Renewable Energy Decision Support Systems: The Challenge of Data Integration

Ioana Andreea STĂNESCU1, Veronica ŢEFAN2, Gabriela NEAGU3, Carmen Elena CÎRNU4

1 Advanced Technology Systems, 1, Tineretului Str., Târgovişte, 130029, Romania ioana.stanescu@ats.com.ro
2 Valahia University of Târgovişte, 2, Carol Bd., Târgovişte, 130024, Romania veronica.stefan@ats.com.ro
3 Research Institute for Quality of Life, 13 Calea 13 Septembrie, Bucharest, 050711, Romania gabi.neagu@iccv.ro
4 The National Institute for Research & Development in Informatics, 8-10, Maresal Al. Avrescu, 011455, Bucharest, Romania carmen.cirnu@ici.ro

* corresponding author

Abstract: The practical benefits of a Decision Support System rely heavily on the system’s ability to efficiently aggregate and manage large volumes of data, information, and knowledge from heterogeneous sources. The authors approach this challenge and explore the concept of flexible software architectures that support highly customizable decision environments at end-user level. The paper proposes a mechanism applied in the Renewable Energy filed that enables end-users to visually describe data and knowledge source structures using a set of data types and semantic annotations, and to integrate them into the DSS logic without intervention from system developers.

Keywords: dynamic software architecture, self-service, INDESEN

1. Introduction

Building upon [1], [2] and the “Europe 2020: A strategy for smart, sustainable and inclusive growth”, the Future Internet Enterprise Systems (FInES) Research Roadmap 2025 [3] envisioned the evolution of key technological areas focusing on networking, knowledge, application, computation and storage technologies, stipulating that:

- The existing understanding of networking will be extended to accommodate advanced forms of collaboration, Interoperability Service Utility, and security;
- High performance content networks will be capable to extensively store, link, integrate, and distribute data and knowledge, coming from any possible entity and source;
- Future computing and storage capacities will be based on smart proactive interconnected objects, able to collaborate spontaneously to create more complex computational entities.

The authors consider such foreseen evolvements and related research [4], [5] from their potential to improve the end-user experiences by increased their independence from information system developers and by enhancing the outcomes of decision processes in terms of speed, efficiency, cost, quality and security.

The paper explores opportunities for Decision Support Systems (DSS) customization at end-user level within the context of the INDESEN project, by employing dynamic software architectures to enable accommodation of new data and knowledge sources, as well as the adaptation of the system logic, without support from DSS developers.

This self-service approach has even higher relevance in decision support, where the efficiency of the decision making processes depends extensively on the timely availability, accuracy, and quality of data and knowledge sources.

Currently, when data and knowledge sources used by a DSS become obsolete or their quality decreases, or when new sources become...
available, end-users cannot update the sources employed by the system without support from developers. This creates a major drawback, which in domains such as Renewable Energy (RE), which employ a large variety of equipment and services, is even more significant.

Increased market dynamics require new abilities to aggregate and process data and knowledge sources. For decision makers it is crucial to be able to easily integrate new data and knowledge generated and collected in heterogeneous environments, by a large number of entities.

To address this issue, DSS developers have sought solutions based on data integration and interoperability practices. However, how to design DSSs with a high capability to use data from a large number of sources remains an open issue. Moreover, even if these solutions have aimed to enable access to a considerably large number of data sources, DSS customization and access to new data sources cannot be implemented without assistance from system developers.

If for larger enterprises that have their own IT department or IT experts this is an issue that can be overcome easier, for micro, small and medium enterprises, the inability to adapt a DSS to their specific needs creates major inefficiencies, especially since RE parks are difficult to manage without monitoring and control systems.

This paper addresses these challenges and builds upon the premise that decision makers need a more personal and a direct control of the DSS [6]. Several data integration challenges are presented and issues specific for the RE field are discussed. The authors introduce a mechanism that facilitates data and knowledge integration at end-user level within the INDESEN System.

2. Technology-driven Evolutions in Decision Support

In this section, the authors discuss key technology evolutions that have critically impacted DSS capabilities, with the purpose of identifying major challenges. The negative impact of inflexible DSS architectures is highlighted, and issues identified at end-user level are considered.

2.1 Data fusion

The ability to share data among multiple applications and organizations coupled with the need to integrate data in a flexible and efficient manner remain a critical issue for enterprises operating in highly dynamic markets fuelled by the Internet technologies.

