful data representation. Besides, such representation is different according to the actors' level. For example, when projecting to build an industrial facility, security rules are not applied on the same spatial objects: regional decision makers think about seismology while local decision makers are interested in distances to surrounding residences.

A second issue deals with the extraction of rules from textual documents or from data. Natural language processing or pattern recognition with a strong analysis on spatial concepts enables rule extraction from documents [1]. Many works on spatial data mining have been conducted for extracting knowledge from data [2]. Besides, some rules are not formally written (e.g., mimicry between close cities, best practices). Evaluation of generated rules is a challenge since both techniques do not guarantee a perfect quality, and a human validation is required (e.g. inconsistency between rules). The definition of a new visual language for designing and correcting these rules could be a major improvement for practitioners. Since many rules are produced at local level, a regional authority could be created between existing levels in order to facilitate the organization and standardization of this information.

Once the rules are available, another challenge is the data collection and integration, which mainly requires specific development [4]. Indeed, data sources are provided by various actors and at different levels, thus causing several issues. One of the most studied problem is heterogeneity: all sources include their own terminology, vocabularies and constraints (schema), and concepts may be represented differently from one provider to another. The same issue occurs for instances and values, which require entity matching or record linkage to detect equivalent entities. Heterogeneity also deals with formats (e.g., Relational, RDF, PDF documents) and categories of data (e.g., structured, textual, multimedia, maps). Data sources may also be more or less trustworthy, for instance according to quality (e.g., bioinformatics data may be automatically or manually curated). Finally, availability depends on the provider. Although the Open Data initiative promoted a better availability, especially in large cities, an effort has to be made on data quality (updates, semantics). Besides, many cities are located on a border. Thus, it is necessary to ensure data and rules continuity between adjacent administrative areas.

Finally, the previous points should lead to the development of applications that enable this transition, for instance by proposing to decision-makers various cartographic visualizations, recommendations or impact predictions of a given project. For a new urban project, an application should store necessary data and enable the selection of relevant and/or prioritized rules. The system can automatically check some of the rules, but others require human interaction to be validated [6]. Therefore the process needs to be incremental. These applications should enable feedback from local levels, where most events occur. Thus, collaborative features could improve the quality and completeness of information and produce new knowledge at higher levels (e.g., when multiple doctors successfully tested a cure, governments could decide to experiment this cure at larger scale).

The recent events brought us to identify another challenge, that is illustrated in the next section: the consequences of the need for important and quick modifications of rules in case of an unexpected and sudden situation.

2 The COVID-19 use case

This section provides application examples of the aforementioned challenges in the context of emergency response scenarios [9]. More specifically, we identify and study impact on knowledge management for the current global health crisis (as of early 2020) caused by the Corona virus outbreak from China[1]. This situation implies need for quick changes of rules and also lead

[1] WHO COVID-19 webpage, http://www.who.int/health-topics/coronavirus

to ambivalent rules. Involved decision levels are global (e.g., World Health Organization), national, regional (or equivalent, e.g., federal) and local. Due to the present situation of outbreak, we can observe issues in ruling system because of technical, political and social issues. Experiences have shown that what worked yesterday may not work today. This is especially true for the present situation where rules need to dramatically change and be adapted from one area to another. Current modifications of rules highly impact society to a large extent, for instance in terms of individual liberty, health-care, transport of goods, job and enterprises' rights and organization, taxes system, privatization/nationalization of enterprises, economic model, etc. They reshape everyday life like during war periods.

