Data Anonymization for Maintenance Knowledge Sharing

Hicham Hossayni
Schneider Electric, Grenoble, France

Imran Khan
Schneider Electric, Grenoble, France

Noel Crespi
Institut Polytechnique de Paris, Paris, France

***Abstract*—Formerly considered as part of general enterprise costs, industrial maintenance has become critical for business continuity and a real source of data. Despite the heavy investments made by companies in smart manufacturing, traditional maintenance practices still dominate the industrial landscape. Maintenance knowledge sharing between industries can significantly optimize the maintenance activity and improve the processes efficiency. Different international standards and initiatives are promoting such approach. However, this trend failed to gain ground in the manufacturing industry. In this paper, we present the results of our investigation about the real roadblocks that obstruct the progress of the maintenance knowledge sharing approach. We determined that the knowledge graphs and, more importantly, the automated data anonymization techniques can facilitate the development of general-purpose solutions to share the maintenance knowledge among concerned actors.**

Index Terms: Data anonymization, industrial maintenance, knowledge sharing

■ **INTRODUCTION** Over the past few decades, knowledge sharing has become both democratized and essential for individuals and professionals. It is a go-to approach in many domains like in computer software through open-source software and tools, open data-sets, crowd-sourcing and collective universal encyclopedia (Wikipedia is just one example). However, this trend has not gained traction in the manufacturing industry yet, especially for the maintenance activity. Typically, industrial maintenance knowledge is inaccessible due to industry policies and practices, with little to no motivation to share it with others. Today, when a new machine is installed in a factory, there is no existing knowledge of its failures unless there is a human expert who has dealt with similar machines before. In any given factory, each failure is only discovered at its first occurrence and requires usually a long and costly curative maintenance operation. The maintenance process includes diagnostics to determine the reasons for a failure, its impact, and to define and apply the correct repair procedures. Therefore, the same failure of a machine, installed in many factories, could lead to high costs and big production losses because every factory, or even every maintenance operator deals with the failure individually, from the diagnostics phase to the problem resolution. A potential solution that could help improve the situation would be to enable the sharing of maintenance knowledge and experiences between the industries or factories that own and operate the same types of machines. The shared knowledge could be used by the machine operators as a guide to reduce the diagnostic and repair times, target the failures' root cause(s), and improve the machines' efficiency. Another important effect of maintenance knowledge sharing is that it will transfer maintenance expertise between experts and novices. This will lead to much faster handling of machine failures, training of new staff and to reduce the dependency on external maintenance companies. Even more importantly, it will reduce the number of accidents related to maintenance activities.

Existing maintenance data exchange standards & projects

There are many international standards and consortium agreements that promote and adopt the maintenance knowledge sharing approach. Some of the most well-known ones are:

- The OREDA project [1]: Offshore and Onshore Reliability Data for oil and gas industries, which led to the ISO 14224 international standard [2]. It is a project organization with 7 to 11 oil and gas companies as members, managing data of 292 installations and 18000 equipment units. Running for more than 35 years, OREDA has led to significant cost savings in the development and operation of platforms [1], and has helped has helped the participating oil companies to save $70M. [2].
- SPARTA [3]: System Performance, Availability and Reliability Trend Analysis adopted by 9 wind energy companies in the UK. It manages and exchanges data of 19 wind farms and 1256 turbines in 2020. It covers more than 60% of all offshore wind power generation in the UK. Thanks to SPARTA, the average number of crew transfers per turbine in the UK fell by 50% between 2014 and 2018, to around six trips per year [3].
- WInDPool [4]: stands for Windenergy-InformationData-Pool, for wind energy industries in Germany. Adopted by 7 members to exchange the maintenance data for a fleet of 640 wind turbines onshore and 297 wind turbines offshore with an expectation of more than 1M€ of costs savings on the maintenance activity [4].
- The Configuration Data Exchange [5]: launched by General Electric Aviation for the world-wide aviation industry. $4 billion savings are anticipated across the sector thanks to the configuration data exchange solution [5].
- OPDE [6]: The International Pipe Failure Data Exchange project; and
- ISO 6527 [7]: A Reliability Data Sharing standard for nuclear energy producers.

