A Survey on Next-generation Power Grid Data Architecture

Shutang You[1], Lin Zhu[1], Yong Liu[1],
Hesen Liu[1], Yilu Liu[1,2]
1. Dept. of Electrical Engineering and Computer Science
The University of Tennessee, Knoxville
Knoxville, TN, USA

Mallikarjun Shankar[2], Russell Robertson[3], Tom King[1,2]
2. Oak Ridge National Laboratory
Oak Ridge, TN, USA
3. Grid Protection Alliance
Chattanooga, TN, USA

***Abstract*—The operation and control of power grids will increasingly rely on data. A high-speed, reliable, flexible and secure data architecture is the prerequisite of the next-generation power grid. This paper summarizes the challenges in collecting and utilizing power grid data, and then provides reference data architecture for future power grids. Based on the data architecture deployment, related research on data architecture is reviewed and summarized in several categories including data measurement/actuation, data transmission, data service layer, data utilization, as well as two cross-cutting issues, interoperability and cyber security. Research gaps and future work are also presented.**

***Index Terms*-- Smart grid, data architecture, survey, information system**

I. Introduction

The goal of power systems is to make full use of all the available resources and information in order to achieve the most efficient, reliable, and sustainable supply of electricity. With the development of technology and the increasing concerns of energy and environment, power system development faces both challenges and opportunities. With the increase of load, wind power, solar power, electric vehicles, and other intermittent generation and controllable devices, power system operation is more complex than ever [1]. On the other hand, the information and communication technologies are also providing new opportunities for power systems to be transformed to a flexible system with two-way energy and information flows, which relies on collecting, processing, and utilizing data at right speeds and scales [2].

Data architecture describes how data is gathered, stored, and used in information systems. It is the basis to deploy information and communication systems and to control data streaming. Since the data sources and the applications of power grids are changing with time, the data architecture should have good flexibility and scalability to provide better support for future functions.

This paper briefly analyzes the requirements and elements of data architecture for future power grids. Then related research and projects on power grid data architecture are reviewed. The rest of this paper is organized as follows. Section II investigates the requirements and challenges of future power grid data architecture, and then it proposes an architecture reference framework and describes its function elements. Section III gives a review of the related research and projects on future power grid architecture design. Discussion and future work are provided in Section IV.

II. Data Architecture Framework

A. Data challenges and requirements

Because the increasing number of measurement devices is generating much more raw data, appropriate architecture is needed to transform the data to information and knowledge within the electricity supply chain. Challenges and requirements in designing the data architecture for future power grids are listed in Table I.

TABLE I. Challenges in Power Grid Data Architecture

Challenges	Explanations
High volume	Increasing number of measurement apparatus and systems in the generation, transmission, and distribution levels.
Diversity and heterogeneity	Various formats of data from different devices, as well as external data sources, such as weather, economy, and other energy systems (such as natural gas).
Timeliness	Latency, inaccuracy, and loss of data may significantly increase system risk and operation cost.
Quality	
Computation burden	Large amounts of data may need to be analyzed for decision making within a very short time, such as a close-loop control.

B. Necessary elements of data architecture

Considering the data architecture requirements and challenges, a reference data architecture framework is presented in Fig. 1. Descriptions on its elements are listed as follows.

1) Data collection/actuation: The data collection and actuation layer contains sensors and actuators in power systems. The technical requirements on this part include the accuracy and timeliness for the data acquisition and control actuation processes. Its development trend is the

This work was supported primarily by the Engineering Research Center Pro-gram of the National Science Foundation under NSF Award Number EEC-1041877 and the CURENT Industry Partnership Program. This work was also supported by the U.S. Department of Energy.

standardization of data formats and communication methods, so that the overall architecture is more scalable.