Governments provide recommendations or constraints to citizens and companies, that could be translated into rules. We mainly refer to Laurini *et al.* for a framework supporting these rules [11, 14]. Let us provide some examples. In France, when the confinement has been decided, only shops related to food, drugs and gas could be opened. Given a set of points of interest $\mathcal{P}$, the rule 1 indicates that a point of interest p which does not belong to the previous categories is closed.

$$\forall p \in \mathcal{P} : p.category \notin \{food, medication, gas\} \Rightarrow p.status \equiv "closed" \quad (1)$$

We note that this rule is ambiguous. Indeed, liquor and wine stores do not really fit in these categories. Open markets (dedicated to food) were also open according to this rule, but the government later decided to close them due to difficulties for limiting contact between people. Thus the previous rule can be updated with a conjunction ($\vee\ p.type = "open\ market"$) in its conditional part. However, in some areas, open markets are essential for obtaining food. For farmers, it can also be the only means of selling their production. Thus, regions could grant opening authorizations to some markets. By establishing a whitelist of regional markets noted $\mathcal{M} = \{market_1, market_2, ..., market_k\}$, rule 2 supersedes the national rule and should be applied afterwards:

$$\forall p \in \mathcal{P} : p.category \notin \{food, medication, gas\} \vee \\ (p.type = "open\ market" \ \wedge\ p.name \notin \mathcal{M}) \Rightarrow p.status \equiv "closed" \quad (2)$$

Finally, citizens are allowed to go out one hour per day, and they have to carry a circulation pass form[2]. Besides, walking or running is limited within one kilometre around home. GIS tools such as Geoportail[3] were heavily used to know the allowed area. Rule 3 illustrates this constraint for any move in the set $\mathcal{M}$.

$$\forall m \in \mathcal{M} : Subtract(current\_time, m.departure\_time) > 60 \vee \\ Dist(current\_location, Coords(m.address)) > 1000 \Rightarrow GiveFine(m.person\_name) \quad (3)$$

This rule was quickly created to fit a generic case, but one may notice that the one kilometre distance is mostly adapted in big cities, but is probably insufficient in countryside areas. Indeed, a local authority may supersede this rule to allow citizens for a larger distance in rural areas. Finally, note that this rule could be automatically verified by checking geo-locations of citizens through their mobile. This would require careful attention to exceptions (e.g., health workers, shop vendors), and would not be acceptable by citizens (against individual liberties in many countries).

It is a common rule makers' practice to adapt a rule to unexpected situations and to their specific territory by superseding rules. The whole process of creating a general rule at the

[2] French circulation pass http://www.service-public.fr/particuliers/vosdroits/R55781

[3] Geoportail, http://www.geoportail.gouv.fr/

national level and its regional adaptation usually takes time. But during an outbreak, rule makers are urged to quickly modify or create rules, especially to fit all kind of situations. Some examples In France deal with the possibility to go out: some cities declared a curfew, thus complementing existing restrictions. To limit the number of people outside in Paris, running was not allowed any more during daytime. As a consequence, most runners practise at 7pm, which tend to concentrate many people outside and new rules may be dictated to avoid this.

Adapting rules to recent events and discoveries is even harder for rule makers due to difficulties in political crisis organization at worldwide level as well as at national and regional levels. Part of the reason is heterogeneity, complexity and today's "hand-made" (non-digital) organization of rules, which does not facilitate simulation.

The next challenge is automatic extraction of rules, mostly from textual documents. The COVID-19 outbreak has triggered the production of a large amount of documents from the scientific community, as illustrated by the 50,000 articles in the COVID-19 open research dataset[4]. Various tools have been developed to facilitate search and retrieval of relevant documents[5] or to visualize relationships between biological concepts[6]. Newspapers articles also contain information like the most recent limitations dictated by governments or latest statistics about the pandemic. Extracting rules from such documents first require named entity recognition (NER) or entity linking techniques [15]. Wang *et al.* have proposed the identification of relevant concepts on the COVID-19 dataset[7][16]. For instance, in the sentence `SARS-COV spike proteins have a strong binding affinity to human ACE2`, the terms *SARS-COV spike proteins* and *ACE2* are identified under the concept *Gene_or_Genome*. This detection is a first step towards rule extraction, but finding the type of relationship between concepts is still a challenging issue [7].