Through these standards and applications, the maintenance knowledge sharing concept has proven its efficiency in the improvement of equipment reliability, in the enhancement of maintenance processes and in the reduction of production costs, leading to savings over a machine's life cycle. However, each of these standards or applications targets a specific domain and has

[1] http://www.datsi.fi.upm.es/~rail/new/WP2/OREDA-history.htm (accessed 17 Feb, 2021)

[2] https://www.oreda.com/join-us (accessed 17 Feb, 2021)

[3] https://ore.catapult.org.uk/wp-content/uploads/2018/11/SPARTA-Portfolio-Review-201718-1.pdf (accessed 17 Feb, 2021)

[4] https://wind-pool.iee.fraunhofer.de/opencms/export/sites/WInD-Pool/img/WInD-Pool-Business-Case_ENG.pdf (accessed 17 Feb, 2021)

[5] https://www.businesswire.com/news/home/20161115005705/en/GE-Aviation-Launches-Configuration-Data-Exchange-Reduce (accessed 17 Feb, 2021)

been adopted by only a handful of participants who exchange their data under multilateral agreements.

Survey of maintenance knowledge-sharing challenges

To understand the perspectives of different actors involved in industrial maintenance operations, we conducted a survey through a set of direct interviews. Plant manager, production manager, machine operators, maintenance engineers, and different other industrial profiles were interviewed. We targeted 30 companies from various domains (Engineering, pharmaceuticals, automotive, HVAC, Food & Beverage, ...etc.) and from 6 countries (UK, Spain, France, Switzerland, China and US). The survey was limited to manufacturing industries for which the equipment maintenance is a major activity. On one hand, the survey brought-forward several issues that are faced by these actors in their daily routine. We found real interest in a maintenance knowledge sharing solution and these actors believed that it will be beneficial in the short term as well as in the long term. On the other hand, the survey allowed us to identify the various challenges that could impede the development of generic maintenance knowledge sharing solutions. Some of the most important ones are:

1) **Business culture**: Sharing knowledge is not a common activity in most organizations. It has been demonstrated that individuals are rewarded mostly for what they know, and not what they share [8]. The competitive environment that encourages individual instead of collective productivity has taught employees to consider their knowledge as their own property, and that to deepen and defend their knowledge is the main way to keep their jobs within the organization.
2) **Maintenance data collection**: The collection of good quality maintenance data[6] is a mandatory step before being able to extract and share useful maintenance knowledge[7] from it. It is a long-term activity, requiring the involvement of competent and well-trained personnel to guarantee the quality and usefulness of the collected data. However, this is usually seen as a marginal and costly process instead of being considered as a solid investment for the future.
3) **Human factors**: Many people may be reluctant to write a detailed report when they are not confident in their own expertise and want to avoid being judged. Also, when failure frequency is high, many maintenance operators believe there is no time to write a detailed report and thus may produce minimal reports with little or no transferable knowledge.
4) **Common maintenance taxonomy**: Collecting good quality maintenance data is not enough to make it shareable. Entities exchanging their maintenance data must have the same understanding of the shared data and hence, use the same vocabulary or taxonomy.
5) **Legal aspects**: Every maintenance report is a potential legal liability. In fact, a report's producer is legally responsible for material damages that may be caused by the application of his or her instructions. The responsibility could be penal in the case of human damages. Due to this issue of responsibility, some large companies destroy their maintenance reports beyond a certain legal period (e.g., 2 years in Europe) to avoid any future problems associated with their reports.
6) **Sensitive data disclosure**: Maintenance data may contain sensitive information that might compromise or negatively impact a company's activity. Thus, companies often choose to keep all their data secret to ensure business stability.

All the standards and projects presented above try to address aspects of these challenges. The existence of international standards that rule the data sharing activity may, itself, has a positive impact on challenges 1,2, and 3, since the existing successful experiences are sufficient to influence the business culture and processes and indirectly reduce the human factors risk. Challenge 4 is ex-

[6] Maintenance data refers to maintenance reports established to document a maintenance operation. These may contain data about the machine's components, description of the failure and details of the maintenance procedure.