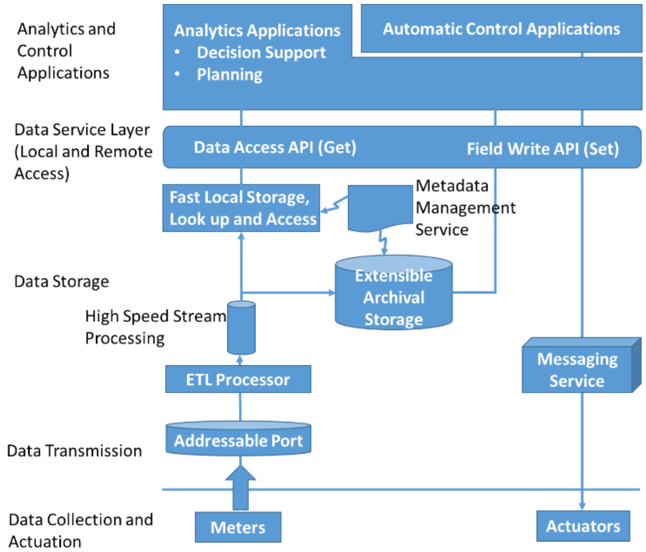

Figure 1. The data architecture reference framework for power grids.

2) Data transmission: Data of power grids are transmitted through wired or wireless networks from sensors to local or regional grid control centers, where data can be utilized. Different technologies have different features, such as costs, bandwidth, security, and latency. There is no one technology that can meet the communication demands in all parts of power grids. The future communication network for power grids tends to be open-standard, secure, and multi-technology based, providing a communication platform for various infrastructures and applications.

3) Data storage: Since data from various resources in power grids are highly heterogeneous, the data architecture needs an Extract, Transform, and Load (ETL) process to format the data for easy storage, querying, and analysis. Real-time data streams need fast data storage like a memory-based database to achieve fast data reading and writing, which is critical for real-time grid applications. Historical data are stored in the archival storage database for non-real-time analyzing purposes. The metadata describes the information of data in power systems, including data organizing information and data content information. It improves the transparency of power grid data sources, by increasing the reuse rates of database interface codes.

4) Data service layer: The data service layer is used to provide data for applications and also for extracting data from applications for actuation. Based on the metadata and generic data interfaces, this layer can have many inherited interfaces to provide specific data services for applications with different requirements. In addition, since this layer implements unified data service management, the bandwidth resource of data buses can be dynamically configured to optimize the data service resources.

5) Analytics and control applications: Due to the deregulation of power markets and the development of distribution generation and demand response, the operation modes of power grid varies more frequently and significantly. This makes the grid operation and control applications more reliant on data with sufficient quality and quantity. These applications, which tend to exist in a modular form to improve scalability, can exchange analysis results between each other, achieving more advanced control and analysis functions.

III. Overview of Grid Data Architecture Research

Based on a common data flow in power grid, this section attempts to summarize the latest research on grid data architecture in the following aspects: data collection, data transmission, data storage, data service layer, and data utilization (e.g., data analytics). In addition, this section also investigates two cross-cutting issues related to grid data architecture: interoperability, and cyber security. An overview of the research projects in this survey is shown in Table II.

A. Data collection

1) Grid Edge: This project focuses on the technology, business, and policy in medium- and low-voltage power grids for the increasing growth of distributed energy, microgrids, and demand response from the customer's perspective [3]. Grid Edge emphasizes the collection and analytic functions of multiple data terminals, e.g., advanced metering infrastructure, especially the data behind meters. In addition, Grid Edge proposes a taxonomy architecture consisting of five layers (the physical asset layer, the network and control layer, the grid-edge application layer, the data analytics layer, and the market evolution layer) to help visualize the various layers and technologies in them.

2) iTESLA: The Innovative Tools for Electrical System Security within Large Areas (iTESLA) project aims to develop a powerful toolbox to support dynamic security assessment and decision making of the pan-European electricity transmission network considering uncertainties [4]. In order to collect and exchange data among different regions, the project defines a set of requirements on external data provided by the different transmission system operators and recommends associated formats to carry the information. This project also provides three open source prototype tools to convert data formats.

