IDENTIFYING THE ROLE OF INFORMATION SYSTEMS IN ACHIEVING ENERGY-RELATED ENVIRONMENTAL SUSTAINABILITY USING TEXT MINING

Nadine Reuter
University of Liechtenstein, Vaduz, Liechtenstein, nadine.reuter@uni.li

Svitlana Vakulenko
University of Liechtenstein, Vaduz, Liechtenstein, svitlana.vakulenko@uni.li

Jan vom Brocke
University of Liechtenstein, Vaduz, Liechtenstein, jan.vom.brocke@uni.li

Stefan Debortoli
University of Liechtenstein, Vaduz, Liechtenstein, stefan.debortoli@uni.li

Oliver Müller
University of Liechtenstein, Vaduz, Liechtenstein, oliver.mueller@uni.li

Nadine Reuter, Svitlana Vakulenko, Jan vom Brocke, Stefan Debortoli, and Oliver Müller, 2014, "IDENTIFYING THE ROLE OF INFORMATION SYSTEMS IN ACHIEVING ENERGY-RELATED ENVIRONMENTAL SUSTAINABILITY USING TEXT MINING", Proceedings of the European Conference on Information Systems (ECIS) 2014, Tel Aviv, Israel, June 9-11, 2014, ISBN 978-0-9915567-0-0
http://aisel.aisnet.org/ecis2014/proceedings/track22/12

IDENTIFYING THE ROLE OF INFORMATION SYSTEMS IN ACHIEVING ENERGY-RELATED ENVIRONMENTAL SUSTAINABILITY USING TEXT MINING

Complete Research

Reuter, Nadine, University of Liechtenstein, Vaduz, Principality of Liechtenstein, nadine.reuter@uni.li

Vakulenko, Svitlana, University of Liechtenstein, Vaduz, Principality of Liechtenstein, svitlana.vakulenko@uni.li

vom Brocke, Jan, Martin Hilti Chair of Business Process Management, University of Liechtenstein, Vaduz, Principality of Liechtenstein, jan.vom.brocke@uni.li

Debortoli, Stefan, University of Liechtenstein, Vaduz, Principality of Liechtenstein, stefan.debortoli@uni.li

Müller, Oliver, University of Liechtenstein, Vaduz, Principality of Liechtenstein, oliver.mueller@uni.li

Abstract

Environmental sustainability is one of the most important topics of our time. Energy consumption is a significant aspect of environmental sustainability, as it is a major source of pollution. In order to determine how information systems can assist organizations in reducing energy consumption, we first identify energy-related activities of organizations by examining the content of 4,999 sustainability reports that communicate corporate activities directed toward sustainability. We use latent semantic analysis (LSA) to extract the most prominent topics discussed in these reports and the concept of affordances in order to portray the potential role of information systems in supporting these activities. We identify five such affordances: (1) balancing supply and demand with consideration of the energy price; (2) capturing, reporting, and sharing energy-related data; (3) reducing energy consumption and emissions; (4) sustainable waste management; and (5) coordination of flexible charging of electric vehicles. These affordances support energy management through information systems in practice and set a foundation for further research in the important field of energy informatics.

Keywords: Environmental Sustainability, Energy Informatics, Latent Semantic Analysis, Sustainability Reports.

1 Introduction

Environmental sustainability is one of the most important topics of our time (Melville, 2010). Being concerned with "minimiz[ing] the negative impacts and maximiz[ing] the positive impacts of human behavior on the environment" (Elliot, 2011, p.208), environmental sustainability asks for fundamental changes in the behavior at the individual, organizational, and societal levels (Elliot, 2011). Several researchers have identified information systems as potential enablers of such sustainable practices (Elliot, 2011; Seidel et al., 2013; Melville, 2010).

Energy consumption is an important aspect of environmental sustainability, as it is a major cause of pollution (Watson et al., 2010). In particular, fossil fuels are a primary cause of global warming

(Hoffert et al., 2002), and the level of energy consumption is not likely to wane (Califf et al., 2012). We refer to environmental sustainability that focuses on the energy as energy-related environmental sustainability. Analogue to Goodland's (1995) definition of environmental sustainability, we understand energy-related environmental sustainability as the practice of using only as much energy as the environment can provide without reducing existing resources to a bigger extent than they can be reproduced and producing thereby only as much emission as the environment can absorb.

Information Systems (IS) research has reacted to this challenge with investigations in the fields of green IS and energy informatics. Green IS investigates the role of information systems in designing and implementing sustainable business processes (Melville, 2010; Watson et al., 2010), while energy informatics focuses on how to reduce energy consumption through information systems (Watson et al., 2010). A first conceptualization of the role of information systems in reducing energy consumption is the energy informatics framework (Watson et al., 2010), which shows the role of information systems as a key element of the energy supply and demand system. Later research focuses on specific aspects of the energy informatics framework, such as the adoption of smart meter technology (Wunderlich et al., 2012) or the improvement of individual energy consumption behavior through social competitions (Yim, 2011) and social normative feedback (Loock et al., 2011). While Watson et al. (2010) focus on a global perspective and Wunderlich et al. (2012), Yim (2011), and Loock et al. (2011) focus on the individual level, organizations have a substantial influence on the ecosystem as well (Elliot, 2011). Therefore, in order to determine how information systems can help to reduce energy consumption, we focus on the organizational level. To do so, we must first answer the question:

What are the most prominent energy-related environmental sustainability topics for organizations?

In answering this question, we conduct a content analysis on sustainability reports, which organizations use to communicate their sustainability activities (Bradford et al., 2012). In particular, we use latent semantic analysis (LSA), a quantitative content analysis technique (Dumais et al., 1988; Landauer et al., 1998) to identify major topics and keywords in the context of energy.

Based on the results of the LSA, we discuss the potential role information system can play in supporting organizations' energy-related activities. To do so, we build upon the concept of affordances (Markus and Silver, 2008) as a theoretical lens, focusing on the actions that information systems should enable.