[7] By maintenance knowledge we refer to the result of an aggregation of maintenance data collected by many entities that is all relative to the same machine type.

plicitly handled by the definition of standard exhaustive taxonomies that describe the equipment components, failure details, and the maintenance procedures of the maintenance key performance indicators (KPIs). This is already possible where the standards are defined for specific industry fields (e.g., oil & gas) with well-known types of equipment. However, this solution cannot be adopted to cover all industries. Consortium agreements are a solution for the legal aspect (challenge 5), as data is contractually agreed to be shared by declining all responsibility relative to its use. Nevertheless, this approach in not scalable and not efficient when targeting a large set of companies. Finally, the sixth challenge is being addressed by the cited standards by their recommendations to apply data anonymization. This process aims to remove all sensitive data from the maintenance knowledge before it is shared. In practice, the anonymization of maintenance data is usually done manually, as no automated approach has been adopted or promoted by the standards and applications.

SemKoRe for maintenance data sharing

In our previous work we proposed ”SemKoRe” [9], a technical solution for maintenance data sharing, with which we aimed to tackle challenges 4, 5, and 6. SemKoRe targets the Original Equipment Manufacturers (OEMs) that produce and sell machines to other industries. An OEM can offer SemKoRe as a supplementary service, connected to a machine, that simplifies the collection and reuse of maintenance data, and then to share the growing body of maintenance knowledge with customers that own the same machines types. our customers showed an interest in using the SemKoRe approach to enhance their industrial maintenance processes. Moreover, the overall machine building process can be optimized. The machine design phase can benefit from the maintenance feedback to identify any weaknesses of a machine and can improve its design. Furthermore, the collected statistics will allow the performance comparison of a particular machine working in different locations and contexts. Thus, additional services and recommendations can be proposed to the customers in order to optimize their manufacturing process.

Technically, instead of proposing an exhaustive taxonomy for all types of equipment, machines or failures for a specific industry field, our approach defines individual taxonomies for each machine type. In fact, SemKoRe defines an ontology-based taxonomy to describe the invariant concepts in maintenance activity. This taxonomy is then extended by the OEMs to build machine-specific taxonomies that describe the exhaustive vocabulary relative to each machine. Moreover, SemKoRe proposes ontology-based adaptive UIs to simplify the filling of failures or maintenance details with more predefined inputs than free texts, which may help to increase the quality of the collected data. To tackle the challenges 5 and 6, we recommend the adoption of an automated anonymization approach [10]. It consists of applying automated tools to remove the sensitive data within maintenance reports so that they can be shared securely between different companies. The removal of sensitive data also removes all information about its origin, and thus the resulting anonymized data assures that no legal responsibility could be engaged against its producer (challenge 5). In the literature, existing automated anonymization solutions for structured or text data achieve high accuracy scores (around 98% [11]). This level is judged to be sufficient [12] to address the 6^{th} challenge (sensitive data disclosure) as the remaining 2% of non-anonymized entities cannot be statistically distinguished from the anonymized ones. In the next section, we describe the automated data anonymization approach in detail. This solution has proven to be the most important enabler of SemKoRe for OEMs and their customers.

Automated anonymization of maintenance data

In the digital era, big data and massive processing capabilities allow us to explore new possibilities. They have become essential for the optimization of our activities and resources, and help us to discover new opportunities. However, the ubiquity of personal information or PII (personally identifiable information) makes this task very challenging [13]. Data anonymization originated from the need to depersonalize processed information while guaranteeing the usefulness and knowledge contained in that information. Around the world, multiple regulations require the anonymization of data before it can be processed, especially in the financial services or the medical field, which is the leader in terms of privacy preservation needs [14]. However, to the best of our knowledge, no initiative for industrial maintenance data anonymization has emerged to date. Several data anonymization techniques can be found in the literature [15], ranging from simple dictionary searches to Deep Learning-based techniques. The main objective of these techniques is to detect sensitive information in a text and classify it according to its nature. Machine learning techniques have shown their absolute superiority in this task. Other methods of pattern-matching or dictionary lookups are complementary and allow to simplify and focus learning models in difficult cases. It is very important to consider the nature of the data to be anonymized. Data can be classified as structured and unstructured (i.e., text data). Structured data is already tagged data that can be stored in a relational database, or in a structured file such as: JSON, XML, CSV, spreadsheets, . . . etc. Unstructured text data might contain any type of text information. In our case, industrial maintenance data is usually a mix of structured data (e.g., dates, machine identifiers, sites, addresses, maintenance operator identities) and unstructured text data (e.g., diagnostic steps, solving procedure, and maintenance operators' observations). Both types of data could potentially contain sensitive information relative to the manufacturing process, production details, machine configuration, or other sensitive data.