Popular events are typically accompanied with fake news, that even the WHO tries to detect and fight[8]. These fake news may produce bad rules, either irrelevant or contradictory. Let us consider the fact `COVID-19 virus cannot be transmitted in areas with hot and humid climates`, which has been considered as a fake news by WHO. It could generate rule 4 using Köppen–Geiger climate classification system [10], whose climate zones are noted as $\mathcal{Z}$: if a geographical area a is included in a climate zone z which is hot and humid (*Cfa* or *Cwa*), then we cannot find the virus in z. This irrelevant rule may incite people not to respect basic safety recommendations in these areas.

$$\forall a \in \mathcal{A}, \exists z \in \mathcal{Z} : Contains(z, a) \ \wedge \ (z.climate \equiv "Cfa" \ \vee \ z.climate \equiv "Cwa") \Rightarrow a.covid\_presence \equiv false \quad (4)$$

A last issue deals with the multiple levels at which rules may be extracted. How to manage complementary or conflicting rules extracted from different sources? The difficulty also arises because some levels (such as WHO) only formulate recommendations, but not restrictions. On the contrary, countries and regions dictate both constraints (to be respected, e.g. with rule 3 about maximum distance and period for going out) and recommendations (e.g., people others than health workers do not need to wear masks). As an example, the French government decided during the first months that wearing masks should not be encouraged. On the contrary, Asian countries have adopted the mask for citizens as a means to limit virus propagation. Several French cities (e.g., Nice, Sceaux) have voted laws to locally enforce mask wearing. Surprisingly

[4]COVID-19 open research dataset, http://pages.semanticscholar.org/coronavirus-research

[5]CORD-19 Explorer, http://cord-19.apps.allenai.org/

[6]CoViz, http://coviz.apps.allenai.org/

[7]CORD-NER, http://xuanwang91.github.io/2020-03-20-cord19-ner/

[8]WHO myth busters about COVID-19, http://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters

these decisions were later cancelled by the regions of concerned cities. The timeline of rules is therefore crucial in multi-level knowledge management.

To conclude the rule extraction, we note that tomorrow's changes may still not work if the modification process is not fast enough. Rules need flexibility and velocity of adaptation, and automation could improve response times. The recent modifications of national laws required by outbreak situation were extraordinary faster than modification times we use to know. As outbreak came from China approximately two months before West, reacting rules of western countries are partly inspired by Asian rules such as lock-down, which is an example of rules' mimic from one region to another.

To apply extracted rules, experts have to collect and integrate data from multiple sources (e.g., statistics about contaminated people, confinement measures by country). An advantage of the strong focus on COVID-19 is that data producers (governments, organizations at different levels) publish their data in hope that a solution is quickly discovered. WHO provides data for each country, including maps, through its global health observatory[9]. It contains more than 2,260 indicators and provide data as JSON or CSV. Another provider is the Humanitarian Data Exchange platform[10], with tens of datasets provided by different actors. At national levels, agencies also offer datasets that are stored on platforms such as the French `data.gouv.fr` website. For these data, levels range from national to local according to the provider and the degree of precision.

Spatial data is very heterogeneous across all these datasets. For instance, let us study measures taken to limit the pandemic. In a dataset which aggregates measures decided by authorities (social distance, lock-down, etc.[11]), only the country name and one other administrative level (e.g., state, region, province, city) are available. Besides, there is no standardization for values in the administrative level field, thus leading to inconsistencies for future exploitation. The Oxford response tracker dataset[12] only mentions the country name, but more specific locations (e..g, airports, schools) are spread out in textual fields (e.g., in newspaper titles,). Binary fields stand for the category of restrictions (e.g., general, workplace, transport, events, fiscal). Comparing these two datasets require processes with variable quality results (mainly entity extraction, deduplication and entity linking). In the school disclosure dataset[13], spatial data includes the country name and an application scale, either national or localized, in which case a note field may provide details (e.g., *"27 of 28 states"*). Finally, the mobility restriction measure is also described by the International Organization for Migration[14]. In addition to country code and name, points of interest have a location name field, which may specify an airport or port name, a land or sea border point. These values are also formatted differently (e.g., using sporadic or detailed address, including useless country name or location type) and would need to be cleaned during an integration process. To compare COVID-19 contamination cases, locations are represented with four administrative levels and city name (based on Google Geolocation API) in the Metabiota dataset[15] while they are represented as coordinates (latitude and longitude) in the WHO dataset[16].