Our study makes three primary contributions:

(1) We identify the most prominent topics of energy-related sustainability in organizations.

(2) We propose how information system can enable activities in these areas.

(3) We show the applicability of text-mining techniques, in particular for the exploration and analysis of organizational reports, for the purpose of IS research.

We proceed as follows. First we present the related research on energy informatics and the concept of affordances, followed by a description of our method. Then we present the results of our analysis. Finally, we discuss these results and conclude the paper with suggestions for further research.

2 Related Research

2.1 Energy Informatics

Energy informatics, a sub-field of IS, was first established by Watson et al. (2010), whose core idea is that energy + information = less energy. Therefore, energy informatics "is concerned with analyzing, designing, and implementing systems to increase the efficiency of energy demand and supply systems" (Watson et al., 2010, p.24). These authors, who focus on the use of information systems in reducing energy consumption, develop an energy informatics framework, the key element of which is an information system that shows the relationship between the demand side and the supply side. The major components of the integrated system are flow networks, sensor networks, and sensitized objects. Flow networks are “a set of connected transport components that support [...] the movement of continuous matter (e.g., electricity, oil, air, and water) or discrete objects (e.g., cars, packages, containers, and people)” (Watson et al., 2010, p.26). These flow networks are optimized by data provided by sensor networks, which report “the status of a physical item or environmental condition” (Watson et al., 2010, p.26). Sensitized objects, on the other hand, are physical goods that report their own use. These three components are managed by information systems and are used by key stakeholders of the energy system: suppliers, customers, and governments. Watson et al. (2010, p.27) define suppliers as those who “provide energy (e.g., natural gas) or services that consume energy (e.g., air conditioning),” while consumers are the ones who consume energy and “bear the ultimate cost of all energy consumption.” While suppliers and consumers are mainly focused on their own interests, governments have to focus on society’s long-term interests through such means as regulations and additional costs for poor environmental behavior (Watson et al., 2010).

Research in the field of energy informatics has focused on particular aspects of the energy informatics framework, primarily power supply, smart grids, and power consumption (Kossahl et al., 2012). Research on power supply has developed tools to help organizations cope with the price volatility that is due to fluctuating supply and demand. For instance, a demand-side management tool could help to adjust demand based on the current energy price (Feuerriegel et al., 2012), while a market for consumption rights could promote shifting demand from peak demand times to times with lower demand (Watson et al., 2013). Smart grids are electricity-distribution networks that manage the supply of and demand for electricity” (Niesten, 2010), and smart metering technology provides information about current processes, energy consumptions, and available energy in the grid (Wunderlich et al., 2012). Adoption of these technologies is an important topic for IS (Wunderlich et al., 2012). However, the adoption of these technologies is of in jeopardy because of data-security concerns (Strüker and Kerschbaum, 2012). Encryption is a possible solution for this issue, but since the retailer must still decrypt the data in order to use it, perfect data privacy cannot yet be ensured (Strüker and Kerschbaum, 2012).

Data gathered by smart meters can be used to improve individuals’ energy consumption. Research on changes in individuals’ energy consumption behavior reveals that goal-setting (Loock et al., 2013), insights into the personal consumption data, guidelines on how to improve the consumption behavior, and incentives lead to energy savings (Graml et al., 2011). Social competitions (Yim, 2011) and social normative feedback (Loock et al., 2011) may also motivate individuals to reduce their energy consumption. On the organizational level, the focus lies on green information technology (IT). Green IT considers the whole lifecycle of IT products and tries to design them in an environmentally friendly way. Such designs include energy-efficient operation of IT resources themselves (Molla et al., 2009), such as through grid technologies (Vykoukal, 2010) and solutions to handle scheduling problems in data centers (Brandt and Bodenstein, 2012).

2.2 The Concept of Affordances

The concept of affordances, which originates in ecological psychology, was developed by Gibson (1986) to describe relationships between animals and the environment. According to Gibson (1986), objects’ affordances (the tasks they facilitate), not their qualities (e.g., color, size), are perceived when they are observed. In comparison to physical properties, which are universal and can be measured objectively, affordances can be measured only in relation to who or what is perceiving that affordance and the environment in which it exists (Gibson, 1986). For example, a human may perceive bark as an affordance for making a canoe, while a deer would perceive bark as a source of nourishment—and neither one would necessarily perceive bark as brown.

Affordances in IS have been suggested to replace the concept of structural features (Markus and Silver, 2008), which are the rules and resources incorporated into the IT system and offered to the user (DeSanctis and Poole, 1994). Functional affordances differ from structural features since they are not perceived as a part of the properties of the technology but as a relationship between the technology (the technical object) and the specified user group. Similar to structural features, functional affordances represent the social structures that may be connected to the IT artifact (Markus and Silver, 2008). Therefore, the concept of affordances avoids a deterministic view of the use of IT (Strong et al., 2009).

The concept of affordances can be also used to explain the imbrication of human and material agencies. The material properties of technologies are interpreted differently and afford different action possibilities based on the context in which they are used (Leonardi, 2011). Therefore, previous research tends to start with IT artefacts and examine the action possibilities that emerge in relationship to certain user groups. For instance, Leonardi (2011) gives an empirical illustration of imbrication by investigating the activities that occur around a computer simulation technology, and Seidel et al. (2013) examine the functional affordances that information systems create in the context of environmental sustainability transformations.

Gibson's (1986) work was also adopted by the human-computer interaction (HCI) community, where Norman (1999) distinguishes between real and perceived affordances. In the case of computers, the real affordances are usually built in (e.g., the component parts of a computer system like the keyboard and screen) to afford pointing, touching, and so on. Designers can change the perceived affordances by, for instance, changing the elements on the screen that advertise the underlying action possibilities (Norman, 1999).

In our research we aim to identify affordances that are important in a particular context of use. Based on the identified affordances, future research can determine which material properties are required to support these affordances. In taking this approach we partially follow Norman (1999), as we accept that designers can create affordances strategically (Leonardi, 2011).