Data sensitiveness determination

An important step for data anonymization is to determine sensitive data and decide if it needs to be removed/replaced or kept without any change. Maintenance reports may contain two types of sensitive data: Personal data and Business data. Personal data or Personally Identifiable Information (PII) includes all data that may refer to a person directly (their name, email address, phone number, social insurance number, etc.), or indirectly, known as quasi-identifiers (such as their birth date, zip code, profession, or gender). Only one direct identifier is needed to identify a person, while a combination of two or more indirect identifiers is required for the same purpose. Business-sensitive data is any information that poses a risk to the company if discovered. Examples of such data include financial data, supplier and customer information, manufacturing secrets, etc. Our survey showed that judging a data item as being business-sensitive varies from one company to another. For example, some users considered that the configuration of their machines was not sensitive, while others (e.g., a flavor manufacturing company, or a tire company) believe that their machine configuration is a manufacturing secret and thus a part of their industrial property, and so it cannot be shared with others. In the industrial context, deciding which data is sensitive and which is not is a challenging task. Consensus on business data sensitiveness among all the actors appears to be impossible to achieve . This situation reinforces the need for an anonymization system with a large set of capabilities to cover all customers' needs, and that allows users to define their sensitive data types as configuration.

Structured data anonymization

The purpose of structured data anonymization methods is to prevent the re-identification of sensitive data from an anonymized dataset by means of matching with external data sources [16]. These methods are classified into two categories: perturbative and non-perturbative methods [17]. Perturbative methods alter the data set to introduce uncertainty around the true values, while non-perturbative methods reduce the details in data by imposing generalization or by suppressing certain values without distorting the data structure. Some of the most well-known structured data anonymization techniques are: Local suppression, Micro-aggregation, Noise-addition and Swapping. These techniques try to satisfy some statistical properties, such as K-Anonymity, l-diversity or t-closeness. The validation of these properties does not mean that the privacy is 100% preserved, but they are a good indicator of the quality of the anonymization process. Some examples of structured data anonymization tools are IBM's ARX Data Anonymization Tool, Amnesia, the Cornell Anonymization Toolkit (CAT), and Aircloak Insights.

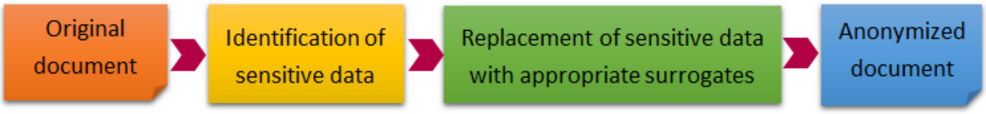
Original document
Identification of sensitive data
Replacement of sensitive data with appropriate surrogates
Anonymized document

Figure 1. Text anonymization process.

Text anonymization process overview

Figure 1 shows an overview of the process of text anonymization. It is composed of two main tasks. The first consists of identifying and classifying the sensitive data within a text. Both personal and business sensitive data are detected during this phase. The second step determines the appropriate surrogates with which to replace the sensitive data detected during the previous phase, thereby producing an anonymized document that still contains valuable data. The next sections present some tools and techniques that are used for both tasks.

Identification of sensitive data process

Different approaches are used for sensitive data detection within a text, all of which are based on the text analysis and rely mainly on Natural Language Processing (NLP) techniques. As shown in **Figure** 2, this process follows the standard schema of an ML Ensembling pipeline. It begins with the pre-processing of the input data before it is simultaneously fed to different classifiers. Each classifier is responsible for detecting a specific type of sensitive data.

This approach is adopted by the Named Entity Recognition (NER) tools [18]. NER consists of

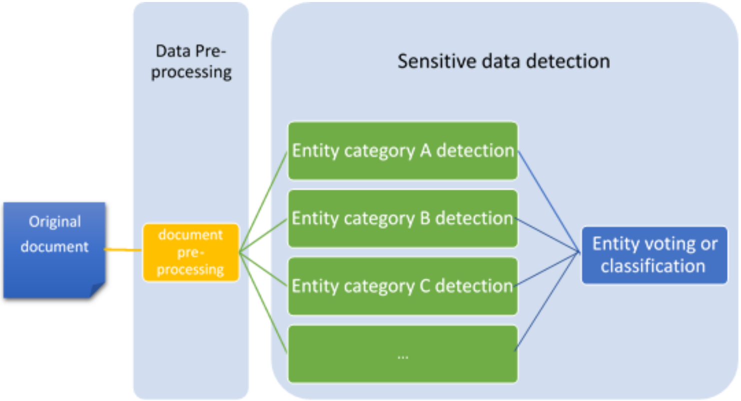
Data Pre-processing
Sensitive data detection
Original document
document pre-processing
Entity category A detection
Entity category B detection
Entity category C detection
...
Entity voting or classification

Figure 2. NLP-based sensitive data detection process

Table 1. Examples of sensitive data surrogates.

