Smart Grids: Challenges of Processing Heterogeneous Data for Risk Assessment

Michael Pacevicius[*a, b], Davide Roverso[a], Pierluigi Salvo Rossi[c] and Nicola Paltrinieri[b]

[a] eSmart Systems, Halden, Norway

[b] Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology NTNU, Trondheim, Norway

[c] Kongsberg Digital, Trondheim, Norway

Abstract: Recent advances in IT-related fields are opening up a broad range of novel applications. This is especially true in the energy sector, where Smart Grid solutions are offering new opportunities for the monitoring of power transmission and distribution in electrical grids. However, optimal use of potentially accessible data sources is challenging, and most of the current Smart Grid projects continue to exhibit suboptimal utilization of heterogeneous information. This situation is also faced when it comes to the assessment of risks associated to operation of electricity transmission and distribution networks. As a consequence, current management systems fail to provide accurate estimations of risk levels in real-world situations. Our paper addresses this issue and contributes to the identification of possible solutions. The paper identifies a number of heterogeneous data sources which could be relevant for risk assessment, but which are currently not fully exploited. Furthermore, the paper points to valuable relations existing across these data sources, that promote a better understanding of real-world situations and empower a more accurate analysis of well-known or newly identified risks within the framework of Smart Grids.

Keywords: Risk assessment, Smart Grids, Heterogeneous Datasets, Data Acquisition, Link Identification.

1. INTRODUCTION

Digitalization offer numerous advantages thanks to the progress made in the IT field (e.g., increase of memory capacities, increase of processing capacities, development of cross-technologies collaboration platforms). For instance, it enables to interconnect devices, to access them through multiple communication networks in real time, and to handle the large quantity of data they generate and transmit. “Smart Grids”, which represent an example of this idea application in the energy sector, offer interesting perspectives for the management of power grids.

There may be several benefits of Smart Grids, such as: higher demand response with minimized costs, reduction of the environmental impact and integration of renewable energy resources, and resilience to disturbances as well as electrical stability in the grid.

Smart Grids may also enable system operators to reduce outage risks by getting access to previously unconsidered data, ranging from weather forecasts to social network data. Combining outage reports with weather reports could for example improve risk monitoring in regions with harsh climatic conditions. This approach has been previously explored in some projects [1], [2], whose focus has especially been set on the impact of climatic conditions. Unfortunately, current researches do rarely fully exploit the real precision degree offered by todays IT solutions and services, and many projects still use averaged data and meta-data for their analyses. Similarly, the full range of various data sources, from which numerous datasets are made accessible thanks to open-access policies (e.g., *OpenAire*, *ENTSO-E Transparency Platform*, *U.S. Open Data platform*), is far from being optimally exploited, especially when it comes to risk assessment.

[*] michael.pacevicius@esmartsystems.com

This challenge is demonstrated by three main arguments:

1. The available frameworks and standards from industrial risk management (e.g., *CSA Q850-97*, *ISO 31000:2009*, *NORSOK Z-013*) are generally overlooked when it comes to the study of Smart Grids. Furthermore, only a few studies propose effective solutions enabling, based on internal as well as external factors, dynamic updates of a risk management framework [3]. This reduces the advantages that can be derived from actual digitalization and hinders the exploitation of the real-time feature offered by Smart Grids.

2. Although Probabilistic Safety Assessment (PSA) approaches advocate for the constitution of cross-disciplinary expert teams (which may include systems analysts, PSA specialists, operators and operational analysts, data scientists and human factor analysts [4]), risk evaluation is mainly based on collections of operating and maintenance data, without taking advantage of additional accessible data. In fact, enlarging the horizon of the observations, in association with new data sources, could allow controllers to detect, observe and potentially predict slow, long-term and non-trivial phenomena (e.g., mechanical fatigue, corrosion, dust accumulation) increasing the failure probabilities.

3. The lack of cross-disciplines experts hinders decision makers from identifying relevant links between data sources, compromising the recognition of efficient combinations of data sources.

