Bringing Analytics into Practice: Evidence from the Power Sector

Completed Research Paper

Gunther Gust
University of Freiburg
gunther.gust@is.uni-freiburg.de

Christoph M. Flath
University of Würzburg
christoph.flath@uni-wuerzburg.de

Tobias Brandt
University of Freiburg
tobias.brandt@is.uni-freiburg.de

Philipp Ströhle
Karlsruhe Institute of Technology
philipp.stroehle@kit.edu

Dirk Neumann
University of Freiburg
dirk.neumann@is.uni-freiburg.de

Abstract

Across industries, the increasing availability of sensor data has created business opportunities for the application of analytical information systems. We shed light on the analytics implementation in practice in the context of a case study in the power sector. Following a design science approach, we present a case study on the implementation of a decision support system (DSS) for grid planning at a large utility. Given the very large number of grids, process automation through analytics promises significant efficiency gains for labor-intensive planning tasks. We demonstrate how the DSS leads to process improvements regarding speed, accuracy, and flexibility. Apart from the benefits for the company, this work contributes to IS practice by deriving general lessons for IS executives facing analytics challenges.

Keywords: Decision Support Systems, Design Science, Case Study, Analytics in Practice, Power Grid Planning

Introduction

Analytics systems are a new, evolutionary step of business intelligence applications that are expected to transform the corporate landscape (Davenport and Harris 2007). New ways to generate business value through analytics have been discussed in both academic and practitioner literature (Chen et al. 2012; Sharma et al. 2014). Chen et al. (2012) structure the emerging field of Big Data and business analytics and outline promising applications. Analytics solutions are geared to look beyond isolated processes and towards shaping decision-making, either with respect to operational decisions or by improving nuanced aspects of strategic decisions (Davenport and Harris 2007).

A case in point is the power sector: Faced by radical technological changes (renewable generation, electric vehicles, smart grids) and new regulatory requirements, utility companies are confronted with many strategic decisions while navigating the uncertain future of energy markets. Consequently, a host of long-established industry procedures need to be adapted and rethought to meet these challenges and to respond to recent developments. To cope with planning uncertainty and to foster better business decisions, data-driven business intelligence approaches are needed (Lim et al. 2013). Such decision-supporting systems could provide grid operators with valuable information, thus enabling better investment and operating decisions. Dawson et al. (2013) note that smart grid investments so far have focused on technologies for communication and data-gathering. They argue that leveraging these new grid capabilities will require back-end analytics systems which allow utilities to extract information from raw front-end data and subsequently optimize processes as well as product and service offerings. Industry estimates on investment savings through intelligent grid planning and through the use of innovative technologies range between 20 and 60% (dena 2012). These potentials underline the importance of information systems in the energy sector, which lends validity to recent manifestos for establishing Energy Informatics (EI) and Green IS as core fields in information systems research (Goebel et al. 2014; Melville 2010; vom Brocke et al. 2013; Watson et al. 2010).

Industry Background

The power grid comprises different layers ranging from the transmission grid, with its extra-high-voltage lines, to low- and high-voltage distribution grids as illustrated in Figure 1. Generation is not limited to higher voltage levels. In particular, renewable energy sources, like photovoltaics or combined heat and power plants, are increasingly integrated into low voltage grids. Therefore, distribution grids are at the forefront of the energy transition because decentralized generation disproportionately affects the distribution end of electrical power systems.

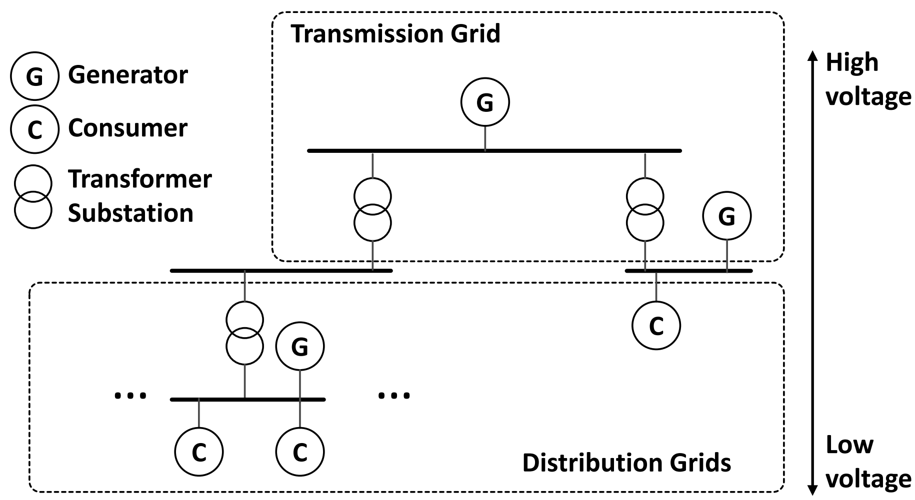


Figure 1. Different layers in a typical power grid

Correspondingly, distribution grids will require large investments in order to provide necessary reinforcements. Investment decisions are virtually irreversible since asset life cycles in the power sector typically span 30–50 years (IEC 2015). Given this long planning horizon, these decisions are greatly affected by uncertainty. Besides the uncertain course of future installed renewable generation capacities and associated generation volatility, the widespread adoption of new electrical loads, such as heat pumps or electric vehicles, introduces additional uncertainty regarding the appropriate level of grid capacity investments. To manage this uncertainty, traditional planning and operation procedures call for significant expansion of distribution grid capacities to establish sufficient operational margins (dena 2012). However, already today, capital requirements for distribution grids are enormous: For the US power system alone, the American Society of Civil Engineers states that "the investment gap for distribution infrastructure is estimated to be $57 billion by 2020" (ASCE 2011, p. 36). Given such capital requirements, even small efficiency improvements will lead to substantial savings.

Analytics for Grid Operators

Against the backdrop of a changing energy landscape, we worked together with one of the largest utilities in Switzerland to support internal decision-making and and boost internal analytics capabilities in line with the corporate mission statement:

> *"We seek to make efficient use of resources. We are committed to developing innovative technologies to support a sustainable, secure future for the supply infrastructure."*

Besides operating power plants and transmission grids, the company is also a large distribution system operator (DSO), owning and operating thousands of low-voltage distribution grids serving more than one million customers. Faced with the challenges of large scale distributed generation and novel loads, there is great potential to improve planning processes. The project's goal is to equip asset management with state-of-the-art analytics and decision support tools for strategic distribution grid planning. The major challenge of distribution grid planning is the large number of low-voltage grids, which makes a comprehensive manual assessment impossible. Decision support tools approach this challenge in two ways. The first approach is to reduce the large number of grids to a few representative grids, which can then be analyzed manually by the grid planners (Dickert et al. 2013; Kerber 2011; Strunz et al. 2009). Second, an increasing share of DSS employ fully-automated procedures (Gust et al. 2016; Hollingworth et al. 2013; Schrader 2014) that are scalable to large sets of grids. The DSS presented belongs to the former class of systems, which have the advantage of increasing users' expertise due to their stronger involvement in the planning process. The DSS is to be primarily used by the company's grid asset management department, which is responsible for long-term strategic planning. The department has decided to move ahead with this flagship project aimed at establishing best-practice grid planning services which may eventually be offered to third-party DSOs.

> *"Now we are having the opportunity to take part in the shaping of the energy transition. We need to think beyond the status quo and develop innovative ideas."* (Senior manager)

The DSS was developed following the design science paradigm (Hevner and Chatterjee 2010). Thereby, we address a gap identified by Gordon et al. (2013) in case-based research in the area of analytics and decision support. We address the following research questions:

- **Q1:**What are the requirements the system needs to satisfy?
- **Q2:**How does the DSS need to be designed?
- **Q3:**How is the DSS field tested?
- **Q4:**Are the requirements satisfied and what are the estimated benefits of the DSS?
- **Q5:**What new knowledge (meta-artifacts, scientific theory, expertise) relevant to research and practice is added to the knowledge base?