3 Method

In order to determine the most prominent energy-related topics in organizations, we take a quantitative approach by performing a content analysis of sustainability reports. We apply a semi-automated method in order to take advantage of the extensive amount of publicly available data. The heterogeneous, unstructured nature of the data requires the use of text-mining techniques (Hotho et al., 2005). Using exploratory statistical analysis of textual data, we aim to identify the common topics in the documents. We employ LSA, which is widely used in text analytics, for topic modelling (Landauer et al., 1998; Dumais et al., 1988; Landauer et al., 2004).

LSA is a statistical technique first proposed for the task of information retrieval (Dumais et al., 1988). In IS research, it was repeatedly applied for content analysis of large document collections, such as, academic publications (e.g., Larsen et al. 2008; Sidorova et al. 2008), social media posts (e.g., Evangelopoulos & Visinescu, 2012), job advertisements (Müller et al. , n.d.), and vendor case studies (Herbst et al., 2014). LSA is a quantitative method for analyzing textual data that exploits latent (not directly observed) dependencies between the words and their contexts (Nakov, 2000). More specifically, LSA is based on the Vector Space Model (Salton et al., 1975), where each document is represented as a vector with values corresponding to the frequencies of terms in the document, resulting in a term-document matrix. The core of LSA is Singular Value Decomposition (SVD) (Landauer et al., 1998; Berry et al., 1995), which splits the term-document matrix into factors that can be interpreted as topics identified in the underlying document collection.

The detailed steps we conducted in our study are (1) data collection, (2) data pre-processing, and (3) analysis (LSA).

3.1 Data Collection

We chose corporate sustainability reports as a source through which to explore the most prominent energy-related topics in companies, as they reveal companies' sustainability-related activities (Bradford et al., 2012). A centralized repository of sustainability reports is maintained by the Global Reporting Initiative (GRI), a leading organization in the sustainability field, which promotes organizations' use of the Sustainability Reporting Framework (GRI, 2013b). Organizations can register their sustainability reports with GRI, which checks their application of the Sustainability Reporting Framework before publishing the reports in the GRI database (GRI, 2013a).

We collected 7,535 sustainability reports published between 1999 and 2013 for our study. We retrieved the PDF documents from the GRI website[1] using a custom automated crawler.

3.2 Pre-processing

We converted the collected PDF documents into plain text, which extracts the textual data and omits the formatting and graphical data, in order to perform automated text analytics.

The majority of the reports are written in English; to avoid translation errors, we decided to focus on the English-language documents only, which reduced the number of reports in our collection to 4,999 English-language reports.

In order to extract the information about energy-related topics that is relevant to our study, we split the original documents into pages and took only those that contained the term "energy." This simple heuristic reduced significantly the amount of textual data to be processed, since a sustainability report usually consists of 60 - 100 pages. With this approach we filtered out the irrelevant parts of the reports, such as parts that were concerned with the economic or social aspects of sustainability. From now on we use the term "document" in relation to the energy-related pages, not the original sustainability reports.

The resulting textual documents were tokenized (split into tokens, which include words and special symbols, such as punctuation), all the characters were brought into lower-case, and the words were lemmatized to reduce the words into their base forms (plural to singular for nouns, verbs to the simple present tense) (Jivani, 2011). We chose the lemmatizing approach over stemming (Porter, 1980), as it provides a smoother way to deal with lexical heterogeneity while retaining the meaning of the original word. Stems are often more difficult to relate to the original concepts, so they are more difficult to interpret (Evangelopoulos et al., 2010). For example, "products" and "production" are both reduced to "product" with stemming, while lemmatization separates the meanings into "product" and "production".

Even after extracting energy-related pages and eliminating standard stopwords (i.e., general-purpose words like articles, pronouns, and conjunctions), the number of terms was overwhelming, so we constructed a term-document matrix by counting the frequencies of the tokens (unigrams) and two-word combinations (bigrams) for each document separately. After filtering out the terms that appeared in less than 1 percent of the documents, we obtained a list of 3,815 terms. Two of this paper's authors then selected the energy-related terms independently from each other, compared and discussed the results, and compiled the final go-list. The degree of agreement between the independent researchers was 79.2 percent, which is a substantial strength of agreement, according to Landis and Koch-Kappa's Benchmark Scale (Landis and Koch, 1977). The go-list, which consists of 722 energy-related terms, frames the scope of our analysis, as only these terms are considered in the further analysis. The go-list

[1] http://database.globalreporting.org/search (accessed in September 2013)

includes unigrams and bigrams that are related to energy consumption, energy sources, waste emission, climate change, recycling, governmental policies, and regulation.

3.3 Latent Semantic Analysis (LSA)

The basis for LSA is a term-document matrix that is constructed using the go-list and that contains the term frequencies for each term-document pair. The term frequency-inverse document frequency (TF-IDF) weighting function is applied to the term-document matrix to normalize term frequencies (Evangelopoulos et al., 2010) by de-emphasizing the role of terms that simply appear frequently and highlighting the terms that are particularly relevant to a specific document.

Next, we performed Singular Value Decomposition (SVD) on this weighted term-document matrix. SVD requires the number of LSA factors to be predefined (Deerwester, Dumais, Furnas et al. 1990). As there is no single methodology on the choice of dimensionality (the number of factors) and LSA performance evaluation established, researchers have been encouraged to experiment with different numbers of dimensions and to apply their domain knowledge when choosing an appropriate LSA model (Evangelopoulos et al., 2010). We chose ten- and fifty-factor LSA models for our analysis in order to obtain a top-level overview of the common topics while also gaining additional insights by observing how the factors are further subdivided.

The results of the SVD are two matrices with equal numbers of columns that represent factors. One of the matrices contains term-loadings for each of the factors, representing the values that reflect the relationship of each of the terms to each of the factors. The second matrix contains document loadings, which are the values that reflect the relationship of the factors to the underlying documents. In order to obtain interpretable dimensions, we followed the recommendation of Evangelopoulos et al. (2010) to apply varimax rotation to the matrices. Varimax rotation (Kaiser, 1958) is used in factor and principal component analysis to recompute the values in a matrix in order to maximize the sum of the variance. This transformation makes the factors more distinct and uncorrelated (orthogonal) and, therefore, easier to interpret.