Sensitive data	**Examples of surrogates**
Date	In many cases, it may not be sensitive data. Otherwise, replace it by the week number or by applying a random shift on [-15days, +15days]; this will maintain the seasonal information that may have an implication in the failure occurrence.
Person names, employee numbers	Replace this information by a person's role within the company, such as: operator, , maintenance manager, etc. When possible, adding information about a person's expertise level could be worthwhile and will contribute to the credibility of a report.
Location, address	Replace all the addresses and locations by ”the factory”. Add the country if there are at least 10 or more companies that own the same type of machine in that country.
Product names	Replace all product names by the keyword ”the Product”, except for the names of the machine components that are common to all customers. For example **Schneider Lexium Servo Drive** is a product name, but it may be common to all the machine owners, making it non-sensitive. This means that this surrogate could behave differently for two different machine types.
Machine configuration(s)	The surrogate of this data type will change depending on the machine type. If the machine configuration is sensitive, then this data will simply be dropped from the content. However, if it is not sensitive but should not be shared as is, then all numeric values must be replaced by adding a random value, for example in the range of [-20%,+20%]. This should give an insight on the range of acceptable values.

identifying, within a text, the entities (words or group of words) that are relative to real-world objects with associated names, and classifying them into pre-defined categories such as person names, organizations, locations, time, quantities, etc. There are many NER tools and APIs, with different capabilities and variable accuracy. As an example, the CLARIN[8] (Common Language Resources and Technology Infrastructure) website references several NER tools, along with their different capabilities and supported languages. Some famous NER python libraries like NLTK[9] , SpaCy[10] or Flair[11] are provided with built-in NER features that support many languages and offer state-of-the-art accuracy and performance. Actually, the well-known NER tools are most-commonly used for personal data identification. Regarding the maintenance data anonymization, additional classifiers should be developed and trained for the detection of sensitive business data, following the same approach in Figure 2. Such classifiers can use different machine learning techniques like: Neural Networks, Recurrent Neural Networks (RNN), Conditional Random Fields (CRF), or transfer learning by using transformer models such as Google's BERT [19]. In practice, the above NLP-based techniques are often used together with the rule-based techniques, that are efficient in the detection of entities with well-known and usually static formats or names. These techniques rely on two approaches: pattern matching using regular expressions to detect entities with specific formats (such as dates, phone numbers, credit cards, social/employee numbers), and dictionary lookups to detect entities with well-known names, including countries, cities, streets, organizations, and products.

Replacement of sensitive data with appropriate surrogates

The generation of surrogates is one of the most challenging problems in automated data anonymization. Unlike the abundant research about sensitive data detection techniques, little progress has been made in surrogate generation [20]. Only some obvious pre-defined substitutions can be found in the literature, such as replacing a person's name by another random name of the same gender, or replacing dates, ages and street numbers by random values. Some works remove all sensitive information and replaces it with their category name (e.g., "Mr. Jack" becomes <PERSON>) instead of using a suitable replacement. However, the risk with such solutions is the loss of vital information and they cannot be generalized and adapted for other use cases, as well as the introduction of confusing lack of detail. Table 1 shows some examples of (generally) appropriate sensitive data surrogates. Ensuring the semantic correctness and usefulness of the anonymized data is a key challenge of surrogate generation [20]. The simplistic solutions, removing the sensitive data or using just the data category as a surrogate, cannot adequately satisfy that requirement. More sophisticated approaches are needed, with a deep dive into each sensitive data type to determine the adequate surrogate for each usage context.