There is thus need for new methods enabling continuous and effective integration of heterogeneous data for accurate risk assessment predictions. For this reason, this paper focuses on the first phases of Smart Grid dynamic risk management: collection and combination of relevant datasets. It is constructed as follows. Following the present introduction, the second section reviews the main concepts defining the Smart Grid technology. It describes Smart Grids, mentions the role of the mains organizations, committees and work groups focusing on the topic, and lists the main challenges faced. The third section focuses on the risk dimension in the context of Smart Grids. It shows how the utilization of Smart Grid technologies affects the risk level in the context of power grid management. It continues by pointing out the role the treatment of heterogeneous datasets can have in this context and underscores the importance of the *Common Information Model* (CIM). It finally brings forward work initiated by different research groups on the use of heterogeneous data for risk assessment in power grids. Section 4 describes the investigation procedure. It reminds the main objectives of our work and shortly describes the principles followed to reach our goal. Section 5 describes the results of this paper. It reports a list of diverse data sources newly identified as being relevant for the enhancement of risk assessment, and provides knowledge for merging this information. Section 6 (discussion) highlights the way forward for risk reduction in Smart Grids, but also provides understanding on what remains to be done in the field. The last section (conclusion) shortly reviews the paper and finally concludes our work.

2. OVERVIEW OF SMART GRIDS

The emergence of the first automatic meters and the broadening of SCADA (Supervisory Control And Data Acquisition) technologies in power grids represent the first examples of Smart Grid concepts. Since the creation of the *European Smart Grids Technology Platform* by the European Commission [5], Smart Grids have gained increasing attention and represent nowadays a well-known topic.

Different definitions are given to Smart Grids by scientific literature and industry [6]–[9]. However, they can be summarized as follows:

The basic concept of a Smart Grid is to optimize the production, transmission and consumption of power and information between the different elements (devices & actors) involved in it. For this purpose, it makes use of the progresses made in the information and communication technology (ICT) sector, as well as those made in the industry of electronic components. Thanks to these advances, it is possible to integrate modern technologies (e.g., photovoltaic panels, batteries, smart meters and other types of devices made accessible through the use of connected sensors) into an existing power grid. This eases and automates management for a number of tasks, such as production coordination, energy distribution and power consumption.

A broadening range of communication channels (internet, satellite, etc.) enable real-time access to the large amount of generated data. The high-resolution picture of interconnected devices that is given empowers improved analyses, finer forecasts of future consumption trends and better predictions of potential outages.

The utilization of low-carbon alternatives for power production (solar panels, wind turbines, water turbines, etc.) and the expanded utilization of batteries have led to a shift from a centralized to a distributed repartition of the production sites. The gained autonomy has, in combination with the intelligence integrated in the power grid, boosted the emergence of micro-grids, which are capable to be islanded and reconnected ("split & merge") to external parts of the network without major disturbances.

Micro-grids present multiple advantages: (1) they provide flexibility in case of outages and enable to maintain power supply during programmed maintenance tasks; (2) the technologies utilized can rapidly be implemented and assimilated; (3) these same technologies can be used for shaving peaks under normal circumstances by compensating the needs of one another. In the same idea, Smart Grids have also the capacity to expand and assimilate new items without hardly impacting the general stability of the grid ("integrate as you grow"). This is enabled by the flexibility acquired, which has led to a dynamization and a decentralization of the decision making. As a consequence, end-users are more involved and share henceforth management responsibilities with power companies, to which they are now capable to smartly and easily provide positive and negatives feedbacks. End-users do not anymore only represent passive consumption nodes, but they become active actors and turn themselves into deciding prosumers (producer-consumer). This completely disrupts the flow management, especially because old infrastructures and protection systems – which were originally conceived to be unidirectional – have now to handle reverse flows of energy. In addition to the challenging multi-directional aspect of power management, multiple other stumbling blocks can also be enumerated when it comes to Smart Grids. A list of the most obvious ones is given in table 1.