B. Data transmission

1) GridStat: GridStat is a data middleware based on a publish-subscribe mechanism, which is designed to deliver real-time operational data of the electric power grid. There are two layers in GridStat middleware: the data transportation layer, and the management layer. The data transportation layer is comprised of routers, which support routing and multicast services. The management layer consists of quality of service (QoS) brokers, which perform resource allocation in the data layer to establish paths that meet the QoS requirements of each subscription request [5]. Recently, GridStat has been integrated into several pilot projects for demonstration and further testing. For instance, GridStat was integrated in GridSim to perform a holistic real-time simulation of power system and communication system in 2011.

2) NASPInet: NASPInet, a project of NASPI (the North American SynchroPhasor Initiative), aims to provide an "industrial grade", secure, standardized, distributed, and scalable data communication infrastructure which supports synchrophasor applications in North America. NASPInet is intended to facilitate the secure exchange of both real-time

TABLE II. OVERVIEW ON POWER GRID ARCHITECTURE RESEARCH AND PROJECTS

Project Name	Objective/Content	Major accomplishments	Focus on data architecture
Grid Edge	· Technology, business, and policy in medium- and low-voltage power grid	· Grid edge taxonomy architecture	· Data collection · Data analytics
iTESLA	· Dynamic security assessment considering uncertainties	· Definition of required external data needs · Conversion tools for format conversion · Uncertainties modeling for security assessments	· Data analytics · Data collection and management
GridStat	· Delivery of power grid operational status information	· Publish-Subscribe based data middleware	· Data transmission
NASPInet	· Data communication infrastructure to support synchrophasor applications	· High-level conceptual synchrophasor communication architecture	· Data transmission
OpenADR	· Standardized and interoperable automated demand response	· An open and interoperable information exchange model for demand response	· Data transmission
GridOPTICS™	· Next-generation concepts and tools for grid operation and planning	· A software framework for power system operations and planning	· Data management · Software framework
INDE	· High performance in smart grid data management	· Accenture Intelligent Network Data Enterprise Platform	· Data management
SERA	· Enables technology innovation and advancements needed to create smart grid ecosystems	· Smart Energy Reference Architecture · Information architecture using Microsoft technologies	· Software framework
SAFE	· Smarter utility solutions across all areas of a utility	· Solution Architecture for Energy and Utilities	· Software framework
NIST framework	· Guidelines for interoperable standards	· Framework and roadmap for smart grid interoperability standards	· Interoperability
IEEE P2030	· Guidelines for smart grid interoperability	· Smart grid interoperability reference model including power system, communication, and information technology architecture	· Interoperability
IEC TC 57	· IEC Smart Grid Standardization Roadmap	· TC 57 reference architecture · Service-oriented architecture	· Interoperability
SGRA	· Provides guidelines for smart grid interoperability	· Smart Grid Reference Architecture	· Interoperability
GridWise	· Establishes interoperability principles	· Interoperability Context- Setting Framework · Smart Grid Interoperability Maturity Model	· Interoperability
IntelliGrid	· Integrates energy delivery systems and information systems	· Power system functions definition and analysis · IntelliGrid Architecture Framework	· Interoperability
GridBlocks	· Integrates information and communication technologies into electric power systems	· Smart Grid Reference Architecture · Cisco GridBlocks Architecture	· Interoperability
SGIP	· Accelerates the implementation of interoperable smart grid devices/systems	· Smart grid logical reference model · Catalog of standards for smart grid	· Cyber security
TCIPG	· Secures low-level devices, communications, and data systems	· Application-aware detection and recovery mechanisms · Protocol definition for secure data transportation · A combined simulation/testbed environment	· Cyber security

streaming data and historical data. It is expected to support both one-tone unicast and one-to-many publisher-subscriber based data sharing in an efficient manner [6]. NASPI also carried out several projects on NASPInet to demonstrate interoperable IEC 61850 90-5 based phasor data exchange over a Wide Area Network (WAN), and IP multicast routing of phasor data across a WAN, etc.

3) OpenADR: Open Automated Demand Response (OpenADR) specifies a set of data models and services for data exchange among utilities, Independent System Operators (ISO) and customer appliances participating in automated demand response programs using a common language and an existing communication infrastructure such as the Internet. Two protocols: hypertext transfer protocol (HTTP) and extensible messaging and presence protocol (XMPP), are supported by OpenADR for different deployment scenarios [7]. Up to now, over 60 utilities and controls vendors have already announced or have deployed OpenADR-based systems.