The higher a value in the loading matrix is, the more relevant the factor is for the respective term or document. We identified high-loading documents and terms using the threshold selection method (Evangelopoulos et al., 2010; Sidorova et al., 2008), based on the assumption that each document-term pairing loads on one dimension on average. The top-n largest values are considered high-loading, and these are the only factors considered in the rest of the analysis.

Two of the authors independently analyzed, interpreted, and labelled the resulting factors. Then the outcomes were compared and discussed, resulting in the interpretations and findings reported in the next sections.

4 Results

We chose a ten-factor analysis since it provides us with a good overview of the most relevant topics for organizations. Table 1 shows the describing label and the most high-loading terms for each factor. The first factor (V1), environmental management, which has the highest loading, explains most of the variance in our data. The highest-loading factor usually contains the most common terms in the collection (Hu et al., 2003). In our case, the highest-loading factor includes such terms as “development”, “product”, and “policy”, which are not particularly useful for our purposes as it is rather a background topic.

Among the remaining nine factors in this analysis, three are dedicated to the area of energy generation: the renewable energy sources (V2) of wind, solar, and water; nuclear energy (V8); and the non-renewable energy sources (V10) of oil, gas, and coal.

Awarding (V3), which is related to prizes and innovations, emerges as another factor in this part of the analysis, as does emission and climate impact (V4), which is one of the most widely discussed topics in the context of energy generation and consumption. Energy network (V5) refers to the transmission of generated energy through the energy grid, while V6 focuses on data, particularly data about consumption, emission, performance indicators, and measurements. V7 is concerned with waste management, enriched with the notion of recycling, as in the context of energy, waste management can refer to energy that is generated from waste. Finally, V9 covers specific terminology related to the costs of, demand for, and trade of energy.

Factor	Factor label	Examples of high-loading terms
V1	Environmental management	development, product, policy, supplier, innovation, environmental, activity, performance, technology, green
V2	Renewable energy generation	wind, solar, wind farm, power, renewable, renewable energy, capacity, wind energy, wind power, plant, hydroelectric
V3	Awarding	award, prize, awarded, green, ministry, contribution, innovation
V4	Emission and climate impact	emission, carbon, climate, ghg, climate change, tonne, gas, green, change, greenhouse
V5	Energy network	solar, electric, grid, power, electricity, transmission, network, vehicle, smart, consumption
V6	Data	indicator, data, indirect, performance, consumption, water, emission, direct, supplier, energy consumption
V7	Waste management	waste, water, recycling, plant, ton, product, recycled, hazardous, production, tonne
V8	Nuclear power generation	nuclear, power, plant, power plant, coal, nuclear power, electric, electricity, generation, electric power
V9	Energy economics	price, loss, cost, electricity, trading, increase, transmission, demand, change, increased
V10	Non-renewable energy generation	oil, gas, natural gas, coal, production, natural, fuel, oil gas, exploration

Table 1 Results of ten-factor LSA

Figure 1 shows the comparative population of the factors in the collection by providing the ratio between the number of high-loading documents for each of the factors and the total number of documents. In other words, the higher the number of documents that are concerned with a topic, the higher the percentage of the corresponding factor is. Thus, we can judge the popularity of an identified topic.

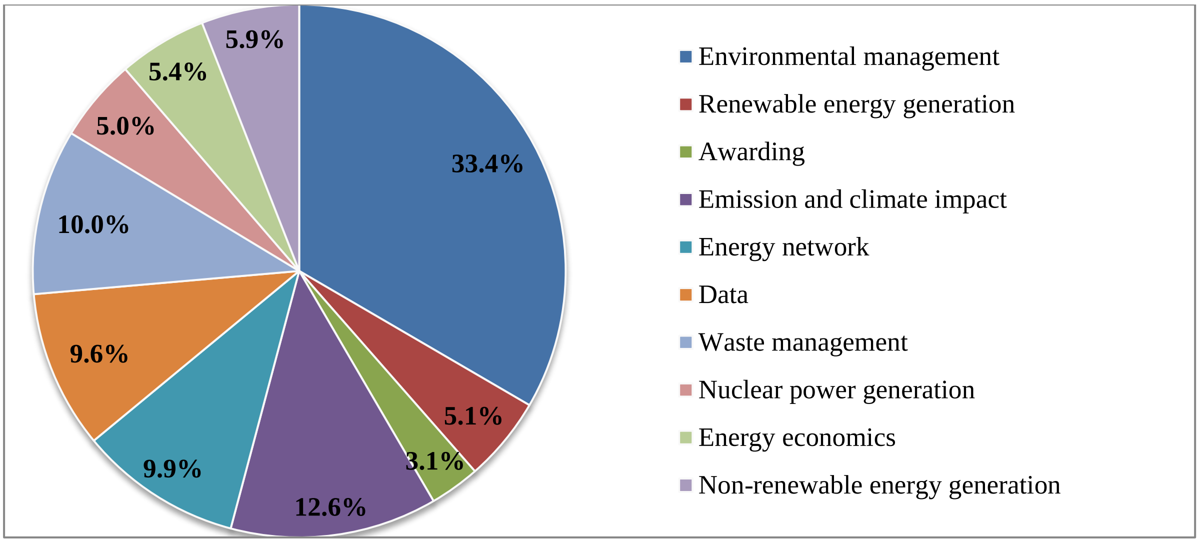

Figure 1. Semantic space structure[2].