The structure of this paper follows the design science publication scheme as defined by Gregor and Hevner (2013) and sequentially addresses the design questions. We conclude with a summary of the project's main contributions and its limitations.

System Requirements

Grid planning decisions require accurate information on electrical loads, generators, and topology for a given distribution grid. Given the large number of individual grids, which amount to thousands, current planning and budgeting processes rely on generic demand projections based on extrapolated historical data and expansion requirements determined through manual assessment of arbitrarily selected individual grids. Subsequently, this information is transferred to other grids deemed similar based on the underlying settlement type (e.g. village or town). Naturally, this approach facilitates neither precise (plans are based on general data) nor fast (significant number of manual tasks) planning processes. Recent approaches to determining representative grids based on a small sample of manually collected network data improve information precision, and yet the overall process remains very time-consuming as it still requires significant involvement from grid planners.

The limited capability of current processes motivates the development and introduction of a decision support system to assist distribution grid planners. For a rigorous definition of systems requirements, we rely on the design framework by Sprague (1980). Besides the organizational and technological context, DSS design is primarily governed by requirements concerning performance objectives and the necessary capabilities to achieve these. Here we discuss the crucial performance objectives as requested by the future system users, i.e. grid planners. These criteria are ultimately determined by the decision task at hand as well as the organizational context and the decision-makers involved (Sprague 1980). In our case, the DSS is used to improve strategic grid planning processes and we seek to make improvements in the following areas:

1. **Faster strategic planning processes.** Current grid planning processes, such as reinforcement planning, are heavily reliant on manual collection of grid data and time-intensive calculations by grid planning engineers. Realizing that the aforementioned future challenges apply to thousands of individual distribution grids, improvements in processing speed are of great importance.
2. **Improved planning accuracy.** Today, grid planning suffers from a lack of detailed information on individual grids and the associated investment requirements. Hence, current practice is often based on subjective expert assessments and historical planning heuristics. This may result in conservative worst-case planning, which entails high investment outlays and yields limited generalizability. Digitization of grid assets, GIS (geographic information system) tools, and the availability of real-time data on grid operating states allow for more accurate representations of the grid population. Using new tools to leverage these inputs should greatly increase planning accuracy and, in turn, increase cost-efficiency.
3. **Increased planning flexibility.** Historically, distribution grid planning was a well-defined and stable activity for which planners would routinely align given standards with the current development of connected loads. The emergence of distributed renewable generation and large-scale flexible loads, such as electric vehicles or storage devices, pose a disruptive change to the status quo. Therefore, the DSS should be a versatile tool suited to a variety of different analyses and scenario contexts.
4. **Seamless process integration.** Sprague (1980) suggests that decision support systems should serve decision-makers from different organizational units and hierarchy levels. We adopt this requirement by striving to develop an accessible and user-friendly system. To this end, the system should offer meaningful input and output functionality, be accessible by all relevant staff, be integrated into current business processes, and should not rely on proprietary planning tools.

DSS Implementation

The DSS was developed iteratively through feedback loops with management and grid planners (Hevner and Chatterjee 2010). In the design cycles, first, different software platforms were tested for congruence with the requirements and, second, the DSS's features were refined. The project sponsor's main concern was alignment with existing processes. Therefore, he argued for a spreadsheet solution based on MS-Excel.

> *"If necessary, we can pass on some functionality for low initial burdens for grid planners."*
> (Project sponsor)

However, the analytic functionalities of MS-Excel proved insufficient and implementation effort very high. Thereafter, the idea of basing the DSS on *R* together with a power-flow calculation software used in the department was evaluated, the advantages being usage experience and highly accurate load-flow calculations. However, the system was not scalable to simultaneously evaluate several grids. Additionally, the power-flow module was offered by a small, external software provider so that implementation knowledge and adaptability would have been limited. Finally, the decision was taken to implement the system entirely in *R* since it displayed the necessary analytic capabilities and flexibility with regards to future adaptations and was also the preferred tool at the company's newly created analytics department.

Figure 2 illustrates the final structure and key parameters of the DSS. The system contains four modules: The preprocessing modules for grid (1a) and service area data (1b) combine raw data from separate repositories to provide aggregated characteristics of both grid and service area. The analytics module (2) turns preprocessed characteristics into information on various strategic use cases ranging from consumer loads to service reliability. The reporting interface (3) consolidates and summarizes this information for the grid

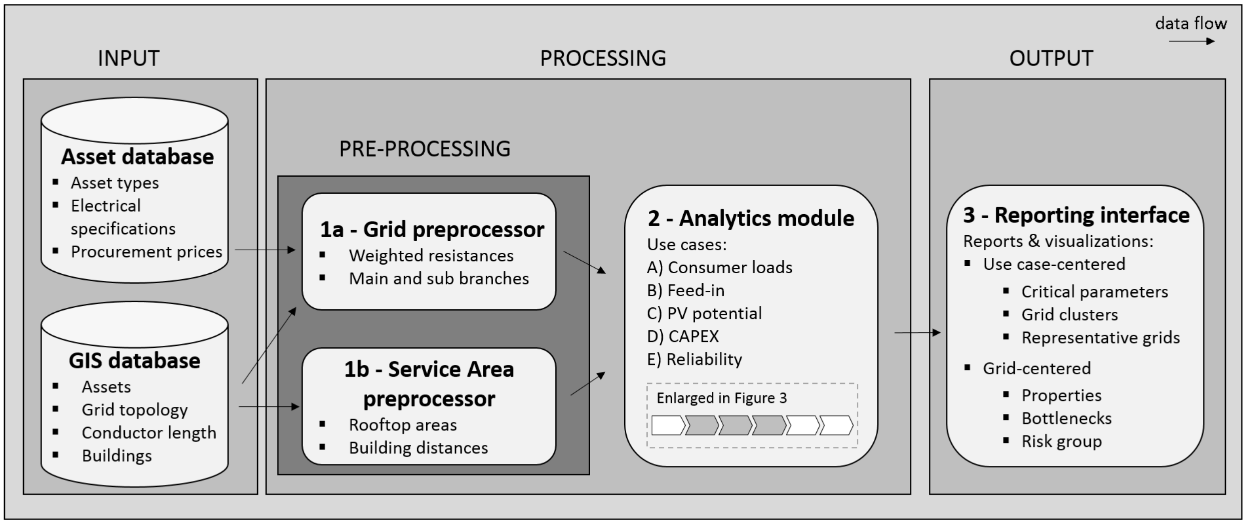


Figure 2. Structure of the decision support system.

planners.

Data Sources

Information on grid characteristics is stored in two places, the geographic information system (GIS) and the asset database. The GIS stores information on installed components in all distribution grids within the operator's service area. This includes technical information on power lines, transformers, and household connections, as well as the corresponding geo-location of these components. Furthermore, information about the service area (e.g. infrastructure such as buildings and roads) is provided, as it may be relevant for different use cases. The quality and validity of GIS data is ensured by the *grid information* department, for example by incorporating updates on the state of the grid generated by the engineering department.

The asset database, on the other hand, is part of an asset management system run by the company's strategic grid planning department. It stores detailed information on asset types and their electrical specifications and serves as a master data store. By separating technical master data from installation data and geo-information, erroneous and redundant data sets are avoided. Beyond technical characteristics, the respective standardized procurement prices are stored in the asset database. These are average prices, including transportation and installation, which are used for grid investment planning.

Preprocessing Modules

The preprocessing modules comprise the grid (1a) and service area (1b) preprocessor. Both modules prepare and transform raw data from the GIS database for use in higher level analyses in the analytics module. Additionally, the grid preprocessor draws on information from the asset database.