C. Data service layer

1) GridOPTICS™: GridOPTICS™ or Grid Operation and Planning Technology Integrated Capabilities Suite is a software framework developed by the Pacific Northwest Nation Lab (PNNL) for power system operations and planning [8]. This effort envisions a flexible, scalable software architecture for integrating a range of data collection, analysis, simulation and visualization technologies. This framework would enable plug-and-play of various analysis, modeling and visualization software tools for fast and accurate control of the power grid. In order to bridge the data access for different control purposes, a scalable and thin layer of event processing that hides the complexity of data storage and management is also provided. Recently, several demo applications have been developed based on GridOPTICS™, e.g., graphical contingency analysis, net interchange schedule, and event detection.

2) SAFE: The IBM Solution Architecture for Energy and Utilities Framework (SAFE) is a powerful software platform to extract actionable information from all data sources and to facilitate faster solution deployment and integration [9]. SAFE helps utilities integrate processes and related information throughout a company. SAFE can extract actionable information from all data sources (e.g., facility infrastructure, assets, customers, and workers) to enable new software capabilities related to smart energy utility management. Moreover, this framework ensures that solutions from IBM business partners can be integrated more easily by providing pre-certified compatibility validation.

3) SERA: The Microsoft Smart Energy Reference Architecture (SERA) provides a valuable reference to accelerate development and guide deployments for the smart energy ecosystem [10]. Its holistic architecture design facilitates data and business process integration. The SERA is intended to address prevailing systems and issues in enough detail to be useful, but without so much detail as to be untenable. In addition, SERA can help identify potential products and solutions that enable this architectural vision.

4) INDE: The Accenture Intelligent Network Data Enterprise (INDE) provides a template architecture based on the best practices and accumulated experience. The standards-based reference architecture consists of several components: a sensor network architecture, a data acquisition, transportation, and storage framework, a data integration platform, a visualization platform, a cyber-security framework, and a grid data analytics catalog, etc. [11]. INDE can be considered as an interlayer software between raw grid data and high-level applications to manage massive real-time data and to transform them into useful, actionable information.

D. Data utilization

In the iTESLA project, the data mining methods applicable to the platform have been defined and described, which mainly fall into three modules: basic statistical, machine learning, and density estimation and sampling [4]. The identified methods have been applied to different examples using real data samples. For instance, a system prototype based on IEEE RTS has been established using these methods and the historical data from Marseille and Nancy. Methodologies for performing data mining has been outlined and applied to sampling, topology prediction, and the identification of positive and negative rules.

E. Interoperability

Interoperability is the ability for all components of a system to exchange information and use that information for the correct execution of specified functions. Since standardization is an effective way to achieve the interoperability, many standard development organizations are focusing on this field.

1) NIST framework: The National Institute of Standards and Technology (NIST) proposes a conceptual architecture of smart grid including seven domains (generation, transmission, distribution, customers, markets, operations, and service providers) and describes the data flow and energy flow among these domains to show they are interrelated [12]. The NIST model has been widely recognized since it is a valuable reference for research and industries to identify white-space standards that are highly desired to realize interoperable smart grid.

2) IEEE P2030: Based on the conceptual architecture proposed by NIST, the IEEE P2030 standard provides a smart grid interoperability reference model (SGIRM), which analyzes the interoperability of smart grids from three perspectives: power system, communication technology, and information technology [13]. Under this framework, IEEE P2030 provides interoperability guidelines to integrate the information and communication technology into the energy systems. These guidelines will support the data management and information exchange for various applications in smart grids.

3) IEC communication architecture: This seamless communication architecture is proposed by IEC TC 57 based on several core standards (IEC 61850, IEC 61970, and IEC 61968), which focuses on the communication requirements of applications in the power systems [14]. This architecture contains power utility communication standards including semantic data models, services and protocols. Furthermore, IEC has developed the smart grid standards map giving the relationship between the smart grid elements and over 100 related IEC standards. It can help researchers and stakeholders easily and instantly identify particular standards required for smart grid implementation.