The Web hosts millions of databases that contain data in a plethora of domains and leveraging. This incredible collection of data and knowledge implies addressing issues such as data heterogeneity, data storage, data extraction, data inconsistency, explicit knowledge capture, knowledge transfer, etc.

Building upon the approach proposed by [7], the key challenges related to an effective data integration can be structured as follows:

- **System challenges.** Systems are developed using different technologies, and even if their development is carried out using the same technologies, design and development practices vary not only from one company to another, but also from one developer to another. For example, if a source is a full SQL engine and can accept very complex queries, another source may be a Web form and therefore accepts only a small number of query templates.

- **Logical challenges.** Even if structured data sources are organized according to a schema, data coming from multiple sources can differ significantly in structure. The schema in the case of relation databases specifies a set of tables, and for each table a set of attributes with their associate data types. The schema in the case of other data models is specified by tags, classes, and properties. For example, given the same specifications for a database application, two individuals can design very different schema.

- **Social, administrative and legal challenges.** Beside technical issues, data integration efforts rely on other factors such as the openness of data owners (people, companies, universities, etc.) to share their data. At times, access to data is restricted or requires special handling (e.g. personal data).

Additional issues emerge out of poor data quality. For example, as enterprises grow, the management structures expand as well, causing the dilution of data management
responsibilities as it passes from one subsystem to another.

Management of metadata relies on data definitions, names, data types, data domains, constraints, applications, database tables, reference repositories and dependence rules associated with different tables and databases, users, access rights, etc. [8].

The ability to integrate data and information from multiple sources within a DSS, even if it is challenging, has greatly enhanced decision making processes. However, the integration of data sources into a DSS is carried out during the development phase. Further customizations of the data sources of a DSS can be carried out only by the system developers. Such initiatives imply the allocation of additional budgets and they can prove time consuming.

Decision makers do not benefit of the possibility to integrate new data sources into a DSS, they cannot customize the data sources that they find most relevant for a certain decision process, and they cannot customize the algorithms and the decision flows accordingly. As some data sources may become irrelevant or obsolete for a certain decision process, or they have been identified as inaccurate, DSS end-users are faced with limited to no capabilities to adapt the system to their current needs without support from the system developers. This greatly hinders the decision making processes and reduces the opportunity to automate the process.

2.2 Knowledge fusion

Knowledge is a strategic tool and it is essential for long-term success of the decisions made within an organization. Knowledge Management (KM) implies the development of repositories that support digital capture, storage, retrieval, and distribution of the documented knowledge within an organisation and beyond. Current KM systems focus on centralized sets of repositories, they are expensive to implement, and they require a long-term commitment to maintaining them [9].

According to [10], key technologies associated with KM can be divided into two categories:

- **Integrative technologies**: document management, knowledge databases, data mining, knowledge repositories, expert systems, IT platforms, groupware, enterprise 2.0, etc.

- **Interactive technologies**: email, collaborative authoring, discussion forums, social networking tools, blogs, wikis, information provision, real time interactions, virtual worlds, multiplayer games, crowdsourcing, co-creation, etc.

As Web 2.0 technologies facilitate knowledge sharing [11], they have the potential to enhance decision making processes. However, at a technical level, knowledge integration has become more challenging in the era of interactive technologies, as the knowledge generated through these new types of collaborations is less structured and more difficult to integrate into current DSSs.

The integration of knowledge from heterogeneous sources enable decision makers to access different types of knowledge and to enhance decision processes. Besides the issues regarding the limited customization capabilities of DSSs at end-user level, the emergence of new knowledge-driven environments (e.g. crowdsourcing, social networks) makes it difficult to integrate these unstructured knowledge sources into normalized systems [12].

2.3 Intelligent systems

Since intelligent capabilities such as the ability to learn, reason, discover, recall data, etc. can be embedded in computer software, research and development efforts have focused on creating better DSSs that employ such capabilities.

The intelligence of a knowledge system is depicted through its capability to reason and to acquire knowledge as efficiently as possible and retrieve it for a specific purpose [13].