The **grid preprocessor (1a)** combines elementary information on the electrical properties of the assets with corresponding topological information (i.e. their location in the grid) and thereby determines the parameters of the overall grid. Topological grid data is available in an edge list (left panel in Figure 3), each row corresponding to a grid asset and containing its technical parameters along with the IDs of neighboring assets. To simplify data operations and to facilitate the extraction of structural information, the list is transformed into a connected graph featuring undirected edges (center panel in Figure 3). The graph edges represent power lines, while the nodes represent connective assets such as transformers, connectors, or generators.

The module's central task lies in extracting network structure in a standardized way. The grid preprocessor first performs a search to locate the transformer that connects the individual grid to the superordinate grid

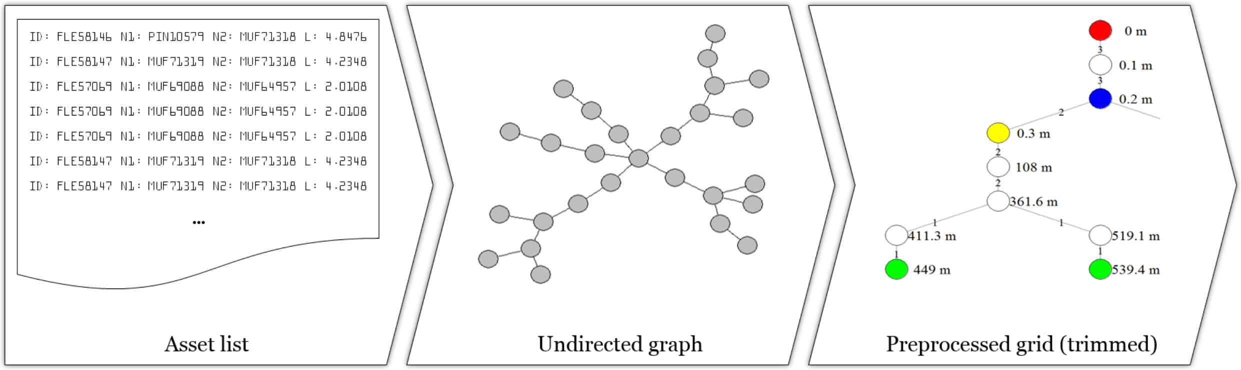


Figure 3. Grid preprocessing

level (top red node in the right panel of Figure 3). Defining the transformer as the root node, the grids can be structured top-down by adding hierarchy levels according to the count of line segments separating each node from the transformer until the household connections (green nodes) are reached. The final graph then resembles a tree. By applying shortest-paths-algorithms, different measures such as distance and electrical resistance between all nodes and the transformer, as well as the number of household connections served by each line segment, can be computed. With this information, the module is able to weigh resistances and accurately classify grid branches into main branches and sub-branches. The preprocessed grid facilitates the evaluation of grid properties by means of power flow calculations in the analytics module.

The **service area preprocessor (1b)** only requires input from the GIS. It structures and transforms data not directly associated with distribution grids, but tied to the surrounding service areas, such as photovoltaics (PV) data. For the analysis of the PV feed-in potential of a specific distribution grid, the service area preprocessor first matches building surface areas to the closest available household connection. Then it derives the rooftop area suitable for the installation of PV panels, taking into account rooftop orientation and inclination. The analytics module subsequently determines the corresponding maximum PV power flows.

Analytics Module

The analytics module serves as the core component of the analytics chain, combining both infrastructure-related and use-case-specific information with the goal of delivering accurate estimates of distribution grids' operational characteristics. The sequence of analyses is visualized in Figure 4.

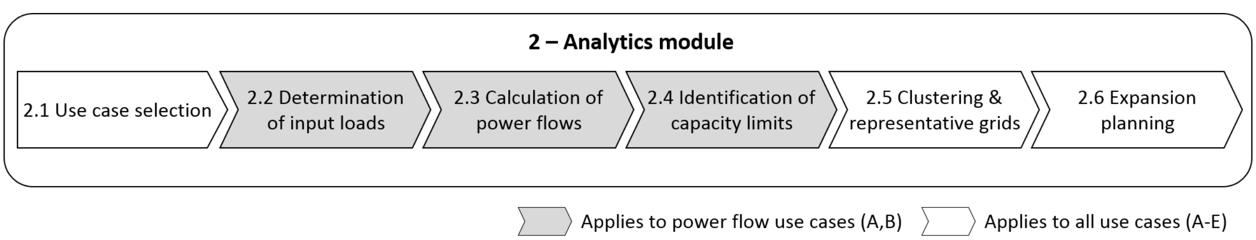


Figure 4. Analytics module process view

Initially, a specific use case is selected from a repository of common analyses performed in strategic grid planning. These standard use cases can be combined to tackle more complex analysis requirements, as we will demonstrate in the next section. The basic use cases cover the following analysis tasks:

- **Capacity limits for consumer loads (A).** Demand for electricity is expected to increase due to growing populations and newly electrified services (e.g. electric vehicles). This analysis produces insights into when grid capacity may be exhausted, potentially requiring intervention by the grid operator.

- **Capacity limits for feed-in (B).** Besides increasing power demand, supply from renewable sources, such as PV, is expected to induce new load peaks and, thus, may occasionally reverse power flows. This use case yields insight into the amount of renewable generation that can safely be integrated into current distribution grids without causing overloads.
- **PV potential (C).** Since generation from PV has been the dominating feed-in technology at the level of power distribution networks, this specific use case has been added. It allows one to determine the PV rooftop potential in the given service areas and to derive corresponding generation levels.
- **CAPEX (D).** Beyond power flows, the analytics module is able to provide an accurate representation of the capital invested in technical grid components such as cables and transformers. Accordingly, accurate estimates of additional grid investments required due to shifts in power demand and supply can be provided.
- **Reliability (E).** System reliability in distribution grids is to a large extent driven by the presence of overhead line segments. Compared to underground cables, these lines are subject to more frequent service interruptions due to adverse weather.Correspondingly, the likelihood of service interruptions can be characterized by means of the share of overhead line segments.

For use cases A and B, which require a power flow analysis, the submodules 2.2 - 2.4 are executed (Figure 4). Initially, the level of aggregate demand (use case A) or supply (B) for which the grids are to be evaluated is specified by the planner (step 2.2). This input may originate from a scenario analysis, as we will outline in the following section. Then the module calculates power flows (2.3) and subsequently identifies the capacity limits for the grids (2.4). The overall capacity limit is determined as the minimum of three specific limits (Table 1). Large electrical currents can cause overloads at the transformer (1), which connects the grid to the superior voltage level, and at individual conductor lines (2). Additionally, voltage distortions (drops or spikes) must not exceed tolerated boundaries (3) along any paths of the grid.

Table 1. Grid capacity limits and handling in expansion planning

Type	Origin	Expansion planning
(1) Transformer overload	Current	DSS
(2) Line overload	Current	DSS (partially)
(3) Voltage boundaries	Voltage	Manually

Expansion planning aims at identifying cost-efficient removal options for grid bottlenecks. To speed up the strategic planning process, the ultimate goal is to implement fully automatic expansion planning. There is significant research dedicated to automatic expansion planning of distribution networks (dena 2012; Hollingworth et al. 2013; Schlömer et al. 2014). Still, current methods are insufficient for application in practice. A key issue is cost determination for line replacements. The majority of Swiss and European distribution grids are underground and line replacement costs are largely driven by excavation work. Costs depend on ground composition — there's an up to forty-fold difference between loose soil and main roads where additional traffic deviation measures are needed. Including these factors in automated expansion planning requires additional GIS information and will be targeted in future updates of the DSS.

The current version of the DSS provides grid planners with assistance in expansion planning, reducing remaining manual work significantly (Table 1). First, the DSS detects *overloaded transformers (1)* and calculates costs of a replacement with increased capacity. This is explained in detail in the next section (Figure 8a). Second, the DSS detects all *overloaded lines (2)* and calculates the length of necessary reinforcements (Figure 8b). The reporting module highlights the bottlenecks for fast identification during expansion planning. Finally, the analytics module clusters all grids and identifies the most representative grids for each cluster. Since the effort associated with manual expansion planning on a large set of grids is prohibitive, a representative grid is used as a substitute for an entire cluster. The grouping can either be performed according to settlement patterns or on *k*-medoids clustering (Kaufman and Rousseeuw 2009). While the former method is the current industry standard due to its simplicity, the latter leads to more homogeneous groups and more accurate investment calculations. The representative grids of the different clusters are then handed to the reporting module along with information on cluster size and similarity.