As the selection of the optimal number of LSA dimensions is still an unresolved methodological issue, it has been recommended that researchers also provide models with alternative dimensionality (Evangelopoulos et al., 2010). Consequently, we conducted a fifty-factor analysis in order to gain deeper insights into the most frequently discussed topics in the sustainability reports. In a fifty-factor analysis the factors in the ten-factor analysis are divided into more specific factors and combinations of terms that were not prominent in the ten-factor analysis occur, allowing us to examine a more detailed calibration of the energy-related factors. Because of space limitations, we do not report on all findings from the fifty-factor analysis but present only those findings that are particularly relevant.

Examples of factors in the ten-factor analysis that are divided into more specific factors in the fifty-factor analysis are the energy sources, where renewable (V2) and non-renewable energy sources (V10) are subdivided into the separate factors for solar energy (high-loading terms: solar panel, sun), wind energy (wind farm, offshore, turbine, capacity), oil and gas (exploration, refinery, petroleum), coal (coal), and natural gas (gas, natural gas, pipeline, distribution). Five topics in particularly emerged in the fifty-factor analysis: (1) energy grid, which is described with terms like transmission, electricity, and distribution; (2) energy saving, with a focus on lighting and energy efficiency; (3) vehicles and the forms of vehicle power (e.g., fuel, hybrid, electric, diesel); (4) smart technology, described by smart grid, meter, technology, and ICT; and (5) performance indicators, which are specified by GRI indicator, key performance, environmental, and social.

5 Discussion

Our analysis reveals ten energy-related topics that are often reported by organizations. Next, we identify issues that might occur in the context of these energy-related topics and then make assumptions about possible affordances that could be enabled by information systems in the context of energy-related environmental sustainability to solve these issues.

Our resulting factors reveal a distinction between external and internal topics, so we structure our discussion based on an external, holistic view and an internal view. The external, holistic group is concerned with the generation, transmission, and distribution of energy and consists of the factors of renewable energy generation (V2), energy network (V5), nuclear power (V8), energy economics

[2] 22 percent of the documents do not belong to any of the factors and are not shown in Figure 1

(V9), and non-renewable energy generation (V10). The internal factors consist of environmental management (V1), awarding (V3), emission and climate impact (V4), data (V6), and waste management (V7).

Our results show that companies' environmental sustainability reports have a strong emphasis on energy generation, as organizations report on topics related to both renewable energy (V2) and non-renewable energy generation (V10). These various sources of energy generation imply several issues. In particular, because of renewable resources' weather dependency, price fluctuations (Feuerriegel et al., 2012; Mattern et al., 2010) and supply fluctuations occur (Mattern et al., 2010). The issue of price fluctuation is reflected by the factor of energy economics (V9), which includes high-loading terms like price, loss, and trading.

Renewable energy has changed the design of the electrical power systems. Traditionally, large generation units were connected to the transmission networks that supplied the power to passive distribution networks, which delivered the power to customers whose demand tended to be passive and uncontrollable (Bayod-Rújula, 2009). In newer electrical power systems, energy is produced by smaller units (e.g., renewable energy sources) that are often located geographically close to the end consumer (Bayod-Rújula, 2009). Consequently, the focus has shifted to the distribution networks. However, information technology is needed in order to create connectivity among the many actors in such distribution networks (Bayod-Rújula, 2009). Furthermore, renewable energy sources require advanced control of the power system, which smart grids can provide (Potter et al., 2009). However, adequate information systems are needed to manage the power system (Mattern et al., 2010). Our analysis of the sustainability reports also shows that the organizations discuss the topic of transmission networks and smart grids, as they are reflected in the energy network factor (V5). Therefore, in order to cope with the challenges of renewable energy generation, supply and demand have to be balanced with consideration of the energy price. Consequently we define the first affordance of information systems for environmental sustainability as:

Affordance 1: Balancing of supply and demand with consideration of the energy price.

The energy informatics framework (Watson et al., 2010) describes such an information system, but our results show that the framework should be extended to include the various ways to generate energy. Because of the weather-dependent fluctuations in the energy supplied by renewable energies, the information systems should address fluctuations in demand and supply. Therefore, information systems must be able to predict the supply based on the expected weather conditions. The framework should also be extended to include the financial part, as renewable energy sources' supply volatility leads to price fluctuations and uncertainty (Feuerriegel et al., 2012). Therefore, information systems should enable demand-side management that shifts electricity use based on the optimal price for electricity (Feuerriegel et al., 2012), such as smart metering technology, which can adjust demand on the individual level. Our fifty-factor analysis reveals that organizations are concerned with smart technology, which includes smart grids, technology, and ICT, an area in which IS can play an important role. As sensitized objects, smart meters are already part of the energy informatics framework (Watson et al., 2010).

Our analysis also shows topics that are not covered by the energy informatics framework, as they are related to a more internal perspective of the organization. This internal view can be separated into two parts, a data-driven part and an activity-driven part. In addition to data (V6), awarding (V3) and emissions and climate impact (V4) also focus on data. Data related to these factors should be gathered, preferably in an automatic way, in order to work on sustainability issues (Kerschbaum et al., 2011), and the relevant information should be shared with appropriate stakeholders. Consequently, we define the second affordance of information systems for environmental sustainability as:

Affordance 2: Capturing, reporting, and sharing of energy-related data.

In order to support organizations in these activities, information systems are needed that can facilitate the capture and reporting of energy-related data, including whether data that should be shared between organizations. For instance, Kerschbaum et al. (2011) develop a solution that provides confidential data-sharing for sustainability benchmarking that could be used for awarding as well.

Emissions and climate impact (V4) is not only related to the reporting of emission data but also to the activity-driven part of reducing emissions. The fifty-factor analysis also shows an interest in energy-saving activities. Energy-saving is related to emission reduction, as less energy consumption reduces the corresponding emissions, but companies are also interested in finding ways to reduce emissions beyond energy savings, such as through the use of renewable energies. Consequently, we need information systems that enable energy savings and emission reduction, so we define the third affordance of information systems for environmental sustainability as:

Affordance 3: Reducing energy consumption and emissions.