2.4 Big data and business analytics

For a business to set the right goals, it is best to start from the analysis of the current market it desires to target. Given the richness of today’s market data, business analytics have been proven to be capable of generating detailed and actionable market knowledge [14]. To fully benefit from business analytics capabilities [15] (Figure 1), critical knowledge from analytics need to be properly managed and actionable market insights need to be made available to the decision makers. In this context, efficient data integration emerges again as a key issue.
The development of the computer environment and of the network environment enables an easier collection and storage of large volumes of data. To manage this data, it is necessary to discover frequent patterns and association rules from the collected data that can be used in decision making processes [16].

2.5 Cloud services

The approach proposed in this paper builds upon cloud computing developments that enables the delivery of computing resources as a service. Cloud solutions implement horizontal scalable architectures, dynamic sizing (automatically adding and removing server instances), zero downtime architectures, dynamic failure detection and recovery architecture, etc [17], [18].

Cloud capabilities, especially dynamic sizing, answer to a critical issue that concern the use of the term “ecosystem” in connection to software: absence of intention [19]. Cloud services are key to building auto-adaptable software systems.

Because cloud environments are able to use as many servers as necessary to optimally respond to the cost and the timing constraints of an application [20], they are able to nurture the natural evolution of DSSs towards software ecosystems.

3. Perspectives on the Use of Data Sources in the RE Field

In the last decades, there has been a massive shift from traditional types of energy generation to distributed generation that has put an enormous strain on existing power grids. The complexities of energy generation, storage, and demand, maintenance of power grid efficiencies have set the stage for the development of more complex energy management systems [21].

The opportunity to harvest energy from renewable sources has brought forward new challenges. Even if the benefits of RE have been demonstrated along years – for example, energy security can be improved through local sourcing, it is necessary to take into account the fact that RE sources are diffuse and intermittent. Even if they can be harnessed using rudimentary technologies, they are less easy to predict and control [22].

Following massive investments in RE projects, RE plants have increased in number and complexity [23]. The operation of these plants is supported by a large number of applications and services. To be able to manage the increasing complexity of the RE applications, the rapid changes and the constant risks in the RE field, decision makers need to be able to use advanced systems that integrate large volumes of data and knowledge, and that provide real-time access to these sources.

Solar energy can be harvested with photovoltaic (PV) cells. A typical PV cell produces approximately 0.5 volts (V) and a current that significantly depends on the intensity of the sunlight and the area of the cell. PV systems are used in many applications such as battery charging, water pumping, home power supply, satellite power systems, etc.
To efficiently operate a PV park, decision makers need to take into account a wide range of variables, such as [24]:

- **The hourly and seasonal variations in solar radiance.** The Earth’s rotation produces hourly variations in power intensities. These are also influenced by location.

- **Direct and diffuse sunlight.** The combination of reflection, absorption, refraction, and scattering results in highly dynamic radiation levels at any given location on the Earth’s surface.

- **Air mass.** It indicates the relative distance that light must travel through the atmosphere at a given location.

- **Peak sun hours.** The number of peak sun hours per day at a given location is defined as the equivalent time in hours at peak sun conditions.

To evaluate the performance of PV cells, modules, panels, and arrays, it is necessary to consider the most important performance descriptors: current-voltage characteristics, the maximum power point, the effect of resistive load on the operating point and so on.

High quality solar irradiation estimate is critical in solar energy production. The use of statistics and machine-learning theories can provide a significant improvement in reducing the prediction errors [25]. Measurements of temperatures can be obtained freely, but resulting databases are not standardized.

When large-scale RE sources are integrated into existing energy systems, DSSs must provide support for coordinating fluctuating and intermittent RE production with the rest of the energy system [26]. Given the nature of PV, wind, wave, and tidal power, it does not suffice to regulate only the renewable source itself to obtain electricity balancing. The possibility of achieving a suitable integration depends on the DSS capabilities.

4. **INDESEN: A Knowledge-based, intelligent Decision Support System**

The research reported in this paper documents the decision-making processes associated with the management of RE plants with the purpose of supporting the design of a cloud- and knowledge-based Intelligent Decision Support and Control System for Low Voltage Grids with Distributed Power Generation from RE Sources (INDESEN) where end-users can integrate into decision flows any data and knowledge sources that comply to accepted formats.

INDESEN is designed to integrate the following main modules:

- **The Monitoring Module.** Enables real-time tracking of the parameters that monitor the electric energy quality and the environmental factors. The data will be stored in the INDESEN database and will fundament the development of the knowledge base and the diagnosis of the hardware equipment.