Table 1: Most Important Challenges in Smart Grids

N°	Stumbling Blocks
1	The collaboration between the main stakeholders and the interconnection of multiple platforms/technologies can be hindered for compatibility reasons or because of conflicts of interest.
2	Although the integration of new technologies enables to compensate some of the fluctuations formerly existing in the power grid, such technologies can also themselves be responsible for new type of peaks. High spontaneous variations of power flows (e.g. because of clouds passing over a solar panel or sudden reduction of the wind intensity in a windfarm) might lead to failures in an aging infrastructure not conceived for such drastic fluctuations.
3	The data made accessible to the customer might lead them to change their habits into unknown ways, which can imply unexpected variation and calls for new profiling analyses.
4	New types of variations might reduce the efficacity of older fault detection methods, localization methods and temporary proxy methods, which are based on assumptions of more stable behaviors in the grid.
5	Smart Grids projects encounter classical *big data* challenges [10] and considerable analytical efforts need to be done in order to transform *data* into *information*. The lack of adapted tools often leads to the ignoring of data because no proper method could enable them to be integrated [11].
6	Real-time or near-real-time access and analysis of data still concerns only a restricted number of information sources. Much information remains post-transmitted and post-processed with more or less high latency (from hours [e.g. with nightly updates of servers] to years [e.g. with the inspection for maintenance of distant power lines]).
7	Optimization techniques for the management of Smart Grids still need to be improved, especially because it is challenging to handle the exponentially growing number of assets integrated in the grid. Finding an optimum between production, transportation, storage and consumption of power in a distributed network is known to be a complicated task.
8	The concept of Smart Grids is closely related to the concept of smart cities, where the management and optimization of water grids, gas pipelines and heating/cooling networks is similarly challenging. This might reinforce every one of the previously highlighted challenges.

To address these numerous challenges, different work groups and committees have been constituted all over the world. Those can exist on global level (North America[1], European Union[2], Asia-Pacific Economic Cooperation[3] or even on larger level[4,5]), but many countries (Korea, Japan, Norway, France, Germany, etc.) do also have national energy regulation committees and Smart Grids associations supporting research in the field. These different organizations contribute to the spreading of knowledge around the topic of Smart Grids, by providing technical reports, protocols, knowledge about standards and by favoring open access demarches and cross-vendor-compatible solutions. The *Smart Electric Power Alliance* (SEPA) provides for example an interesting and regularly updated Catalog of Standards (CoS) on their webpage[6], which enables one to rapidly be aware of the most relevant standards in the field of Smart Grids. The listed standards emanate from standardization organizations and research institutions (e.g., the *Electric Power Research Institute* (EPRI), the *National Institute of Standards and Technology* (NIST), the *International Electrotechnical Commission* (IEC), the *European Committee for Standardization* (CEN), the *European Committee for Electrotechnical Standardization (*CENELEC), the *European Tele-communications Standards Institute* (ETSI)), which permanently provide updates for old and new standards.

[1] North American Electric Reliability corporation (NERC) Smart Grid Task Force
[2] EU Smart Grids Task Force
[3] Energy Smart Communities Initiative (ESCI)
[4] International Smart Grid Action Network (ISGAN)
[5] Global Smart Grid Federation (GSGF)
[6] https://sepapower.org/knowledge/catalog-of-standards