4) Smart Grid Reference Architecture (SGRA): European standard development organizations extend the NIST conceptual model by adding two important elements: distributed energy resource (DER) and flexibility. DER allows addressing the very important role that it plays in the European objectives, while the flexibility concept can group consumption, production and storage together in a flexible entity (still under study) [15]. In addition, the extended European conceptual model indicates that interoperability should be implemented in different layers (Business, Function, Information, Communication and Component), hierarchical levels of power system management, and the complete electrical energy conversion chain.

5) GridWise® Alliance: GridWise® Alliance proposes an interoperability context-setting framework to provide a context for identifying and debating interoperability issues [16]. There are eight interoperability categories related to systems integration and interoperation in the electric industry, which can be divided into three aspects: technical, informational, and organizational. In addition, a smart grid interoperability maturity model is developed by GridWise, which provides a means to measure interoperability maturity for interoperability progress assessment, gap analysis, and interoperability improvement.

6) Intelligrid: The project Intelligrid divides power system functions into six categories to address the business needs of power system operation requirements for today's and future power systems [17]. Intelligrid also defines a communication and information environment including configuration, quality of service, security, and data management requirements in each function categories for interoperability.

7) GridBlocks: The Cisco GridBlocks™ architecture provides a forward-looking view on how the electrical grid can be integrated with digital communications across the entire power delivery chain. The model is a starting point for creating utility-specific designs, and offers guidance on deployment of grid-specific applications [18]. It also lays out a framework for designing and deploying comprehensive management and security solutions across the grid. This will help utilities to lower the total cost of ownership of their communication infrastructure, as well as create additional value by helping to enable new utility services.

F. *Cyber security*

1) SGIP: The Smart Grid Interoperability Panel (SGIP) published the guidelines for smart grid cyber security, which proposes a logical reference model that shows logical interfaces linking actors and suggests the types of information exchanged [19]. SGIP extends each of the seven major domains of the NIST conceptual model to several sub-domains that exchange information and make decisions for smart grids. SGIP also defines the logical interfaces among the sub-domains. These logical interfaces describe where the smart grid should provide security and what technology requirements the smart grid needs to ensure security. Also, a Risk Management Framework is proposed to assess the risk of smart grid components and systems, providing guidance on security practices.

2) TCIPG: TCIPG, or the Trustworthy Cyber Infrastructure for the Power Grid, focuses on the security of data transmission and computation systems in power utilities [20]. It studies the methods to increase the ability of power grid infrastructures to withstand cyber-attacks and other failures due to natural causes or operation errors. TCIPG research covers the following categories: a). wide-area and local-area management, monitoring and control security b) detection, response and management of cyber events. c). methods and tools for assessment power grids' resilience to cyber-attacks and cyber-events. The security research in TCIPG covers different data layers vulnerable to security issues in future power grids.

IV. Summary and Future Work

Current research projects cover many important aspects of power grid data architecture, such as data transmission, the data service layer, interoperability, and network security. Extensive research results have been achieved through multiple conceptual designs, as well as some prototype system development and demonstration. New requirements for the data architecture due to renewable integration and demand response are well-studied. Nevertheless, some research gaps still need to be explored in the future.

(1) Most current research is in the theoretical and methodology research stage. Since a specific technology or an architecture design needs a corresponding use case for verification, practical projects to demonstrate the performance of conceptual designs are necessary. In addition, interoperability experiments are needed to testify if an architecture meets interoperability requirements, and detailed technical implementation guidelines should also be formulated.

(2) Structural requirements for emerging applications like distributed control, real-time closed-loop control remain to be addressed. For instance, designing communication protocols for control message transmission in real-time wide-area closed-loop control is still a challenge.

(3) An open source power grid application development platform can be established based on standard data sources and interfaces. This will help developers to focus more on algorithms and control strategies designing, and not worrying about interfaces and data formats.