- **The Diagnosis Module.** The diagnosis of the equipment and of the causes that led to dysfunctions is carried out in the Diagnoses Module against reference standard values collected from the equipment specifications, the regulatory requirements issued by competent authorities, standards, national and European legislation, etc.

- **The Prediction Module.** The short-, medium, and long-term prediction is based on internal and external data sources. Data are constantly collected from the equipment and sensors installed in the RE parks and from external weather services. The analysis takes into account the historical data and expert knowledge collected in the system. The capacity to produce electric energy is determined based on factors such as: the periodicity (day/ night; summer/ winter) of the environmental factors (solar radiation, temperature, wind speed), the equipment wear in time, etc.

- **The Decision Module.** The system will generate alternative decision scenarios that the decision-makers can choose from. For each scenario the system will present a synthesis of the rationale behind it build on
historical data, current state of the equipment, weather prediction services, expert knowledge, etc.

- The Control Module. The system will allow real-time control of the equipment, in order to increase its efficiency. Decision-makers can intervene to optimize the energy supply/storage ratio.

INDESEN will be offered as a cloud-based service. The service will be open and it will interconnect several data sources with the actors that take the decisions based on the analyses carried out by the system. Data communication is based on open Web services standards SOAP (Simple Object Access Protocol), REST (Representational State Transfer).

The system is designed as a cloud centric solution that stores and processes data centrally. This approach increases the adoption by lowering upfront costs and overall Total Cost of Ownership. Another contributing factor in choosing a cloud based solution is the large geographical distribution of RE power plants which can only be efficiently managed through a centralized solution.

For data collection purposes, the system defined a set of API functions that can be used by third-party developers, as well as a set of standard software agents that can be deployed on site and feed data from proprietary hardware to the cloud.

End users are able to integrate their own data sources in any of the above mentioned modules of INDESEN. The customizable decision layer maintains the consistency across the system, as end users can define correlations between the new data sources and the decision flows supported by the system.

As INDESEN is designed as a cloud-based system, this enables the solution to scale according to the demands. End users will not encounter hardware or software capabilities issues when connected to a large number of new data sources.

5. Data Integration Scenario to Support Decision Making in the RE Field

Decision processes in the RE field are subject to challenge as they are based on masses of details concerning various different issues that require analysis and management: status of the equipment, sensor data, weather prediction services, and so on. For example, energy prediction should be as accurate as possible to minimize additional costs and risks.

RE decision makers use different tools to carry out activities such as monitoring and control, diagnosis, prediction, simulations and decision-making. They also rely on a wide range of data sources to fundament their decisions. These data sources are difficult to estimate when the decision support tool is developed. Even if initial, generic data sources exist, for better decisions DSS end users need to be able to customize the data sources they use based on priority, accuracy and other user-defined parameters. They should also be able to integrate a larger number of data sources if needed, without assistance from system developers.

The availability of a large number of data sources, especially concerning weather prediction and next-day pricing, coupled with the need to remove obsolete sources and add new data and knowledge sources that can improve decision processes in the RE field, increase the complexity of the decision making environments and highlight the necessity to advance new solutions that give more independence from developers and enable end-users customize the DSS to better answer their specific needs.

The research carried out in this paper builds upon the concept of dynamic software architectures that empower end-users to have a more sustainable control over the DSS architecture. In the case of a dynamic system architecture, end-users are able not only to integrate their own data and knowledge sources, but also to customize the DSS algorithms and decision flows to integrate them.

INDESEN has been designed to enable end-users connect new data and knowledge sources and integrate them into the system logic without support from developers. End-users are able to connect new monitoring equipment or sensors they use into the system. For direct connected devices that support direct IP addressing (Internet of Things), the end-users provide the device endpoint address and select the protocol that will be used to communicate with the device (SOAP, REST or flat file transfer using HTTP, FTP or S/FTP). The
system can then retrieve raw data from these devices and feed it into the semantical mapping engine. For devices that do not support direct addressing, the system allows end-users to push data into the cloud using one of the above-mentioned transfer protocols.

One of the key barriers that need to be overcome is the fact that data sources differ in the format they provide data. Even if these differences are minor, they can have a significant impact on the data integration process and on the quality of the outcomes.

Not all the databases have the same structure. Some may provide fewer information, or even if they provide similar information, the data can be structured differently. For example, weather databases manage data on temperatures, latitude, longitude, date, etc. These fields may not be in the same order or they may not use the same format.