[9] WHO Global Health Observatory, http://www.who.int/data/gho

[10] COVID-19 pandemic locations, http://data.humdata.org/event/covid-19

[11] Government measures dataset, http://data.humdata.org/dataset/acaps-covid19-government-measures-dataset

[12] Oxford response tracker dataset, http://data.humdata.org/dataset/oxford-covid-19-government-response-tracker

[13] Global school closures dataset, http://data.humdata.org/dataset/global-school-closures-covid19

[14] Mobility restriction dataset, http://data.humdata.org/dataset/country-point-of-entry-mobility-restriction-covid-19-iom-dtm

[15] COVID-19 cases from Metabiota, http://data.humdata.org/dataset/2019-novel-coronavirus-cases

[16] COVID-19 cases from WHO, http://data.humdata.org/dataset/coronavirus-covid-19-cases-data-for-china-

Integrating data at lower levels becomes more complex. For instance, consider food services that either are open or deliver during the pandemic. A German dataset is available for Berlin[17] and a French one covers the city of Poitiers[18]. The comparison of schemas require to solve multilingual issues. Besides, data are represented with different structures. For instance, French opening hours may be detailed in a textual field (*specificities*) while the German dataset includes seven fields, one per week day. Both provides food categories, addresses and contact information, but geolocation is only available in the French dataset, which would require geocoding if one plans to present data on a map. Instances also have to tackle multilingual issues, overlapping categories, field merging for address, etc. Note that we have chosen cities of Berlin and Poitiers, but datasets exist for other areas such as Nevers, Corsica, all of them with different schemas. There is even a dataset for the whole territory based on OpenStreetMap[19]. Some metrics (e.g., estimation of completeness) could be useful to select the most relevant datasets and minimize the integration task.

When dealing with integration, it appears that finer levels, which are the most interesting due to their accuracy, are also the most difficult to fusion. Yet, this process is essential to ensure that rule's application produces the most relevant and reliable results.

The last challenge is related to application and transition towards automatic knowledge management. We first discuss the evolution of rules. In a normal situation, hand-shaking is a common practice in Western countries for salutation. Such rule could be inferred from multiple sources, but when a change occurs, a new conflicting rule (e.g., no physical contact for salutation) is produced, and the system should decide which one to apply. Dates of rule extraction are reliable indicators in that case. However, rules extracted in a very short time period may be contradictory, thus leading to ambivalent situations. This occurred in France when the confinement has been decided. On the one hand, people were encouraged to work from home when possible or stay at home for non-essential jobs. But governments are also worried about the economic impact. Other ministers claim that workers unable to do home activities should therefore go to their workplace (e.g., building labourers). Although social distance has become the norm, another minister announced that 200,000 unemployed people (due to COVID-19) should go and help farmers for producing food. Similarly, transportation should be limited, but health workers have to be able to go to their workplace. A last example of possibly conflicting rules is about democracy. This political system goes along with information transparency and individual liberty. Yet, during troubled times, governments communication first trended to minimize or dissimulate danger in order to avoid panic, and then they set rules which restrict liberty such as confinement and geo-tracking. Our societies have to find a trade-off to respect these ambivalent rules, maybe by using uncertainty theories like in the data uncertainty domain [5].