Reporting Interface

The reporting interface provides grid planners with summarized information in graphical and tabular form. It aims at providing a quick overview of future (hypothetical) grid states according to different scenarios. Reports provide a use-case-centered, macro perspective on the impact of different scenarios on the entire population of grids. This includes distributions of key parameters of the use case, e.g. the display of capacity limits. Moreover, it provides information on the clustering and selection process for the representative grids.

Second, a grid-centered, micro perspective is focused on the impact of the scenarios on the individual grid. It enables grid planners to quickly assess the situation and speeds up expansion planning. Figure 5 shows a graphical micro-level report. At a single glance the report conveys the essential information on a particular grid. Mortenson et al. (2015) note that meaningful visualizations facilitate effective communication, increase the impact of analytical findings, and can even serve as an analysis tool in their own right.

The column on the left sketches the physical properties such as length and count of the main branches and the number of house connections served. The second frame displays the grid topology. Grid elements of special interest, such as branches suffering overloads, can be highlighted based on personal preferences. The third frame displays an aerial image of the grid. During expansion planning, grid planners use aerial images to estimate the cost of line replacements, which depend on ground composition. The bottom frame compares the current grid (orange dot) to its cluster (box plot) in terms of different characteristics. Depending on the characteristic, the grid planner can quickly decide which type of expansion measure is most promising.

An additional risk map serves a similar purpose (Figure 6). It summarizes the properties of each individual grid (rows) with regards to the previously introduced use cases A - E (columns). The cells indicate the grids' cluster membership using five clusters. This will be illustrated in detail in the PV integration analysis in the next section. The risk map assists grid planners in detecting grids with similar behaviour with regards to multiple use cases. Assessing similar grids, planners can evaluate the scope of a particular reinforcement strategy — whether it applies to several grids or is just to some special cases. This is needed for standardization of grid expansion measures, as well as the evaluation of new technology options for grid reinforcement. Furthermore, the risk map also supports grid planners in prioritizing investment decisions. If, for instance, multiple grids needed investment due to expected PV overloads, the ones in a worse state regarding reliability could be considered first.

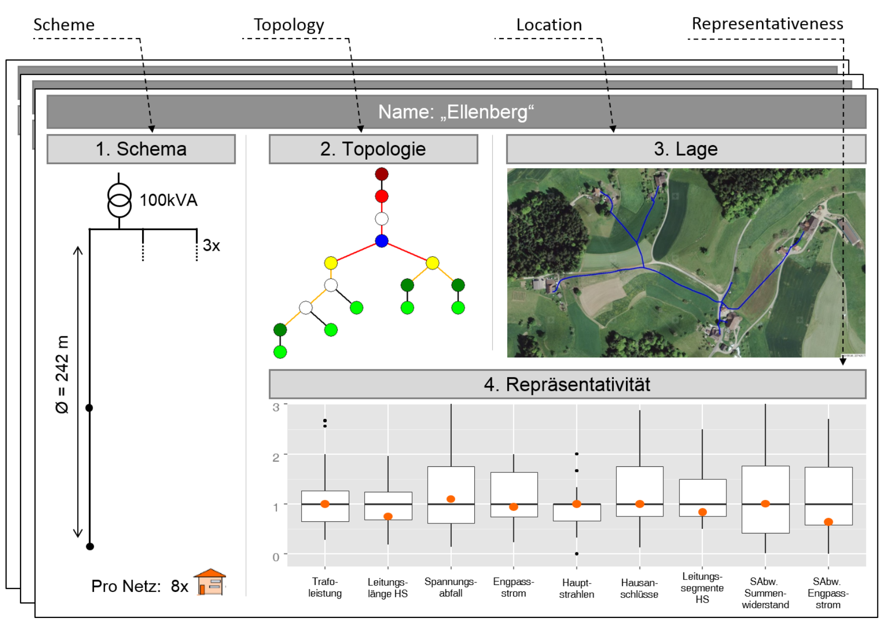


Figure 5. Screenshot of a grid report

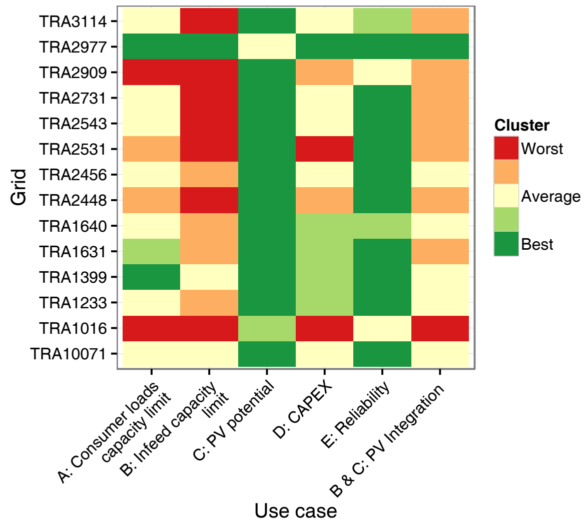


Figure 6. Risk assessment framework (excerpt) for individual grids and various use cases

Process Demonstration: Photovoltaics Integration

The strategic grid planning department employs the DSS to analyze the ability of the distribution grid to integrate solar power from distributed PV generators. We demonstrate the functionality of the DSS throughout the course of this analysis process. For confidentiality reasons, we only report the results for a non-representative sample of 143 distribution grids, focusing on the description of the process steps (2.1–2.6) from Figure 4.

Use Case Selection and Determination of Photovoltaics Feed-in Levels

The goal of the analysis is to determine the grid reinforcements necessary to absorb the expected future amount of PV feed-in. This analysis combines the basic use cases B and C of the analysis module: Feed-in is defined relative to the available PV potential in each service area (use case C: *PV potential*) and the capacity limits are calculated by means of use case B: *feed-in capacity* (Figure 2). Solar power generation may substantially contribute to overall power generation in the future. The Swiss Federal Energy Agency (SFEA) defined four scenarios pertaining to the future of the Swiss energy generation mix (BFE 2013). According to these scenarios, 2–8% of gross PV rooftop potential will be tapped into by 2020, growing to 15–45% by 2035.

We use this data to determine growth scenarios for PV feed-in within the service area. For ease of exposition, we use average scenarios with PV levels of 4% and 28% in 2020 and 2035, respectively, in addition to the upper and lower bounds. In a first step, the analytics module converts the predicted PV levels into input loads, which are the basis for power flow calculations.

Power Flow Calculation and Bottleneck Detection

The DSS subsequently computes power flows in the distribution grids that result from the feed-in and compares them to the capacity limits for PV generation. Figure 7 illustrates grid overloads for the low, average and high PV growth scenario. The difference between computed PV feed-in (use case C) and the capacity limits (use case B) of the grids, the overload, is displayed on the vertical axis. Horizontally, grids are sorted in increasing order of the computed overloads. Grids below the 0-kW line can fully absorb the PV feed-in and do not require any expansion.