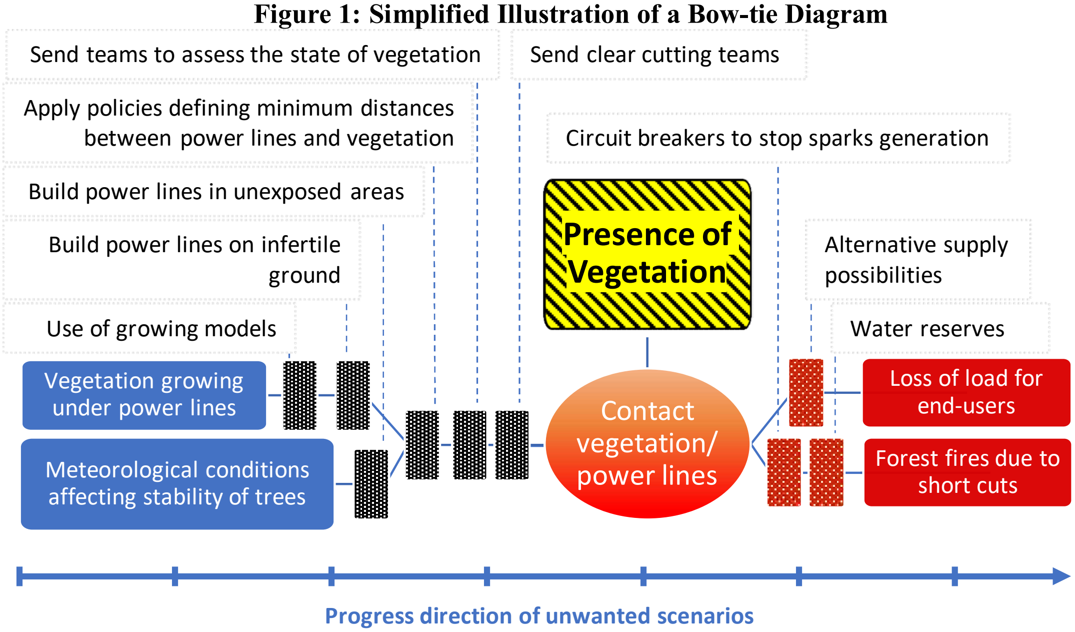
3. RELEVANT RISKS & RELATED INFORMATION

3.1. Risks in the Power Sector

Kaplan and *Garrick* [12] define the general notion of risk by characterizing the outcome of a specific action. More especially, they consider risk as a triplet of (1) a potential unwanted event, (2) its likelihood of occurrence and (3) the consequences of this event happening.

The veracity of the three dimensions depicting risk is dependent on the level of knowledge of a specific situation one may have at a specific moment. This defines the notion of uncertainty, which highlights the incapacity to perfectly characterize a scenario (some events are not anticipated; some probabilities are inadequately trusted; and some consequences are not foreseen). The process of risk assessment aims to reduce this uncertainty by periodically reviewing risky situations and suggesting barriers. Therefore, it uses tools such as the bow-tie diagram [13], which allow for identification and characterization of possible scenarios associated to a hazard [3]. Figure 1 schematically represents the principle of such a diagram. It shows how the hazard loss of control – also referred to as a critical event – can be prevented or mitigated by appropriate countermeasures. For this, it focuses on the risk lead by vegetation growing in the neighborhood of power lines, lists potential threats and consequences, and enumerates existing barriers.[1]

Figure 1: Simplified Illustration of a Bow-tie Diagram



Multiple standards (e.g. *CSA Q850-97*, *ISO 31000:2009*, *NORSOK Z-013*)[2] support the process of risk assessment by providing frameworks and guidelines depicting the best practices in this field. These represent a solid basis for companies that aim to reduce the impact of unwanted and uncontrolled factors on their business.

Depending on what is at stake, power industry risk is expressed in different and often partial ways. For instance, potential socio-economical costs for stakeholders in the case of adverse event provide inputs for decision makers [14]. However, they only address the consequence factor from a financial perspective. Similarly, other methodologies focus on field-oriented indicators by evaluating the physical consequences of an event on the different assets present in the grid [2].

[1] The number of threats, barriers and consequences has been limited for consistency.
[2] An extended list is available at www.ntnu.edu/ross/info/standards

Different types of management systems (e.g., energy management systems, asset management systems, outage management systems) have been set up to minimize the occurrence and consequences of negative outcomes, as well as to ensure an optimal demand/supply-storage balance. These tools are related and affect each other's performance. Data management systems reflect the infrastructure life-cycle and are the core information sources to consider in risk monitoring [3]. They allow creating proxies to assess the performance of such countermeasures.

3.2. Shift to Smart Grids

Guidelines and practices underscore the importance of business knowledge, identification of environmental factors affecting the integrity of facilities, context knowledge of occurred events, information communication, continuous update, and, most importantly, risk reduction. In that sense, Smart Grids represent a great opportunity. Improved information transmission and automatized recovery processes aim to reduce frequency and duration of outages.

Smart Grids imply potential integration of new barriers as well as proxies. Nevertheless, in the current situation, grid companies face compatibility problems and monitoring challenges. Especially, it is difficult to integrate different systems, processes and datasets.

Moreover, Smart Grids may represent further vulnerabilities, which lead to the risk of new unwanted scenarios, such as hacking – key-names in the history of energy-related hacking are Stuxnet (Worm), BlackEnergy (Trojan), Industroyer (malware framework), DragonFly (Hacking group). As the privacy of customers represent a main concern due to the increase of hacking probability, a lot of attention has been given to cyber-security and cyber-attacks when it comes to risk management of Smart Grids.