Several studies have focused on information systems that help to reduce energy consumption and emissions, such as a tool to improve the energy consumption of private households. Loock et al. (2012) examine the influence of a reference frame and population density on the effectiveness of social normative feedback on electricity consumption, while other authors examine the effects of user identity and sanctions in online communities on real-word behavior (Baeriswyl et al., 2011), the role of goal-setting (Loock et al., 2013), and the potential of virtual communication to replace travel (Seidel et al., 2013; Ijab et al., 2012).

Another action-related factor is waste management (V7). We interpret this factor as energy-related because waste can be a source of energy. However, the main focus of waste management remains the reduction of waste, and information systems should enable sustainable waste management. Possible solutions are information systems that enable separation of waste through sensor-based sorting (Erdmann and Hilty, 2010) and information systems that provide detailed information for electronic devices that support the recycling process (Thoroe et al., 2011). Therefore, we define the fourth affordance of information systems for environmental sustainability as:

Affordance 4: Sustainable waste management.

Finally, we address the application of one of the factors from the fifty-factor model: vehicles and their various forms of power (e.g., fuel, hybrid, diesel, electric). The switch from vehicles powered by conventional or diesel fuel to electric vehicles may reduce the emissions from traveling and commuting. The increased use of electric vehicles may also be a way to balance supply and demand because electric vehicles can act as storage devices (Wagner et al., 2013; Mattern et al., 2010). In order to perform this function, electric vehicles must be connected to the power grid and their charging must be aggregated and coordinated (Wagner et al., 2013). The charging behavior of electric vehicles needs to be adjusted to the supply and demand situations and their users’ requirements (Wagner et al., 2013). Consequently, we define the fifth affordance of information systems for environmental sustainability as:

Affordance 5: Coordination of the flexible charging of electric vehicles.

As in affordance 1, the energy informatics framework (Watson et al., 2010) can serve as an implementation framework (Wagner et al., 2013). However, the required information systems remain to be defined in detail.

These five affordances support energy management through information system in practice and set a foundation for further research in the important field of energy informatics. Table 2 provides an overview of these affordances and their related factors.

No.	Affordance	Related factors
1	Balancing of supply and demand with consideration of the energy price	Renewable energy generation (V2), non-renewable energy generation (V10), energy economics (V9), energy network (V5)
2	Capturing, reporting, and sharing of energy-related data	Awarding (V3), emission and climate impact (V4), data (V6)
3	Reducing energy consumption and emissions	Emission and climate impact (V4)
4	Sustainable waste management	Waste management (V7)
5	Coordination of flexible charging of electric vehicles	Vehicles (fifty-factor analysis)

Table 2 Identified affordances and related factors

6 Conclusion

6.1 Implications

Environmental sustainability and energy consumption in particular are important topics for today's companies. Our research examines the energy-related areas on which organizations worldwide are reporting in their sustainability reports. Adopting the LSA technique for this purpose, we identify five possible affordances of information systems for environmental sustainability. Through enabling these affordances, information system can support organizations in their environmental sustainability efforts.

The results of our study showcase the topics on which organizations concentrate when reporting on energy and the areas in which information systems can play an important role in supporting organizations' energy-related activities. The five information systems affordances we propose and motivate are (1) balancing of supply and demand with consideration of the energy price; (2) capturing, reporting, and sharing of energy-related data; (3) reducing energy consumption and emissions; (4) sustainable waste management; and (5) coordination of flexible charging of electric vehicles.

Our research has several implications for practice and theory. To our best knowledge, this is the first study to provide such an overview of energy-related areas that are relevant for organizations based on organizations' sustainability reports. Organizations can use the results shown in table 1 to identify areas in which to engage, and based on the energy-related topics we identified, they can consider the use of information systems to cope with the challenges they face regarding energy-related sustainability.

Our results are also relevant for research. Based on the results of our analysis, we propose an extension of the energy informatics framework (Watson et al., 2010) to include various ways to generate energy in consideration of the energy price. In addition, we identify potential affordances of information systems in regard to energy-related sustainability in organizations. These affordances may guide further research in this field, provide grounds for discussion, and motivate further developments in the field of energy informatics. Researchers may also choose among the most prominent topics of energy-related environmental sustainability revealed in this study to study them in depth.

Finally, our study is one of the first studies in IS that employs LSA. In particular, to our best knowledge, it is the first study to use LSA to explore sustainability reports or, more generally, annual reports published by organizations. We hope that our study will motivate other researchers to employ this method as well.

6.2 Limitations and Future Work

Our research has a few limitations. First, our data source, the collection of publicly available sustainability reports, are a means for organizations to publish their sustainability activities (Bradford et al., 2012), so they are also marketing instruments and might not present the complete reality. As a result, the vocabulary used in our study is inherited from the organizational rhetoric, which may not reflect the real-life situation well (Astley and Zammuto, 1992). Despite these limitations, we decided to use these sustainability reports because of the significant number of sustainability reports that are available from various industries worldwide and because they report on activities that may cast light on the issues that companies are already dealing with. This study may serve as an initial step in the extraction of factors in the context of energy informatics in organizations, so future research could verify and expand these factors through the use of other data sources.

A second limitation of this study is that the evaluation of the quality and interpretation of the LSA results is influenced by the researchers' bias. The results depend on the choice of the dimensionality for the latent semantic space and the go-list of terms used in the analysis, which is determined by the researchers. We worked to overcome this issue by having the LSA results evaluated independently by two researchers. Despite this limitation, we contend that LSA—and topic modelling in general—is a valid and efficient approach to exploratory analysis in the presence of large volumes of data, and it can provide valuable insights and summarize unstructured textual data. Further analyses and application of other algorithms may provide different results. For example, in future analyses the seeded LDA approach (Jagarlamudi et al., 2012) may be applied instead of LSA, as this algorithm directs the focus of the analysis in a more natural way than a go-list restriction does through incorporation of the "seed words" a user provides (Jagarlamudi et al., 2012).