Evidently, most grids have sufficient capacity to absorb PV generation regardless of the scenario in 2020. By 2035, the number of grids where PV feed-in exceeds capacity limits is significantly higher. In the low

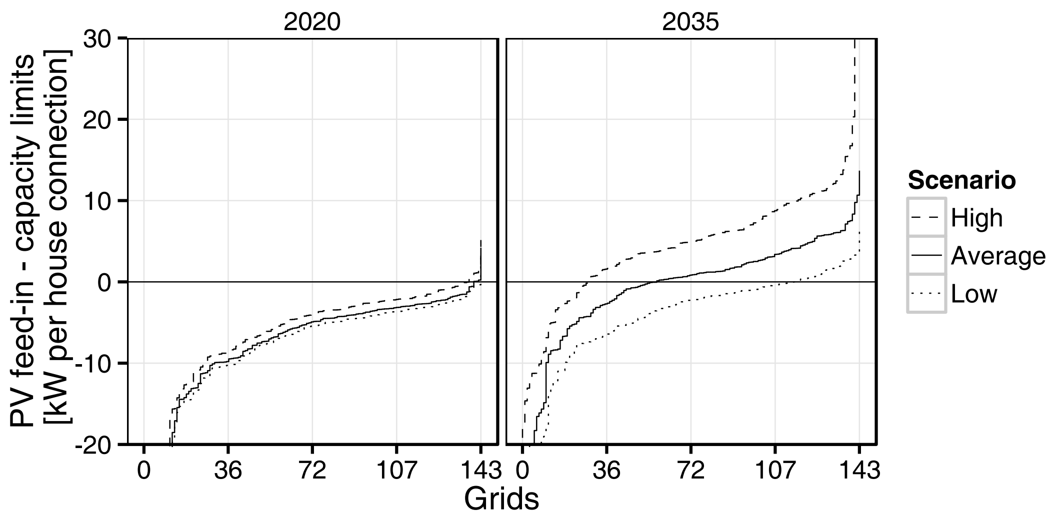
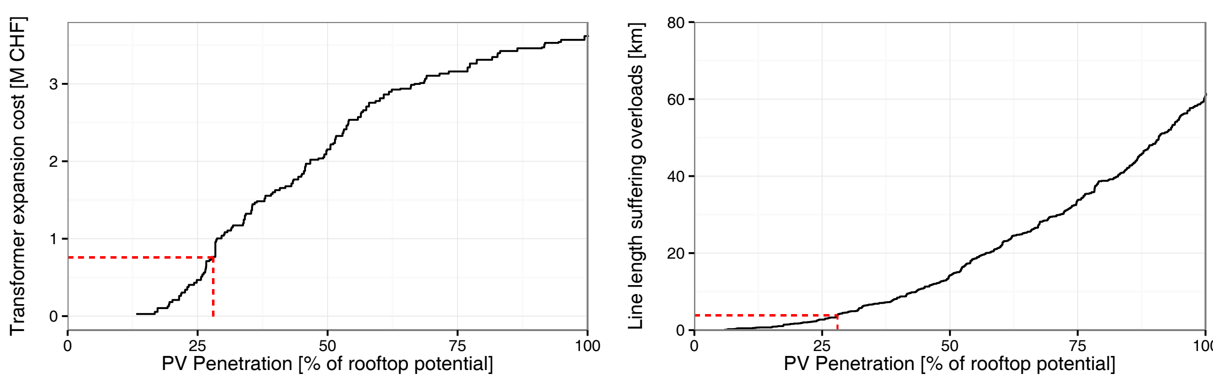


Figure 7. PV overloads for the low, average and high scenario

feed-in scenario, 115 grids (80% of the sample) remain capable of fully absorbing expected generation from PV sources. This number falls to only 60 grids (42%) in the average scenario. Finally, in the scenario with the highest PV penetration level, only 28 grids (20%) will not suffer from bottlenecks.

Automated Expansion Analysis and Grid Clustering

After the bottleneck detection, the DSS quantifies necessary grid reinforcement investments. Figure 8 displays the results of this automated expansion analysis for transformer and power line capacity limits. In order to avoid repetitive replacements, we implemented the following basic strategy for *transformer expansion*: Whenever PV feed-in exceeds the capacity limit of a transformer, it is replaced by the smallest (least expensive) transformer supporting the maximum feed-in level of 100% PV penetration. We are currently working on implementing strategies that identify cost-optimal transformer replacement by incorporating timing and non-trivial expectations about the future. For the *conductor lines* — as mentioned earlier — expansion cost cannot be accurately determined in an automated fashion. Therefore, only the length of overloaded line segments is reported. The necessary expansion measures for the average scenario until 2035 are highlighted in the figures.



(a) Transformer expansion cost

(b) Line length suffering overloads

Figure 8. Automated expansion cost analysis for varying PV penetration levels

The analysis suggests that expansion measures become necessary if PV penetration levels exceed 10%. In the average scenario (28% penetration) transformer expansion costs of about CHF 0.8M are required for the sample of 143 grids. Additionally, about 5 km of lines need to be reinforced, corresponding to 0.05% of total sample line length. The decision, which power line segments to reinforce, as well as the determination of associated expansion costs, need to be made manually by grid planners. To support the planners in this task, the system clusters similar grids based on their experienced overloads and selects representative grids (Figure 9). To derive robust expansion estimates, planning is based on the average scenario. For each cluster the median grid (horizontal line in the box plot) is chosen as the representative grid. In contrast to treating each grid individually, such grouping reduces complexity and manual labor, while retaining high accuracy with respect to expansion planning decisions. The number of clusters can be specified by the grid planner. Generally, decision accuracy increases with the number of clusters formed. The optimal count depends on the maximum dissimilarity tolerated in the clusters and hence will typically be use-case-specific. While we do not evaluate the cluster number choice, the decision needs to strike a balance between the tolerated dissimilarity and the additional cost of manually planning additional representative grids.

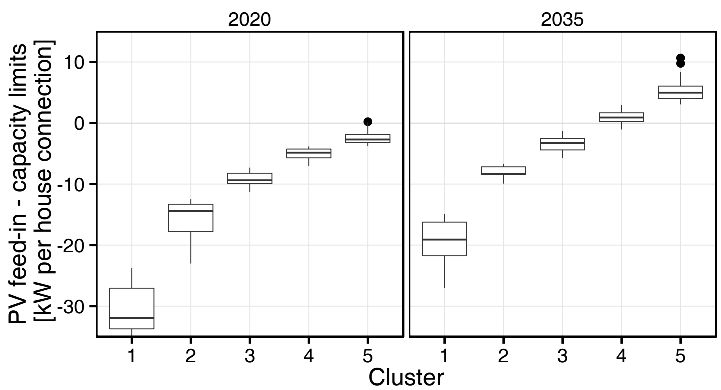


Figure 9. Grid clustering based on overloads for the average scenario of PV feed-in

Expansion planning

As described before, in 2020, all representative grids can cope with the estimated power flows and no expansion planning is necessary (Figure 9). In contrast, for 2035, two representative grids (representing clusters 4 and 5) experience overload and consequently need to be considered for expansion planning.

Table 2. Stylized grid expansion cost estimation using representative grids

	Scenario I			**Scenario II**		
	Cluster X	Cluster Y		Cluster X	Cluster Y	
Representative grid capacity sufficient?	✓	✗		✗	✗	
Required investment [CHF]	0	50k		40k	100k	
Number of grids in cluster	50	20		50	20	
Extrapolated investment [CHF]	0	1.0 M		2.0M	2.0M	
Estimated total investment [CHF]			**1.0M**			**4.0M**

Table 2 exemplifies the expansion cost estimation using a stylized example. For each scenario, clusters are formed and representative grids are selected. If the capacity of the representative grids is insufficient, expansion planning is performed and the investments required for reinforcement are calculated. Assuming representativeness, the expansion costs are applied to all other grids in the given cluster. By multiplying the investment by the number of grids in the cluster we obtain the extrapolated investment for the entire cluster.

Applying the scheme from Table 2 to the real case, grid PV penetration reinforcements will require investments of about CHF 8M until 2035 for the sample of 143 grids. Extrapolating the costs to the entire DSO service area leads to investment requirements of several hundred million Swiss francs, emphasizing the strategic and financial challenge.

DSS Evaluation

The DSS is evaluated regarding the criteria specified in the system requirements section. Thereby, we describe the impact of the DSS on the DSO's strategic planning processes. Additionally, we consider the industry-standard planning process, as well as alternative planning approaches proposed by research.