3.3. The Common Information Model (CIM)

The *Common Information Model* (CIM) addresses the topic of intersystem communication. It was developed by EPRI in the 90's and aims to enable exchanges of information between the different systems involved in a power grid. It is an object-oriented standard – based on *Unified Modelling Language* (UML) – which aims to represent the different objects of interest, as well as their relations in terms of electric generation, distribution, transmission [15]. The development of this model is ensured by the *IEC Technical Committee 57*, especially through the development of the standards *IEC 61970: Common Information Model (CIM)/Energy Management*, and *IEC 61968: Common Information Model (CIM)/Distribution Management* which are also closely related to the standard *IEC 61850: Power Utility Automation*, focusing on the communication networks and systems in substations and power utility systems. The book of *Uslar et al.* "The Common Information Model CIM" [6] as well as the website of IEC[1] and the one of the CIM User Group[2] provide an interesting basis for anyone looking for additional documentation on the topic.

The broadening of CIM among stakeholders in the power industry has been facilitated in 2009, as the *European Networks Transmission Operators – Electricity (ENTSO-E)* approved CIM as a standard exchange format. As a broadly accepted model in the field of energy management, the CIM became an important element for the digital connection between datasets. However, acquiring the knowledge enabling to obtain information out of the data lake created remains a challenging task. Moreover, modelling, integration of information, model implementation and result interpretation still require important research progress to be effective.

[1] www.iec.ch/smartgrid/standards
[2] http://cimug.ucaiug.org

3.4. Related Work

Most of the methods currently used for risk assessment are based on the N-1 criterion [1], [2], [14], [16]–[19]. To paraphrase the sources referenced, this means that the companies in charge of the power transmission or distribution have to be able to ensure full power delivery, even if a main element of their network came to fail. The main downside of such deterministic methods is the lack of flexibility. Fine variations of uncertainty may be disregarded, while binary behavior may be represented by considering all events as equally likely.

Condition based risk assessment and other probabilistic methods represent interesting alternatives. As these methods base their efficiency on the amount of input data, they have logically suffered from poor computing performances of past IT technologies. Recent ICT evolutions led to new relevant projects [2].

Several EU projects (e.g. *AFTER*, *iTesla*, *Umbrella*, *HyRiM*) have focused on the topic of risk in electrical grids. *GARPUR*, one of the most recently completed project (November 2017), focused on asset management policies and outage scheduling assessment. Table 2 lists the main recommendations and calls for new research reported in the different deliverables of the project[1].

Table 2: Main Recommendations/Calls for Research of the *GARPUR* Project

N°	Recommendations/Calls for Research
1	Collect more asset-related and outage-related data (**failure rates**, degradation rates, estimation of expected life time, outage data, context dependent outage data, **costs, consequences**, impact of pre/post-event maintenance tasks, **restoration rates, failures of corrective controls/actions**) to improve currently applied models.
2	Add new types of data if the ratio (complexity added/information gained) is interesting; evaluate how these can be clustered with existing datasets.
3	Diversify imagined scenarios; better anticipate future challenging situation; consider large scale as well as more located scenarios (relevant in the context of micro-grids).
4	Increase the knowledge on smart technologies newly integrated into the grid; analyze the new features they enable; examine the way they interact with other devices; evaluate the way they should be modelled.
5	Consider environment-affected variable failure rates to better model the effects of slowly evolving conditions (e.g. pollution, dust, corrosion, vegetation).
6	Improve the quality and the availability of the data to gather.
7	Improve the coordination between Transmission System Operators (TSO's); Improve the coordination between Distribution System Operators (DSO's); Improve the coordination between TSO's and DSO's.
8	Sustain the definition of standards; improve the knowledge-sharing of standards.
9	Determine the value of potentially unreliable data (in the context of reliability metrics) by applying sensitivity analyses.
10	Define frameworks and guidelines for collecting, maintaining, and sharing data and models
11	Deepen the research for the automatization of processes related to risks by using machine learning techniques.
12	Sustain a gradual development, implementation and testing of methods relevant to the definition of a framework, rather than to wait for a final holistic solution to be proposed.
13	Develop new tools and sustain the formation of cross-disciplinary experts capable of providing risk-oriented analyses of heterogeneous datasets.

One of the initial and core topic to address is the access to data. As it is also further stated in the literature [10], [16], [17], [20] there is need for research on what type of datasets to gather, as well as a need for clarification on the way to intelligently combine data sources for the improvement of existing models in the field of risk assessment. Nevertheless, no clear guidelines have been found on the way to merge a large number of heterogeneous sources of information for improved risk assessment in the context of power grids - and more especially in the context of Smart Grids.