Conclusions and perspectives

There are many standards for maintenance data sharing. Despite their limited application scope, their positive impact on the enhancement of the maintenance activity makes the generalization of their approach more attractive. Given this perspective, we developed our solution, 'SemKoRe', to simplify the collection, reuse and sharing of maintenance data through powerful services relying on semantic web technologies. Before adopting the maintenance data sharing approach, our customers had expressed the need for automated anonymization features to avoid sensitive data disclosure. This requirement came up as the make or break point. However, we've found that the application of data anonymization for manufacturing industries is still a challenging task. On one hand, a consensus about the sensitiveness of data seems unreachable for the different industrial actors, since each data fragment could be considered as sensitive for some and non-sensitive for others, which makes a one-solution-fits-all approach impossible. On the other hand, the replacement of sensitive data with appropriate surrogates is still an open challenge. Tackling these challenges is our cur-

[8] https://www.clarin.eu/resource-families/tools-named-entity-recognition (accessed 8 Jan, 2021)
[9] https://www.nltk.org/ (accessed 8 Jan, 2021)
[10] https://spacy.io/ (accessed 8 Jan, 2021)
[11] https://github.com/flairNLP/flair (accessed 8 Jan, 2021)

rent priority for the SemKoRe project. We are working on a novel solution, combining the efficiency of the ML techniques and the semantic web technologies to reconcile the data protection and context-awareness. Furthermore, collecting diversified maintenance data is another issue that needs to be considered, since a single source of data (e.g. Schneider-Electric) is not representative enough to cover all the needs and specificities of the various industrial domains.

■ REFERENCES

1. SINTEF Technology and Society, Norges teknisk-naturvitenskapelige universitet, OREDA. Offshore Reliability Data Handbook, OREDA participants, 2002.
2. ISO 14224: 2016. Petroleum, petrochemical and natural gas industries – Collection and exchange of reliability and maintenance data for equipment. Third edition, Geneva, Switzerland.: International Organization for Standardization (ISO).
3. Portfolio Review 2016, System Performance, Availability and Reliability Trend Analysis (SPARTA), Northumberland, UK, 2016.
4. ”WInD-Pool: Wind-energy-Information-Data-Pool,” (Online). Available: http://www.wind-pool.de. [Accessed 18 11 2020].
5. GE Aviation, ”GE Aviation launches Configuration Data Exchange to reduce maintenance costs,” [Online]. Available: https://www.geaviation.com/press-release/systems/ge-aviation-launches-configuration-data-exchange-reduce-maintenance-costs. [Accessed 18 11 2020].
6. J. R. Bengt Lydell, ”OPDE—The international pipe failure data exchange project,” Nuclear Engineering and Design, vol. 238, pp. 2115-2123, 2008.
7. Nuclear power plants - Reliability data exchange - General guidelines, International standard ISO 6527, 1982: International Organization for Standardization.
8. K. Dalkir, ”Knowledge Management In Theory And Practice,” Oxford, Elsevier Inc: Jordan Hill., 2005, p. 132–133.
9. H. Hossayni, I. Khan, M. Aazam, A. Taleghani-Isfahani and N. Crespi, ”SemKoRe: Improving Machine Maintenance in Industrial IoT with Semantic Knowledge Graphs,” Applied Sciences, vol. 10, no. 6325, 2020.
10. Mamede, N., Baptista, J., & Dias, F. Automated anonymization of text documents. In 2016 IEEE congress on evolutionary computation (CEC), 2016, p. 1287-1294.
11. Marimon, M., Gonzalez-Agirre, A., Intxaurrondo, A., Rodriguez, H., Martin, J. L., Villegas, M., & Krallinger, M. (2019, September). Automatic De-identification of Medical Texts in Spanish: the MEDDOCAN Track, Corpus, Guidelines, Methods and Evaluation of Results. In IberLEF@ SEPLN (pp. 618-638).
12. Personal Data Protection Commission, Singapore, ”Guide To Basic Data Anonymisation Techniques”, 2018 (Online). Available: https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Other-Guides/Guide-to-Anonymisation_v1-(250118).pdf?la=en [Accessed 16 02 2021].
13. P. M. Heider, J. S. Obeid, and M. Meystre, ”A Comparative Analysis of Speed and Accuracy for Three Off-the-Shelf De-Identification Tools,” AMIA Summits Transl. Sci. Proc., pp. 241–250, 2020.
14. Kuschner, ”Anonymizing and Sharing Medical Text Records,” Physiol. Behav., vol. 176, no. 3, pp. 139–148, 2017.
15. Goswami, P., & Madan, S. (2017, May). Privacy preserving data publishing and data anonymization approaches: A review. In 2017 International Conference on Computing, Communication and Automation (ICCCA) (pp. 139-142). IEEE.
16. Scaiano, M., Middleton, G., Arbuckle, L., Kolhatkar, V., Peyton, L., Dowling, M., ... & El Emam, K. (2016). A unified framework for evaluating the risk of re-identification of text de-identification tools. Journal of biomedical informatics, 63, 174-183.
17. Domingo-Ferrer, J., Sánchez, D., & Soria-Comas, J. (2016). Database anonymization: privacy models, data utility, and microaggregation-based inter-model connections. Synthesis Lectures on Information Security, Privacy, & Trust, 8(1), 1-136.
18. Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. Lingvisticae Investigationes, 30(1), 3-26.
19. Mao, J., & Liu, W. (2019). Hadoken: a BERT-CRF Model for Medical Document Anonymization. In IberLEF@ SEPLN (pp. 720-726).
20. Yogarajan, V., Pfahringer, B., & Mayo, M. (2020). A review of automatic end-to-end de-identification: Is high accuracy the only metric?. Applied Artificial Intelligence, 34(3), 251-269.