Another common phenomenon, which is not specific to outbreak situation, is about rules' discontinuity at borders. This occurs when neighbour local authorities emit discontinuous rules, e.g., fishing may be prohibited in a local area because of river's pollution but authorized in a downriver neighbouring region. Figure 1 illustrates a similar situation related to maximum travelling distance: a road to home may cross another region (in rose) with a different maximum authorized distance from home. Thus, a citizen may become an outlaw due to rules' discontinuity. In France, some local authorities have taken very restrictive rules (e.g., residents

and-the-rest-of-the-world

[17] Berlin Gastronomien dataset, http://www.govdata.de/web/guest/daten/-/details/gastronomien-und-ladengeschafte-mit-liefer-und-abholdiensten

[18] Food shop and producers in Poitiers dataset, http://www.data.gouv.fr/fr/datasets/commerces-alimentaires-et-producteurs-locaux-ouverts-covid19/

[19] Open places in France, https://www.data.gouv.fr/fr/datasets/lieux-ouverts-ou-fermes-pendant-le-confinement-covid-19/

allowed to go out only 10 meters far from their home). An application could be useful in these cases to recommend new paths and avoid a fine. A consequence of border discontinuities is implication of another rules. If two countries follow different strategies against outbreak, then border trespassing is prohibited or restricted with virus testing. Note that in this situation, a rule is conditioned to the existence of another one ("IF $rule_x$ *at neighbour* THEN *local action/fact*", based on format in [12]).

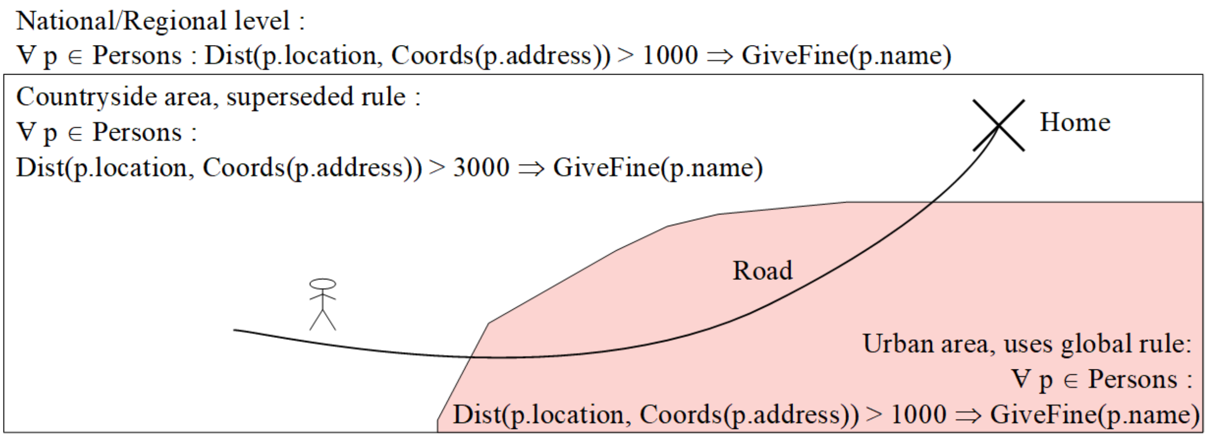

Figure 1: Example of discontinuous rules at borders

A formal and digital knowledge application could also demonstrate to decision makers the uselessness or limited impact of a new rule. The COVID-19 crisis has seen many local rules that have been quickly removed due to citizen or government pressure (e.g., the possibility to relax only two minutes on benches or prohibition of alcohol). The collaborative aspect is also crucial in this kind of crisis. Many data sources originate or are built from local information (e.g., number of deaths in local hospitals, open food shops in a city). This means information is accessible at higher levels and may be combined to derive new knowledge (e.g., estimate when a region has reached the peak of the pandemic according to local counts).