Faster Strategic Planning Process

The DSS drastically increases the speed of strategic grid planning at the DSO (Table 3). Using established software, power flow analysis and bottleneck detection have always been possible for *individual* grids. However, there was no software for evaluating a larger number of grids and for adding related data, such as the PV potential. Such analyses required the collection of data on the assets' topological characteristics, i.e. each component's length as well as technical predecessor and successor in the respective power grid. This data collection process was a time-consuming manual effort requiring up to five man-hours for a single grid (depending on grid size). Even today, many distribution grid operators, particularly the smaller ones, lack digital access to their grid data. Grid planners have to manually take measurements from grid plans, consuming even more time. This requires an effort of several man-weeks for gathering the sample of 150 grids used in the process demonstration. Correspondingly, the industry standard for simple if-then analyses, common to grid planning, mostly comprises laborious and mundane data collection processes.

In the research domain, Kerber (2011) improves upon this industry standard by employing a semi-automatic approach for grid digitization. Grid plans are printed and manually redrawn using standardized symbols in order to make them machine readable and to determine key parameters, such as line length, automatically. However, Kerber (2011) developed this approach against the background of a single use case (PV integration). Alternative use cases continue to require significant amounts of additional grid data gathering and preparation, severely limiting the flexibility and applicability of this approach. Dickert et al. (2013) describe a support tool (representative grids) for *automated* grid data aggregation and evaluation. However, the analyzed grid characteristics are fixed and cannot be adjusted to different use cases (no runtime measures are given to serve as a comparison). Moreover, there is a large stream of use-case-centered research (mostly PV integration) in the electrical engineering domain (e.g. Degner et al. 2011; dena 2012; Schlömer et al. 2014). Since its focus is not decision support, runtime comparisons are not possible.

The automated procedures of the DSS allow the scenario-specific evaluation of parameters within seconds for an individual grid. Additionally, the automatic procedure is modular and highly flexible, enabling a rich variety of use cases, scenarios, and parameters to be added with minimal overhead. Even for large service areas the calculation of additional parameters is typically possible within minutes.

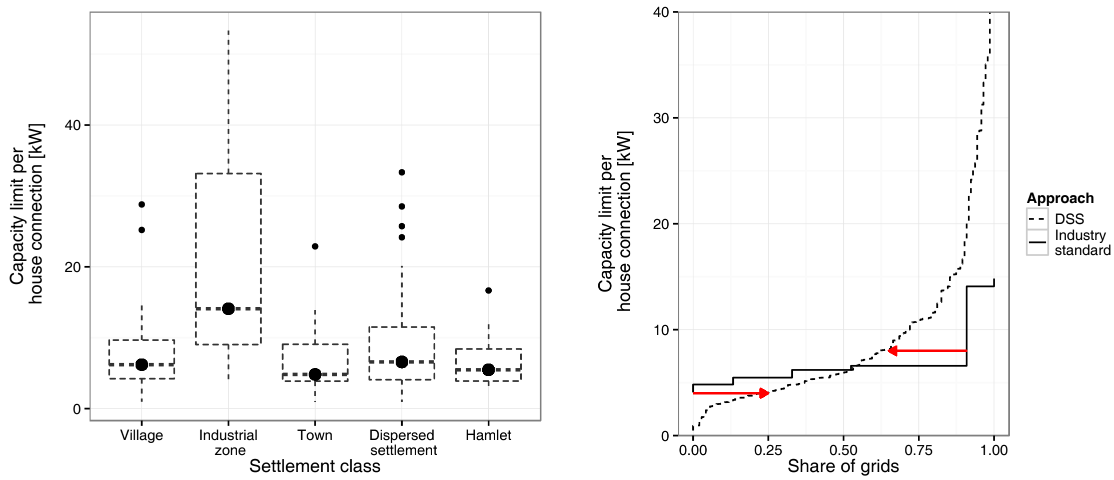
Table 3. Comparison of representative grid generation approaches.

Procedure	Time per grid	Use cases
Manual (industry standard)	1-5 hours	General, fixed
Semi-automated (Kerber 2011)	Hours	Fixed (PV integration)
Automated (Dickert et al. 2013)	—	General, fixed
Automated (our system)	**Seconds**	**General, adaptable**

Improved Planning Accuracy

Through more precise estimation of **grid capacity limits** and **expansion costs**, the DSS has significantly enhanced overall planning accuracy. We illustrate this gain in accuracy by means of an ex-post assessment of the previous planning procedures. While the improvements apply to all use cases, we focus on the example of

PV capacity limit analysis here. Both charts in Figure 10 vertically display the capacity limits for PV feed-in. Figure 10a illustrates the approach of classifying distribution grids according to the settlement patterns in the area which is the current, standard approach in the industry. For each settlement pattern, representative grids (dots) are defined. The grids are either defined by expert opinion or constructed by averaging some easily measurable parameters such as line lengths and resistances. In the best case, the capacity limit of the representative grid will be exactly at the median level of the cluster — as assumed in Figure 10a.



(a) Industry standard (settlement classes) (b) Industry standard vs. DSS

Figure 10. Determination of grid capacity limits

Applying the industry-standard, manual grid planning process, it was virtually impossible to determine the capacity limit for a large number of grids. Therefore, the capacity limits of all grids in the cluster were assumed to be similar to the representative grid. *(Ex-post the box plots provided by the DSS highlight actual heterogeneity in grids' specific capacity limits.)* Figure 10b compares the share of grids with PV capacity limits below the indicated level using both the DSS and the standard approach. Since the limits are known only for the representative grids, the standard approach yields a step function with five levels.

The more precise determination of **grid capacity limits** with the DSS translates into more accurate expansion planning: The deviations of the step function from the exact curve signify previous planning inaccuracies. In areas where the step function is above the exact curve, the number of grids experiencing a bottleneck was underestimated. For instance, at a PV feed-in of 4 kW per house connection, no capacity limits were detected. Now the DSS reveals that at this level as many as 25% of the grids need expansions (as highlighted with the lower arrow). On the contrary, at a PV penetration of 8 kW, bottlenecks used to be detected in 91% of the grids while in fact only 65% needed reinforcements. In this case, necessary investment would have been overestimated (upper arrow).

An improved estimation of grid **expansion costs** is achieved by virtue of a more suitable clustering of the grids. Figure 10a reveals that grouping according to settlement patterns is not ideal for the evaluation of PV capacity limits. The medians in all classes, except for industrial zones, are almost identical, while intra-class deviation is quite high. With groups based on settlement type, grid capacity limits differ an average of 65% from their corresponding representative grid. In contrast, using the clustering approach with an identical number of clusters, the relative gap to the representative grid can be reduced to 16 % on average, a four-fold difference. Since the representative grids are subsequently used for manual expansion planning, it directly follows that we obtain a more accurate expansion-cost estimate for the entire cluster.

Increased Planning Flexibility

The DSS is highly flexible as it implements use cases encompassing various dimensions of grid planning which go far beyond standard power flow calculations. Since strategic grid planning is concerned with a range of challenges, considering the presented use cases separately may be inadequate to arrive at robust expansion and reinforcement decisions. Consequently, the DSS also facilitates the combination of use cases, illustrated by the PV integration example. Even if use cases cannot be combined analytically, the risk map provides a tool to simultaneously account for competing use cases. The DSS is also flexible regarding the addition of new analyses. Given standardized data gathering and preprocessing modules, new use cases can be added to the DSS and calculated within minutes for the entire service area. This stands in strong contrast to the previously discussed manual or semi-automatic approaches still prevalent in the industry.