Third, the identification of required affordances based on the identified topics is not based on empirical evidence but on issues reported in the literature on the identified topics. Thus, only the use context—an organization interested in energy-related environmental sustainability—is defined and not the IT. This is a new approach to employing the concept of affordances. Future research might use our results and evaluate empirically whether these affordances occur and whether they have the intended effects. Future research may also use the identified affordances to design the corresponding information systems.

References

Astley, W., Zammuto, R. (1992) Organization science, managers, and language games. Organization Science. 3(4), 443–460.

Baeriswyl, M.C., Staake, T.R., Loock, C.-M.L. (2011) The effects of user identity and sanctions in online communities on real-world behavior. In Proceedings of the Thirty Second International Conference on Information Systems, Shanghai, China.

Bayod-Rújula, A.A. (2009) Future development of the electricity systems with distributed generation. Energy. 34, 377–383.

Berry, M.W., Dumais, S.T., O'Brien, G.W. (1995) Using Linear Algebra for Intelligent Information Retrieval. SIAM Review. 37(4), 573–595.

Bradford, M., Earp, J.B., Williams, P. (2012) Sustainability: An Analysis of Organizational Reporting and Implications for ERP systems. In Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA. pp. 30–35.

Brandt, T., Bodenstein, C. (2012) Evaluating Scheduling Methods for Energy Cost Reduction in an Heterogeneous data center environment. In Proceedings of the Twentieth European Conference on Information Systems.

Califf, C., Lin, X., Sarker, S. (2012) Understanding energy informatics: A Gestalt-fit perspective. In Proceedings of the Eighteenth Americas Conference on Information Systems, Seattle, Washington, USA.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R. (1990) Indexing by latent semantic analysis. Journal of the American Society for Information Science. 41(6), 391–407.

DeSanctis, G., Poole, M. (1994) Capturing the complexity in advanced technology use: Adaptive structuration theory. Organization Science. (2), 121–148.

Dumais, S.T., Furnas, G.W., Landauer, T.K., Deerwester, S., Harshman, R. (1988) Using latent semantic analysis to improve access to textual information. In Proceedings of the SIGCHI conference on Human factors in computing systems - CHI' 88. New York, New York, USA: ACM Press, pp. 281–285.

Elliot, S. (2011) Transdisciplinary perspectives on environmental sustainability: a resource base and framework for IT-enabled business transformation. MIS Quarterly. 35(1), 197–236.

Erdmann, L., Hilty, L.M. (2010) Scenario Analysis - Exploring the Macroeconomic Impacts of Information and Communication Technologies on Greenhouse Gas Emissions. Journal of Industrial Ecology. 14(5), 826–843.

Evangelopoulos, N., Visinescu, L. (2012) Text-mining the voice of the people. Communications of the ACM. 55(2), 62–69.

Evangelopoulos, N., Zhang, X., Prybutok, V.R. (2010) Latent Semantic Analysis: five methodological recommendations. European Journal of Information Systems. 21(1), 70–86.

Feuerriegel, S., Strüker, J., Neumann, D. (2012) Reducing Price Uncertainty through Demand Side Management. In Proceedings of the Thirty Third International Conference on Information Systems, Orlando, USA. pp. 1–20.

Gibson, J.J. (1986) The Theory of Affordances. In The Ecological Approach to Visual Perception. pp. 127–143.

Goodland, R. (1995) The concept of environmental sustainability. Annual Review of Ecology and Systematics. 26, 1–24.

Graml, T., Loock, C., Baeriswyl, M., Staake, T. (2011) Improving residential energy consumption at large using persuasive systems. In Proceedings of the Nineteenth European Conference on Information Systems. pp. 1–15.

GRI (2013a) GRI: Application Level Check. [online]. Available from: https://www.globalreporting.org/reporting/report-services/application-levels/Pages/default.aspx [Accessed December 1, 2013].

GRI (2013b) GRI: What is GRI. [online]. Available from: https://www.globalreporting.org/information/about-gri/what-is-GRI/Pages/default.aspx [Accessed December 1, 2013].

Herbst, A., Simons, A., vom Brocke, J., Müller, O., Debortoli, S., Vakulenko, S. (2014) Identifying and Characterizing Current Topics in Enterprise Content Management: A Latent Semantic Analysis of Vendor Case Studies. In European Conference on Information Systems. Tel Aviv.

Hoffert, M.I., Caldeira, K., Benford, G., Criswell, D.R., Green, C., Herzog, H., Jain, A.K., Kheshgi, H.S., Lackner, K.S., Lewis, J.S., Lightfoot, H.D., Manheimer, W., Mankins, J.C., Mauel, M.E., Perkins, L.J., Schlesinger, M.E., Volk, T., Wigley, T.M.L. (2002) Advanced technology paths to global climate stability: energy for a greenhouse planet. Science (New York, N.Y.). 298(5595), 981–987.

Hotho, A., Andreas, N., Paaß, G., Augustin, S. (2005) A Brief Survey of Text Mining. LDV Forum - GLDV Journal for Computational Linguistics and Language Technology. 20(1), 1–37.

Hu, X., Cai, Z., Franceschetti, D., Penumatsa, P., Graesser, A. (2003) LSA: The first dimension and dimensional weighting. In Proceedings of the 25th Meeting of the Cognitive Science Society. pp. 1 – 6.

Ijab, M.T., Molla, A., Cooper, V. (2012) Green Information Systems (Green IS) Practice in Organisation: Tracing its Emergence and Recurrent Use. In Proceedings of the Eighteenth Americas Conference on Information Systems, Seattle, Washington, USA. pp. 1–12.

Jagarlamudi, J., Daume, H., Udupa, R. (2012) Incorporating Lexical Priors into Topic Models. Proceedings of the Thirteenth Conference of the European Chapter of the Association for Computational Linguistics, 204–213.

Jivani, A.G. (2011) A Comparative Study of Stemming Algorithms. International Journal of Computer Technology and Applications. 2(6), 1930–1938.

Kaiser, H.F. (1958) The varimax criterion for analytic rotation in factor analysis. Psychometrika. 23(3), 187–200.