3 Conclusion

In this paper, we have presented the main challenges related to knowledge management with multiple decision levels, especially for urban planning or emergency response support. Currently, most of this knowledge is described in textual documents or only known by experts. But in case of major crisis such as the COVID-19, there is a strong need for quick decisions and rules' adaptation. Automatically gathering, formalizing and exploiting domain knowledge is therefore crucial to facilitate decision-making. Sharing best practices in multi-level knowledge management is also at the heart of scientific questions, as shown by a Kaggle competition[20]

We have identified four main challenges: modelling of rules, extraction of rules from documents, gathering and integration of relevant data and design of applications. They have been illustrated on the current COVID-19 pandemic, for which much knowledge and many data sources are released at different levels. We insisted on several issues such as quick rules' modifications, ambivalent rules, and discontinuity at borders. We believe that a formal and digital knowledge system of rules would contribute to meet the presented requirements. It would facilitate organization of rules between different levels of governance, make modifications of rules

[20]Information sharing task, http://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/tasks?taskId=583

faster, enable the detection of ambivalence and discontinuity at borders and possibly infer new knowledge or recommendations.

References

[1] Charu C Aggarwal and ChengXiang Zhai. *Mining text data*. Springer Science & Business Media, 2012.

[2] Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, 51(4):1–41, 2018. Link to CSUR18.

[3] Mickaël Brasebin, Julien Perret, Sébastien Mustière, and Christiane Weber. A generic model to exploit urban regulation knowledge. *ISPRS International Journal of Geo-Information*, 5(2):14, 2016. Link to ISPRS16.

[4] AnHai Doan, Alon Halevy, and Zachary Ives. *Principles of data integration*. Elsevier, 2012.

[5] Didier Dubois and Henri Prade. Formal representations of uncertainty. *Decision-making process: concepts and methods*, pages 85–156, 2009. Link to DMP09.

[6] Susa Eräranta and Aija Staffans. From situation awareness to smart city planning and decision making. In *Computers in Urban Planning and Urban Management*, volume 197, pages 1–17, 2015. Link to CUPUM15.

[7] Oren Etzioni, Anthony Fader, Janara Christensen, Stephen Soderland, and Mausam Mausam. Open information extraction: The second generation. In *IJCAI*, volume 11, pages 3–10, 2011. Link to IJCAI11.

[8] Manfred Fischer. From conventional to knowledge-based geographic information systems. *Computers, Environment and Urban Systems*, 18:233–242, 07 1994. Link to CEUS94.

[9] Vagelis Hristidis, Shu-Ching Chen, Tao Li, Steven Luis, and Yi Deng. Survey of data management and analysis in disaster situations. *Journal of Systems and Software*, 83(10):1701–1714, 2010. Link to JSS10.

[10] Markus Kottek, Jürgen Grieser, Christoph Beck, Bruno Rudolf, and Franz Rubel. World map of the köppen-geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259–263, 2006. Link to MZ06.

[11] Robert Laurini. A mathematical language for the modeling of geospatial static rules. In *DMS Conference on Visualization and Visual Languages*, pages 1–20, 2019. Link to DMS19.

[12] Ronald Ross. More on the if-then format for expressing business rules: Questions and answers. *Business Rules Journal*, 12(4), 2011. Link to BRJ11.

[13] B Rubenstein-Montano. A survey of knowledge-based information systems for urban planning: moving towards knowledge management. *Computers, Environment and Urban Systems*, 24(3):155–172, 2000.

[14] Sylvie Servigne, Franck Favetta, and Robert Laurini. An introduction to geographic rules. In *Distributed Multimedia Systems (DMS)*, pages 91–97, 2016. Link to DMS16.

[15] Wei Shen, Jianyong Wang, and Jiawei Han. Entity linking with a knowledge base: Issues, techniques, and solutions. *Knowledge and Data Engineering, IEEE Transactions on*, 27(2):443–460, 2015. Link to TKDE14.

[16] Xuan Wang, Xiangchen Song, Yingjun Guan, Bangzheng Li, and Jiawei Han. Comprehensive named entity recognition on cord-19 with distant or weak supervision, 2020. Link to CORD-19.