Our future work will mainly focus on: 1) Revising and refining the data architecture for the next-generation power grid, based on reference data architecture, to make it accommodate future power grid features and applications. 2) selecting multiple representative applications in different operation and control domains of future power grids, such as state estimation, AGC, data analytics, distributed coordinated control, and wide-area oscillation damping control, and then investigating how to coordinate multiple applications in different domains within a unified data architecture.

References

[1] F. Hassan, "The path of the smart grid." *IEEE Power Energy Mag.*, vol. 8, no. 1, pp. 18-28, 2010.

[2] F. Li, Q. Wei, H. Sun, H. Wan, J. Wang, Y. Xia, Z. Xu, and P. Zhang. "Smart transmission grid: vision and framework." *IEEE Trans. Smart Grid,* 1, no. 2, pp. 168-177, 2010.

[3] Grid Edge - Utility modernization in the age of distributed generation. [Online]. Available: www.greentechmedia.com/gridedge#download

[4] Innovative Tools for Electrical System Security within Large Areas. [Online]. Available: www.itesla-project.eu/

[5] C. H. Hauser, D. E. Bakken, and A. Bose, "A Failure to Communicate: next generation communication requirements, technologies, and architecture for the electric power grid", *IEEE Power Energy Mag.*, vol. 3, no. 2, pp. 47-55, March/April, 2005.

[6] P. T. Myrda, K. Koellner, "NASPInet - The Internet for Synchrophasors", *in Proc. 43rd Hawaii International Conf. on* System Sciences (HICSS), pp. 1-6, Jan. 5-8, 2010.

[7] The OpenADR Primer - An introduction to Automated Demand Response and the OpenADR Standard, [Online]. Available: www.openadr.org/assets/docs/openadr_primer.pdf

[8] GridOPTICS: A Software Framework for Power System Operations and Planning. [Online]. Available: gridoptics.pnnl.gov/docs/FPGI-FA1-gorton.pdf

[9] The IBM Solution Architecture for Energy and Utilities Framework. [Online]. Available: www.ibm.com/smarterplanet/us/en/smart_grid/nextsteps/solution/S410300L69965D41.html

[10] Microsoft's Smart Energy Reference Architecture. [Online]. Available: www.microsoft.com/enterprise/industry/manufacturing-and-resources/power-and-utilities/reference-architecture/performance-oriented-infrastructure.aspx#fbid=RLhf3oZtqA1

[11] Accenture Intelligent Network Data Enterprise. [Online]. Available: www.accenture.com/us-en/Pages/service-intelligent-network-data-enterprise.aspx

[12] NIST framework and roadmap for smart grid interoperability standards, Release 2.0. [Online]. Available: www.nist.gov/manuscript-publication-search.cfm?pub_id= 910824

[13] *IEEE Guide for Smart Grid Interoperability of Energy Technology and Information Technology Operation with the Electric Power System (EPS), End-Use Applications, and Loads,* IEEE Std 2030™-2011. Sep. 2011.

[14] IEC Smart Grid Standardization Roadmap [Online]. Available: http://smartgridstandardsmap.com/.

[15] CEN-CENELEC-ETSI Smart Grid Coordination Group Smart Grid Reference Architecture. [Online]. Available: ec.europa.eu/energy/gas_electricity/smartgrids/doc/xpert_group1_reference_architecture.pdf

[16] GridWise® Interoperability Context- Setting Framework. The GridWise Architecture Council, March 2008. [Online]. Available: www.gridwiseac.org/pdfs/interopframework_v1_1.pdf

[17] Intelligrid. [Online]. Available: www.intelligrid.info/IntelliGrid_Architecture/Overview_Guidelines

[18] Industry Solutions - GridBlocks Architecture. [Online]. Available: www.cisco.com/web/strategy/energy/gridblocks_architecture.html

[19] Cyber Security User's Guide (NISTIR 7628) - Smart Grid Cyber Security Implementation Guidelines. [Online]. Available: http://www.nist.gov/smartgrid/upload/nistir-7628_total.pdf

[20] Overview of the TCIPG Project. [Online]. Available: tcipg.org/sites/default/files/factsheets/FactSheet_TCIPG-Overview.pdf