[1] www.sintef.no/projectweb/garpur/deliverables

The lack of maturity of methods focusing on degradation processes over long time periods and focusing on impact evaluation (positive-negative) of maintenance activities on components life-time highlight furthermore the need for more research in the field of probabilistic methods [1]. In order to provide such knowledge, research needs to focus on the nature of the relations existing between various phenomena and components' behavior to better consider obvious links (e.g. effects of storms on power lines) as well as slighter links (e.g., effects of dust, rust, and pollution on insulating properties) [10] in risk assessment and monitoring.

4. INVESTIGATION PROCEDURE

Considering the situation described and the challenges faced, we aim, with this work, to reduce the gap existing regarding risk assessment in Smart Grids. For this reason, we identify and report a list of data sources that can be used to better characterize risks. Moreover, we intend to highlight the links existing between the different sources in order to understand how the data should later be aggregated.

A deepened review of storm and outage reports, as well as an intensive research among the existing literature and among online websites of power management stakeholders has allowed identifying main elements and factors involved in the emergence of risks, outages and accidents in power grids. This research has enabled to identify main categories of directly related data sources, as well as tools that are used to reduce the severity of such outcomes.

An additional analysis has ultimately consisted in identifying the data sources that could have increased the knowledge level in one of the negative situation encountered, by providing identical, improved or complementary information, compared to the data initially considered.

5. RESULTS

Data and services initially defined for a specific purpose may lead to additional interesting features [10]. For instance, correlations among datasets enable creation of proxies and finer situation understanding. Table 3 provides a non-exhaustive list of data sources that could be used for more effective risk assessment. It illustrates the benefits of such information when combined with asset management and daily monitoring data from the grid, such as:

- grid topology & asset information (age, location, failure rates, initial life-time models, etc.),
- inspection and maintenance reports (preventive/corrective approaches),
- outage reports (context dependent outage data, consequences, costs),
- customer feedback and crew management decisions,
- power-flow forecasts & real-time power flow measurements in the grid,
- past, present and future design documents of the power grid.

The two first columns of table 3 identify the field of information considered; the third column highlights a relevant use case and proposes an access to data; and the last column suggests additional datasets to use for supporting the use case highlighted.

Table 3: Useful Sources of Information for Improved Risk Assessment

N°	Topic	Use Case	Add. Dataset
A	Vegetation	Anticipate risk induced by trees along power lines using growth models, health analysis and clearcutting reports. **Data access:** https://kilden.nibio.no	B, C, E, F, I, J, Q
B	Meteorological data, lightnings	**Historical data:** Identify risky areas (wind, ice-rain, etc.) for the planification of power grids - **Real-time data/forecasts:** Team management & material supply before and after events (snow, storms, etc.). **Data access:** www.yr.no, https://api.met.no, www.senorge.no	A, C-F, M, N, P
C	Satellite images	Enable real-time & post-event situation analysis; identify potential location of outages. **Data access:** https://open.nasa.gov, https://scihub.copernicus.eu	A, B, D, F, M, N
D	Region dependent disaster data	Evaluate risk level due to landslides, avalanches, floods, earthquakes, hurricanes, etc., for the planification of power grids. **Data access:** https://atlas.nve.no	A-C, E, F
E	Climate change forecasts	Anticipate snow melting for dams' reserves; Anticipate pests migrations. **Data access:** https://climate.copernicus.eu/seasonal-forecasts	B, C, F, I, M, N, Q
F	Terrain model, mineral composition	Anticipate evolution of streams and random water flows during floods; prevent short-circuits due to salinity. **Data access:** https://hoydedata.no/	A-E, J, N, P, Q
G	Forest fires	**Historical data:** Identify risky areas for the planification of power grids - **Real-time data/forecasts:** Team management & material supply during and after events. **Data access:** http://gwis.jrc.ec.europa.eu/	A-C, E, F, H, I, M, N, Q
H	Pollution, dust, smoke	Evaluate short-circuit probabilities and degradation due to aggressive chemicals [18], [21]. **Data access:** www.environment.no	B, C, M, N, P, Q, S
I	Animals, insects & fungi	Predict species migrations to estimate and respond to risk induced by animals coming in contact with electrical components (e.g. birds, squirrels) [22] or by pests deteriorating the infrastructures. **Data access:** https://gd.eppo.int/	A-C, E, H, M, N, P, Q, S
J	Laws & standards	Anticipate security requirements. **Data access:** List of standards, national/international regulations	B, D, E, H, K, L, O, Q-S
K	Union trade contest	Anticipate strikes. **Data access:** Internal data	J, L, M, O, R, S
L	Security and surveillance	Protection against potential espionage/sabotage/terrorism [16], [21]. **Data access:** Internal data, external audit reports, national security reports	C, K, M-R
M	Social networks	Anticipate risks due to social contest (organization of protest-events), detect/locate outages with clusters of dissatisfaction-hashtags, manage emergency-related outages (live messages, pictures and videos). **Data access:** Public feed API (Facebook), PowerTrack API (Twitter)	J-L, N, Q-S
N	Collaborative platforms	Increase risk-related knowledge thanks to voluntary contribution of anonym public. **Data access:** www.regobs.no	B-D, M, Q
O	Market, finance	Anticipate partners/customers behaviour due to market price fluctuations. **Data access:** Internal/partner/customer data	B, G, J-M, R, S
P	Simulated environment	Improve risk models thanks to simulated environments and simulated data. **Data access:** Internal/partner/customer data	B, D, F, Q
Q	Open-access data sources	Improve company's knowledge thanks to open-data politics of governments and external companies. **Data access:** https://open-power-system-data.org, https://rte-opendata.opendatasoft.com, https://transparency.entsoe.eu, www.data.gov	/
R	Geo-political information	Anticipate variations of power costs and consumption due to geo-political conflicts (e.g., Europe and its dependency to Russian gas). **Data access:** Various newspapers & other media	J, L, M
S	Other energy related data	Anticipate variation of power costs and consumption due to outages/problems in other energy utilities (e.g., explosion on a gas pipeline during the winter which leads to a pic of power consumption [21]). **Data access:** Internal/partner data, media	B-G, J-O, Q, R