- Latitude and longitude can be decimal or can be expressed using minutes and grades.
- Numerical values can use a comma or a dot.
- Data formats can be day/ month/ year or month/ day/ year.

There are many different units that are employed in energy work. Even if systems of measurement as the International System of Units (SI - Le Système International d'Unités) are used, a greater consistency in the choice of names and symbols of the different prefixes adopted by the SI is needed [27].

Even if pascals are commonly used to express pressures, atmospheres are also used. Electron-volts are more convenient than joules. Energy in barrels of oil or kWh may convey better the idea of costs.

INDESEN has been designed to enable each user to connect his own data sources into the system (equipment data, sensor measurements, weather data, etc.) without support from developers, using a wide range of formats, and then transform and give meaning to this data employing the INDESEN Integration Layer, in order to process it and support enhanced decision processes across the INDESEN modules (Figure 2).

Before this raw data can be used in the decision process, end-users must provide a format for interpreting it. For this, INDESEN provides a tool that allows users to visually map data source fields with system data points, and at the same time define conversion rules for value formats, units and scales, if required. At the end of this stage, data is clean and coherent and it can be grouped into fuzzy sets and processed by INDESEN.

By using this approach, end-users are provided with the necessary integration glue that is required to cover the multitude of RE devices and services available, without depending on
developers or official Original Equipment Manufacturer support for this operation.

The next step into the direction of a self-service DSS is to allow users to change the rule base so that it can incorporate new data points and also fine tune existing algorithms for each individual user’s need. For this, INDESEN has been designed to allow end-users direct control over the entire decision logic. INDESEN rules and algorithms can be individually managed for each user, leading to an enhanced user experience.

6. Discussions and Conclusions

Today’s enterprises cannot be effective without high quality data that can support more informed and more efficient decisions. Data, information and knowledge quality depends on a variety of factors starting from creation, storage, usage, ownership, the existence of metrics for measuring and assessing quality, error detection and correction.

Even if advancements in information technology bring new opportunities to improve DSS capabilities, the development of new DSS features cannot be carried out without the direct involvement of the system developers. As the redesign process requires time and implies additional costs, the adoption rate of emerging technologies remains low.

The authors approach the challenge of enabling intelligent fusions of massive amounts of heterogeneous data, information and knowledge from a wide range of distributed sources, in order to support problem solving and decision making.

The paper promotes the concept of dynamic software architectures that enables the transition from inaccessible, inaccurate, out-of-date, unmanageable, redundant information; inconsistency with other sources; incomplete data; and incomprehensible data to a consistent integration of data, information, and knowledge.

The authors explore the opportunity to build systems geared for flexible sharing and integration of data across multiple autonomous data providers. A data integration scenario is presented for the RE field. Access to timely information is crucial in order to promote efficiency and quality across RE decision making processes. DSS should integrate the critical mass of data sources to support performance.

Based on the concept of dynamic software architectures, one of the goals of the INDESEN project is to develop a mechanism that enables end-users to visually describe the data structures of the files using a set of data types and semantic annotations that allows the DSS to process and validate the data.

INDESEN is capable of integrating data from customer systems using a multitude of data formats such as XML, JSON and delimited flat files, as well query relational databases and parse non-conventional data sources such as HTML.

For automated time data upload, the system also provides connectors that can be installed and configured by users to collect information and push it at regular intervals.

This approach builds upon the next generation of DSSs that shift from inflexible to flexible software architectures and empower end-user to personalize their decision environments.

Further work will focus on extending and testing current functionalities and limitations of the data integration mechanism with end-users, in order to validate the proposed approach.

Acknowledgements

The work presented herein is partially funded by Unitatea Executiva pentru Finantarea Invatamantului Superior, a Cercetării, Dezvoltării și Inovării (UEFISCDI) through the Romanian Partnership Program, Contract nr. 42/2012 Intelligent Decision Support and Control System for Low Voltage Grids with Distributed Power Generation from Renewable Energy Sources (INDESEN).

REFERENCES


23. STANESCU, I. A., A. STEFAN, F. G. FILIP, C. KITTL, T. LIM, Interoperability Scenarios in Serious Games Ecosystems: The Impact on FInES, 7th IFAC Conference on Manufacturing Modelling, Management, and Control, Volume # 7,