Seamless Process Integration

During the development of the DSS, competing methods involving synthetically constructed representative grids (Dickert et al. 2013; Kerber 2011) have been evaluated. The drawback of expansion planning based on synthetic grids is, first, that results would differ substantially from approaches relying on actual networks, and, second, that grid planners would not be able to use their standard software tools. Additionally, when discussing such alternative implementations, grid planners kept challenging the analyses' results since they were not familiar with the methodology. In order to foster acceptance and usage of the DSS, we paid particular attention to the detail of integrating the DSS into existing processes of the strategic grid planning department. Grid planners can conduct expansion analyses with the same tool chain, as they are used to from operative grid expansion planning. Moreover, the DSS relies on the same databases that the department already accesses for their power flow and GIS software solutions.

Implications for Practice and Research

Based on the experiences and interactions with the stakeholders during the design process of the DSS, we were able to derive general organizational implications that we translate into three lessons for CIOs and IT managers facing analytics challenges. Thereafter, we describe the implications for research, i.e., additions to the knowledge base (Gregor and Hevner 2013; Hevner and Chatterjee 2010) established in the course of the artifact design process.

Lesson 1: Identify the Business Values of Analytics Now

"Every Swiss franc we invest now, will yield a highly leveraged future return." (Project sponsor)

The energy sector is faced with disruptive changes from new technologies, such as distributed renewable generation, battery storage and electric vehicles. Additionally, a policy push towards a more sustainable energy system, as well as regulatory measures to promote market liberalization, lead to lower margins and higher competitive pressure. At the same time, the ongoing inclusion of sensors and actors into the power grid, such as smart meters and controllable devices on the demand side, offers increased availability of data — which, in turn, can create new business opportunities. Data-driven decision support systems assist in both quantifying future uncertainty and in developing effective, actionable strategies to respond to the changes ahead. Such systems streamline the analysis of future developments. Therefore, they enable managers to tackle the associated challenges both in a more efficient and more nuanced fashion.

The increased permeation of IT is not unique to the energy sector. Recent practice-oriented research from different domains demonstrates how new data sources leveraged with analytics can create business value. Adopting the Mooney et al. (1996) framework can help IT managers to identify and structure business value opportunities of analytics within their company (Table 4) by mapping the value of IT to the corresponding business processes — i.e., at the operational level (activities of the firm's value chain) or at the management level (administrative, allocative and controlling activities). Besides the process level, IT business value opportunities can further be clustered by their scope—from informational via automational to transformational changes. This fits well within the analytics domain, where this scope is spanned by descriptive, predictive and prescriptive analytics (Holsapple et al. 2014).

For instance, at the utility, the DSS provides automated access and evaluation of grid data, thereby reducing the overhead for manual data-collection at the operational level. At the management level, the aggregated evaluation of grid data enables more informed decision-making. Both functionalities provide informational business value. Additionally, the DSS speeds up the utilities' planning processes by standardizing scenario

Table 4. Analytics business opportunities (adapted from Mooney et al. 1996)

Scope	Process level	Functionality and Impact
Informational (Descriptive)	Operational	Automated grid data access reduces overhead
	Management	Evaluation of aggregated grid data improves decision quality
Automational (Predictive)	Operational	Assisted/ automated reinforcement planning leads to higher decision throughput and quality
	Management	Standardization and routinizing of investment scenario analysis improves decision quality
Transformational (Prescriptive)	Operational	Proactive grid reinforcement planning redefines planning processes and reduces costs Technology consulting services create new revenue potentials

analyses for multiple use cases — which creates business value as a result of increased automation. However, not all potential benefits from analytics were realized during the project. Further business value in the automational and transformational dimensions, such as fully automated reinforcement planning, was identified, which may be tapped into in follow-up projects.

Lesson 2: Digitize your Corporate Knowledge

> *"The profound knowledge of our assets is our core competence. Especially with this [uncertain] future ahead."* (Senior manager)

Data availability is a prerequisite for the effectiveness of analytics systems. Many firms do not have ready access to essential data on assets and processes. By pushing forward digitization initiatives, firms can lay the foundations for future improvements in planning and operations. In the case study, the utility decided to proceed as follows: First the accuracy and completeness of historical asset data had to be improved. During the project, some information about the power grids proved incomplete and erroneous. For instance, for a large share of assets, a default value (1.1.1900) was stored in the system as the installation date. To facilitate use cases, such as condition-based maintenance, for the DSS, the company sent engineers out to the field to re-inspect the assets in order to gather or estimate the missing information. However, for predictive analytics, static data is insufficient. Going forward, the firm wants to use dynamic information on the actual use of the assets. This information is contained in consumer contract data, which is already digitally available in the sales department's information system. To this end, the implementation of an interface for data transfer was initiated. In the final stage, the aim is to process live data on the grids' assets, e.g. from recently rolled-out smart meters and other controllers, in order to actively monitor and optimize their usage.

This digitization strategy can serve as a blueprint for other firms facing similar challenges. The complexity of data processing increases incrementally as one expands the scope of the applied analytics. Therefore, data assets and analytics capabilities will have to be jointly developed. At first, only static, historical data is digitized. In the second step, dynamic consumption data is added and the volume of the data is significantly increased. The final step is of even larger complexity, since high-volume data needs to be processed at a high velocity in real time. The resources exploited in the digitization process do not have to be entirely internal. In the case at hand, geo-referenced data regarding the service area was purchased from a public cartography service provider. Consumer research data or economic indicators could be integrated in a similar fashion.

Lesson 3: Introduce Analytics Stepwise

> *"I do not follow a planning procedure as long as I do not understand and believe its methodology."* (Grid planner)

The adoption of analytics systems is not a binary yes-no decision, but is rather a question of investment timing (when to adopt the system) and scope (which system to adopt). Due to limited absorptive capacity of the system users, the implementation of analytics systems can face significant obstacles. Analytics systems constitute a complex technology that requires advanced user skills, e.g. in the areas of statistics and operations research. Thus, gradually implementing and upgrading an analytics system enables learning-by-doing (Fichman 2004) and may result in lower roll-out costs due to less training being required. Additionally, data quality impacts expected benefits from any analytics application (Hazen et al. 2014; Warth et al. 2011). An initial investment in a descriptive system permits one to get an overview of the available data and will help to better assess the expected additional benefits from predictive and prescriptive systems. Lastly, the value of analytics capabilities can also depend on external factors — such as changes in the regulatory environment. In the case of the utility, the benefits of the prescriptive system could not yet be realized due to missing regulatory incentives. Thus, the implementation of the functionalities was put on hold.

For the aforementioned reasons, we put forward *stepwise implementation* as a superordinate design paradigm for introducing analytics into practice. This principle is congruent with previous findings in practice-oriented research, referred to as "soft-OR" (Mingers 2000). In the early stages of the implementation, it may be worthwhile to pursue a hybrid approach which combines "hard," data-driven analytics methods (automated grid data evaluation) with subjective, expert opinions and manual analyses (expansion planning). At later stages, the latter components may be gradually replaced by more advanced analytical systems.

Implications for research

Rigorous design science research provides additions to the knowledge base in the form of lessons based on field experience, as well as new scientific methods, theories or meta-artifacts (Hevner and Chatterjee 2010). The artifact developed in this study belongs to the group of support tools based on representative grids. However, it is neither restricted to a particular use case, such as a large set of tools for investigating the impact of distributed generation (e.g. Hollingworth et al. 2013; Rudion et al. 2006), nor does it provide representative grids that apply universally to a wide range of use cases (Dickert et al. 2013; Strunz et al. 2009). To the best of our knowledge, it is the only system that allows the user to flexibly select representative grids according to critical parameters for a particular use case or for combined use cases. Additionally, by describing the design process, including user and management perspectives, we provide a contribution to the general IS and Energy Informatics domain. Finally, the experiences of this project have led to follow-up research in both electrical engineering and IS. Gust et al. (2016) describe a prototype for automated distribution grid reinforcement planning that may eventually replace the manual expansion cost projections. In a research project directed at the IS community, we are currently investigating how future opportunities for system upgrades can be accounted for in present analytics investment decisions using a real-options approach.