6. DISCUSSION

The results highlight the plurality of data sources that can actually be relevant for risk assessment. Doing so, they enable to imagine the diversity of existing possibilities for the creation of proxies; which is a main asset for the increase of resilience in a power grid. The sample of suggested scenarios also highlight the plurality of the applications a data source can be useful for. Future work will need to focus on additional scenarios and need to look for more links across the datasets. This implies cross-disciplinary teams and the possibility to access and exchange datasets among the principal stakeholders.

It should be mentioned that the order in the table does not reflect the importance level of the data sources for the estimation of risk level, which is a matter of geographical localization and contextual configuration. Moreover, for the purpose of this work, a focus has been set on data relevant for a Norwegian power company. Some of the datasets suggested are thus not directly relevant for companies located outside of Norway and additional sources of information would have to be found in such circumstances. Finally, it is known to the authors that there is already room for optimization in the manipulation of the data used for asset management and daily monitoring. As, in such a case, the focus has to be set on the model and not on the access of data – which goes beyond the scope of this paper – this task has not been further deepened here.

Even if some online sources of information are known to be relatively less reliable (especially social networks and collaborative platforms) we stress that such sources should not be ignored for risk assessment, but rather integrated with precautionary methods enabling to estimate the value of the information before adding it into the models.

The manual combination of datasets described in this work is a pre-step for automated merging, which is relevant in the context of machine learning. However, further work still needs to be addressed before accessing to this automatization phase. A focus has especially to be set on aggregation models, which will enable to provide the first outputs directly utilizable by decision makers for the management of the power grids. This implies nevertheless to overcome many remaining challenges (variety of temporal resolutions/boundaries, variety of spatial resolutions/boundaries, restriction policies for data access, etc.) and calls for further research regarding data management solutions in Smart Grids.

7. CONCLUSION

Smart Grids have shown to be particularly promising. They come nevertheless with numerous challenges, from which we listed the most important ones. In this paper, we mentioned multiple relevant projects and publications focusing on risk assessment in modern power grids. We also brought forward a wide range of types of data sources that should be considered in this context and suggested different scenarios utilizing such sources of information to reduce the global level of risks in electrical grids. We finally highlighted which data sources could be combined to obtain more accurate estimations of actual risk levels existing in a real-world situation in the framework of Smart Grids.