HICHAM HOSSAYNI R&D Engineer on Semantic Web & Industrial IoT at Schneider Electric, Grenoble, France. Previously, he worked in Orange Applications for Business (OAB), Grenoble as Embedded Software Engineer, and in CEA-Leti, Grenoble as a Cryptography and Embedded software R&D Engineer. He received M.S. Eng. in networks and distributed systems from ENSEIRB-Bordeaux (France) and ENSIAS-

Rabats (Morocco), and M.S. in Cryptography, Security and Information Coding from Joseph-Fourier, Grenoble. His current research interests are Internet-of-Things, Semantic Web, Data Mining and Cloud & Edge computing. (hicham.hossayni@se.com)

IMRAN KHAN, Senior Member IEEE, is currently working as Senior Principal Architect at Schneider Electric France, leading Ontology Program for the IoT Platform of Schneider Electric. Previously, he worked as Innovation Project Leader working on efficient data and information management in Industrial IoT space. He received Ph.D. degree from Institut Mines-Télécom, Télécom SudParis jointly with UPMC Paris VI, France, M.S. degree from M.A. Jinnah University, Pakistan and B.S. degree from COMSATS Institute of IT, Pakistan. During his Ph.D., he worked as collaborating researcher at Concordia University, Montreal, Canada to lead a 3 year Cisco funded project. He was also involved in several European research projects funded by ITEA2 and H2020. During M.S., he was member of Center of Research in Networks and Telecom (CoReNeT) and worked on projects funded by French Ministry of Foreign Affairs and the Internet Society (ISOC). He has number of publications in peer reviewed conferences & journals and patents. He has also contributed to IETF standardization activities. His current research interests are IoT, Knowledge Graphs, Data & Information Management, Cloud & Edge Computing and Intelligent Systems. For additional details: http://www.imrankhan1984.com, (imran@ieee.org).

NOEL CRESPI Prof. Noel Crespi holds Masters degrees from the Universities of Orsay (Paris 11) and Kent (UK), a diplome d'ingénieur from Telecom ParisTech, and a Ph.D and an Habilitation from UPMC (Paris-Sorbonne University). From 1993 he worked at CLIP, Bouygues Telecom and then at Orange Labs in 1995. He took leading roles in the creation of new services with the successful conception and launch of Orange prepaid service, and in standardization (from rapporteurship of the IN standard to the coordination of all mobile standards activities for Orange). In 1999, he joined Nortel Networks as telephony program manager, architecting core network products for the EMEA region. He joined Institut Mines-Telecom SudParis in 2002 and is currently Professor and Program Director at Institut Polytechnique de Paris, leading the Service Architecture Lab. He coordinates the standardization activities for Institut Mines-Telecom at ITU-T and ETSI. He is also an adjunct professor at KAIST (South Korea), an affiliate professor at Concordia University (Canada), and a guest researcher at the University of Goettingen (Germany). He is the scientific director of ILLUMINE, a French-Korean laboratory. His current research interests are in Data Analytics, the Internet of Things and Softwarization. http://noelcrespi.wp.tem-tsp.eu/, (noel.crespi@telecom-sudparis.eu).