Kerschbaum, F., Strüker, J., Koslowski, T. (2011) Confidential Information-Sharing for Automated Sustainability Benchmarks. In Proceedings of the Thirty Second International Conference on Information Systems, Shanghai, China. pp. 1–17.

Kossahl, J., Busse, S., Kolbe, L.M. (2012) The Evolvement of Energy Informatics in the Information Systems Community - A Literature Analysis and Research Agenda. In Proceedings of the Twentieth European Conference on Information Systems.

Landauer, T.K., Foltz, P.W., Laham, D. (1998) An introduction to latent semantic analysis. Discourse Processes. 25(2-3), 259–284.

Landauer, T.K., Laham, D., Derr, M. (2004) From paragraph to graph: latent semantic analysis for information visualization. Proceedings of the National Academy of Sciences of the United States of America. 101 Suppl, 5214–5219.

Landis, J.R., Koch, G.G. (1977) The measurement of observer agreement for categorical data. Biometrics. 33(1), 159–74.

Larsen, K.R., Monarchi, D.E., Hovorka, D.S., Bailey, C.N. (2008) Analyzing unstructured text data: Using latent categorization to identify intellectual communities in information systems. Decision Support Systems. 45(4), 884–896.

Leonardi, P. (2011) When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. MIS Quarterly. 35(1), 147–167.

Loock, C.-M., Staake, T., Landwehr, J. (2011) Green IS Design and Energy Conservation: An Empirical Investigation of Social Normative Feedback. In Proceedings of the Thirty Second International Conference on Information Systems, Shanghai, China. pp. 1–15.

Loock, C.-M., Staake, T., Thiesse, F. (2013) Motivating Energy-Efficient Behaviour with Green IS: An Investigationof Goal Seeting and the Role of Defaults. MIS Quarterly. 37(4), 1313–1332.

Markus, M., Silver, M. (2008) A foundation for the study of IT effects: A new look at DeSanctis and Poole’s concepts of structural features and spirit. Journal of the Association for Information Systems. 9(10), 609–632.

Mattern, F., Staake, T., Weiss, M. (2010) ICT for green: how computers can help us to conserve energy. In Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking. pp. 1–10.

Melville, N. (2010) Information systems innovation for environmental sustainability. MIS Quarterly. 34(1), 1–21.

Molla, A., Cooper, V.A., Pittayachawan, S. (2009) IT and Eco-sustainability: Developing and Validating a Green IT Readiness Model. In Proceedings of the Thirtieth International Conference on Information Systems, Phoenix, USA. pp. 1–17.

Müller, O., Schmiedel, T., Gorbacheva, E., vom Brocke, J. Toward a Typology of Business Process Management Professionals: Identifying Patterns of Competences through Latent Semantic Analysis. Enterprise Information Systems.

Nakov, P. (2000) Latent semantic analysis of textual data. Proceedings of CompSysTech, 3–5.

Niesten, E. (2010) Network investments and the integration of distributed generation: Regulatory recommendations for the Dutch electricity industry. Energy Policy. 38(8), 4355–4362.

Norman, D.A. (1999) Affordance, conventions, and design. interactions. may + june, 38–42.

Porter, M.F. (1980) An algorithm for suffix stripping. Program: electronic library and information systems. 14(3), 130–137.

Potter, C.W., Archambault, A., Westrick, K. (2009) Building a smarter smart grid through better renewable energy information. In IEEE Power Systems Conference and Exposition. Ieee, pp. 1–5.

Salton, G., Wong, A., Yang, C.S. (1975) A vector space model for automatic indexing. Communications of the ACM. 18(11), 613–620.
Seidel, S., Recker, J., vom Brocke, J. (2013) Sensemaking and Sustainable Practicing: Functional Affordances of Information Systems in Green Transformations. MIS Quarterly. 37(4), 1275–1299.
Sidorova, A., Evangelopoulos, N., Valacich, J.S., Ramakrishnan, T. (2008) Uncovering the Intellectual Core of the Information Systems Discipline. MIS Quarterly. 32(3), 467–482.
Strong, D., Volkoff, O., Johnson, S. (2009) A Theory of Clinic-EHR Affordance Actualization. In Proceedings of JAIS Theory Development Workshop. Sprouts: Working Papers on Information Systems. pp. 9–47.
Strüker, J., Kerschbaum, F. (2012) From a Barrier to a Bridge: Data-Privacy in Deregulated Smart Grids. In Proceedings of the Thirty Third International Conference on Information Systems, Orlando, USA. pp. 1–16.
Thoroe, L., Appelhanz, S., Schumann, M. (2011) RFID-based individualization of extended producer responsibility and recycling for WEEE. In Proceedings of the Nineteenth European Conference on Information Systems.
Vykoukal, J. (2010) Grid Technology as Green IT Strategy ? Empirical Results from the Financial Services Industry. In Proceedings of the Eighteenth European Conference on Information Systems. pp. 1–14.
Wagner, S., Brandt, T., Neumann, D. (2013) Beyond Mobility - An Energy Informatics Business Model For Vehicles In The Electric Age. In Proceedings of the Twenty First European Conference on Information Systems.
Watson, R.T., Lawrence, T.M., Boudreau, M.-C., Johnsen, K.J. (2013) Design Of A Demand Response System : Economics And Information Systems Alignment. In Proceedings of the Twenty First European Conference on Information Systems.
Watson, R.T.R., Boudreau, M.C.M., Chen, A.A.J. (2010) Information systems and environmentally sustainable development: energy informatics and new directions for the IS community. MIS Quarterly. 34(1), 23–38.
Wunderlich, P., Veit, D., Sarker, S. (2012) Examination of the Determinant of Smart Meter Adoption: An User Perspective. In Proceedings of the Thirty Third International Conference on Information Systems, Orlando, USA. pp. 1–17.
Yim, D. (2011) Tale of Two Green Communities : Energy Informatics and Social Competition on Energy Conservation Behavior. In Proceedings of the Seventeenth Americas Conference on Information Systems,.