With our work, we provide pieces/elements for the construction of a framework supporting better decision making in the management of risk in modern power grids. In order to provide actionable intelligence, there is nevertheless a need for better sharing of best practices and for better sharing of data. This will enable the principal stakeholders to get a more accurate overview of their infrastructures and so to better make decisions when it comes to risk management in their power grids.

Acknowledgements

This work is part of the project "*Dynamic risk management for Smart Grids in large-scale interconnected power systems.*" funded by *eSmart Systems* and the *Norwegian Research Council* (NæringsPhD program - Project No.: 276404), which the authors would like to thank for their support.

References

[1] M. Rekinger, M. Hofmann, A. Chaouachi, E. Karangelos, L. Wehenkel, and G. Kjølle, "A Transition Roadmap Towards Probabilistic Reliability Management," 2017.

[2] S. A. B. De Almeida, R. Pestana, F. P. M. Barbaosa, and N. Machado, "Operational Risk Assessment Methodology for the Portuguese Transmission System" *Electra*, no. 252, p. 11, 2010.

[3] N. Paltrinieri and F. Khan, *Dynamic Risk Analysis in the Chemical and Petroleum Industry: Evolution and Interaction with Parallel Disciplines in the Perspective of Industrial Application*. Elsevier, 2016.

[4] O. Nusbaumer, "Introduction to Probabilistic Safety Assessments (PSA)." p. 20, 2010.

[5] ETP SmartGrids, *European Technology Platform Smart Grids: Vision and Strategy for Europe's Electricity Networks of the Future*, vol. 19, no. 3. 2006.

[6] M. Uslar, M. Specht, S. Rohjans, J. Trefke, and J. M. Vasquez Gonzalez, *The Common Information Model CIM*, vol. 66. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[7] R. W. Y. Habash, V. Groza, D. Krewski, and G. Paoli, "A Risk Assessment Framework for the Smart Grid," in *2013 IEEE Electrical Power & Energy Conference*, 2013, pp. 1–6.

[8] National Institute of Standards and Technology, "NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0," 2014.

[9] S. Borlase, *Smart Grids: Advanced Technologies and Solutions*, Second. CRC Press, 2017.

[10] S. R. Khuntia, J. L. Rueda, and M. A. M. M. Van Der Meijden, "Smart Asset Management for Electric Utilities: Big Data and Future," in *WCEAM 2017*, 2017.

[11] L. Al-Hakim, *Challenges of Managing Information Quality in Service Organizations*. Idea Group, 2007.

[12] S. Kaplan and B. J. Garrick, "On The Quantitative Definition of Risk," *Risk Anal.*, vol. 1, no. 1, pp. 11–27, 1981.

[13] C. Delvosalle, C. Fievez, A. Pipart, and B. Debray, "ARAMIS project: A comprehensive methodology for the identification of reference accident scenarios in process industries," *J. Hazard. Mater.*, vol. 130, no. 3 SPEC. ISS., pp. 200–219, 2006.

[14] L. Wehenkel, M. Hofmann, A. Vafeas, E. Karangelos, and E. Jordanger, "Results and recommendations towards stakeholders," 2017.

[15] E. C. for E. S. (CENELEC), *Energy Management System Application Program Interface (EMS-API) - Part 301: Common Information Model (CIM) Base (IEC 61970-301:2011)*. 2011, p. 253.

[16] S. Perkin *et al.*, "Modelling weather dependence in online reliability assessment of power systems," *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, vol. 231, no. 4, pp. 364–372, 2017.

[17] V. V. Vadlamudi, C. Hamon, O. Gjerde, G. Kjolle, and S. Perkin, "On Improving Data and Models on Corrective Control Failures for Use in Probabilistic Reliability Management," in *2016 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2016 - Proceedings*, 2016.

[18] N. Machado, S. A. De Graaff, and R. Pestana, "Risk assessment methodology – Running tests at the Portuguese TSO," in *CIGRE 2014*, 2014.

[19] O. Gjerde, A. Vafeas, C. Badajoz, S. Dourlens-Quaranta, and S. H. Jakobsen, "Final Conference Proceedings," 2017.

[20] TECHNOFI, "Technical analysis of on-going projects," 2015.

[21] NERC, "State of Reliability 2017." p. 208, 2017.

[22] NRECA, "Animal-Caused Outages," 2005.