Conclusion

Following a design science approach, this paper presents the iterative implementation of a DSS for grid planning at a large Swiss utility. Based on a requirements analysis, we describe the design process and the structure of the resulting decision support system. The core part of the DSS, the analytics module, evaluates key characteristics of very large numbers of grids with regards to a variety of use cases, including electricity feed-in and consumption, CAPEX and reliability. Since the DSS is implemented in practice, we demonstrate the operation of the DSS analyzing necessary grid investments due to projected future growth of distributed photovoltaic generation.

The project's contributions are threefold. First, at the company level, the system evaluation demonstrates how the speed of the planning process, as well as its accuracy and flexibility, could be improved. The system helps to minimize manual data collection activities and replaces planning decisions based on expert opinion with a data-driven approach. Second, for engineering and DSS research, the system extends existing support systems for grid planning by providing a flexible, use-case-centered approach according to the needs of the system users. Third, for IS practice, the experiences gained from the implementation translate into general lessons for IS executives who face analytics challenges outside the energy sector.

Regarding the limitations, integration into existing processes has reduced the design alternatives of the DSS.

Additionally, the DSS does not provide the degree of process automation of recently developed prototypes that employ fully-automated planning procedures (Gust et al. 2016; Hollingworth et al. 2013; Schrader 2014). Such extensions were discussed for follow-up projects with the management. However, these have not been pursued (yet), mainly because part of the cost-savings could not be internalized due to a lack of regulatory incentives.

References

ASCE 2011. *Failure to Act: The Economic Impact of Current Investment Trends in Electricity Infrastructure*. American Society of Civil Engineers.

BFE 2013. *Energieperspektiven 2050*. URL http://www.bfe.admin.ch/themen/00526/00527/.

Chen, H., Chiang, R. H., and Storey, V. C. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS Quarterly* (36:4), pp. 1165–1188.

Davenport, T. H. and Harris, J. G. 2007. *Competing on analytics: The new science of winning,* Harvard Business Press.

Dawson, D., Simpkins, E., and Stillmann, J. 2013. "Waiting for the Digital Grid," *strategy+business* (73).

Degner, T., Arnold, G., Reimann, T., Engel, B., Breede, M., and Strauss, P. 2011. "Increasing the photovoltaic-system hosting capacity of low voltage distribution networks," in *Proceedings of 21st International Conference on Electricity Distribution,* pp. 1243–1246.

dena 2012. *dena Distribution Grid Study*. Deutsche Energie Agentur GmbH.

Dickert, J., Domagk, M., and Schegner, P. 2013. "Benchmark low voltage distribution networks based on cluster analysis of actual grid properties," in *Proceedings of IEEE PowerTech 2013,* pp. 1–6.

Fichman, R. G. 2004. "Real options and IT platform adoption: Implications for theory and practice," *Information Systems Research* (15:2), pp. 132–154.

Goebel, C., Jacobsen, H.-A., del Razo, V., Doblander, C., Rivera, J., Ilg, J., Flath, C. M., Schmeck, H., Weinhardt, C., Pathmaperuma, D., Appelrath, H.-J., Sonnenschein, M., Lehnhoff, S., Kramer, O., Staake, T., Fleisch, E., Neumann, D., Strüker, J., Erek, K., Zarnekow, R., Ziekow, H., and Lässig, J. 2014. "Energy Informatics," *Business & Information Systems Engineering* (6:1), pp. 25–31.

Gordon, S., Blake, R., and Shankaranarayanan, G. 2013. "Case-based research in information systems: Gaps and trends," *JITTA: Journal of Information Technology Theory and Application* (14:2), p. 47.

Gregor, S. and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact." *MIS quarterly* (37:2), pp. 337–355.

Gust, G., Biener, W., Brandt, T., Dallmer-Zerbe, K., Neumann, D., and Wille-Haussmann, B. 2016. "Decision support for distribution grid planning," in *2016 IEEE International Energy Conference (ENERGYCON),* IEEE, pp. 1–7.

Hazen, B. T., Boone, C. A., Ezell, J. D., and Jones-Farmer, L. A. 2014. "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications," *International Journal of Production Economics* (154), pp. 72–80.

Hevner, A. and Chatterjee, S. 2010. *Design research in information systems,* vol. 22. Springer Science & Business Media.

Hollingworth, D., Mukherjee, R., Hodges, G., Miller, D., and Lyons, P. 2013. "A Network Planning and Decision Support Tool for Integration of Low Carbon Technologies and Solutions," in *22nd International Conference on Electricity Distribution (CIRED 2013),*

Holsapple, C., Lee-Post, A., and Pakath, R. 2014. "A unified foundation for business analytics," *Decision Support Systems* (64), pp. 130–141.

IEC 2015. *Strategic asset management of power networks*. International Electrotechnical Commission.

Kaufman, L. and Rousseeuw, P. J. 2009. *Finding groups in data: an introduction to cluster analysis,* vol. 344. John Wiley & Sons.

Kerber, G. 2011. "Aufnahmefähigkeit von Niederspannungsverteilnetzen für die Einspeisung aus Photovoltaikkleinanlagen," PhD thesis. Technische Universität München.

Lim, E.-P., Chen, H., and Chen, G. 2013. "Business intelligence and analytics: Research directions," *ACM Transactions on Management Information Systems (TMIS)* (3:4).

Melville, N. P. 2010. "Information systems innovation for environmental sustainability," *MIS quarterly* (34:1), pp. 1–21.

Mingers, J. 2000. "Variety is the spice of life: combining soft and hard OR/MS methods," *International Transactions in Operational Research* (7:6), pp. 673–691.

Mooney, J. G., Gurbaxani, V., and Kraemer, K. L. 1996. "A process oriented framework for assessing the business value of information technology," *ACM SIGMIS Database* (27:2), pp. 68–81.

Mortenson, M. J., Doherty, N. F., and Robinson, S. 2015. "Operational research from Taylorism to Terabytes: A research agenda for the analytics age," *European Journal of Operational Research* (241:3), pp. 583–595.

Rudion, K., Orths, A., Styczynski, Z. A., and Strunz, K. 2006. "Design of benchmark of medium voltage distribution network for investigation of DG integration," in *IEEE Power Engineering Society General Meeting, 2006*. IEEE, pp. 6–11.

Schlömer, G., Reese, C., and Hofmann, L. 2014. "Methode zur automatisierten Bewertung des zukünftigen Ausbaubedarfs in der Niederspannungsebene unter Berücksichtigung verschiedener technischer Konzepte," in *13. Symposium Energieinnovation, Graz/Austria,*

Schrader, S. 2014. "Netzplanung 2.0 – Simulationstool ersetzt lange Planungszeiten," *ew - Magazin für die Energiewirtschaft* (10).

Sharma, R., Mithas, S., and Kankanhalli, A. 2014. "Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations," *European Journal of Information Systems* (23:4), pp. 433–441.

Sprague, R. H. 1980. "A Framework for the Development of Decision Support Systems." *MIS Quarterly* (4:4), pp. 1–26.

Strunz, K., Fletcher, R., Campbell, R., and Gao, F. 2009. "Developing benchmark models for low-voltage distribution feeders," in *IEEE Power & Energy Society General Meeting, 2009. PES'09*. IEEE, pp. 1–3.

vom Brocke, J., Fridgen, G., Hasan, H., Ketter, W., and Watson, R. 2013. "Energy Informatics: Designing a discipline (and possible lessons for the IS community)," in *Proceedings of the 2013 International Conference on Information Systems (ICIS 2013),*

Warth, J., Kaiser, G., and Kügler, M. 2011. "The impact of data quality and analytical capabilities on planning performance: insights from the automotive industry." in *Wirtschaftsinformatik,* p. 87.

Watson, R. T., Boudreau, M.-C., and Chen, A. J. 2010. "Information systems and environmentally sustainable development: energy informatics and new directions for the IS community," *MIS Quarterly* (34:1), pp. 23–38.