Taking Advantage of Data Generated by Products
Trends, Opportunities and Challenges

van der Vegte, Wilhelm Frederik

DOI
10.1115/DETC2016-59177

Publication date
2016

Document Version
Final published version

Published in
Proceedings of the ASME 2016 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference

Citation (APA)
van der Vegte, W. F. (2016). Taking Advantage of Data Generated by Products: Trends, Opportunities and Challenges. In *Proceedings of the ASME 2016 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference : 36th Computers and Information in Engineering Conference* (Vol. 1B, pp. 1-10). [DETC2016-59177] ASME. https://doi.org/10.1115/DETC2016-59177

Proceedings of the ASME 2016 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2016
August 21-24, 2016, Charlotte, North Carolina, USA

IDETC2016-59177

Taking Advantage Of Data Generated By Products: Trends, Opportunities And Challenges

Wilhelm Frederik van der Vegte
Delft University of Technology
Delft, The Netherlands

ABSTRACT

Now that all kinds of products are increasingly getting connected to the Internet, it is expected that it will become easier to collect data on how they are actually used during the middle-of-life stage of their product lifecycles. At the same time, a growing number of data analytics technologies offers opportunities to transform this data into actionable knowledge. Over the years, such knowledge extracted from usage data has already become a reliable input for managing maintenance and related services, but other uses such as feedback to design – where product data management systems have started to offer support for data collection practices – and providing advice to end users are now also being considered. Most data from sensors and other product-embedded information devices are collected in batches and analyzed retrospectively. In order for companies to further benefit from data collection in terms of efficacy and acceptance in society, two key challenges are (i) finding ways to effectively use data analytics techniques – which currently do not seem to be used to their full potential, and (ii) finding a good trade-off between respecting privacy and yet producing useful knowledge.

INTRODUCTION

Analysis of how different customers use products can provide valuable insights for companies which depend on revenues generated directly or indirectly from those products. Among these companies are manufacturers, resellers and third parties such as maintenance providers and insurance companies. In addition, non-commercial parties such as law enforcement authorities and NGOs may have interest in how particular products are being used.

The collection of usage data to obtain particular insights has been common for a long time in the exploitation of websites and software, as well as hardware such as computers, smartphones and digital cameras [e.g., 1,2,3]. Now that the Internet is evolving "from a network of interconnected computers to a network of interconnected objects" [4], also referred to as the Internet of Things (IoT), more and more categories of products offer opportunities for collecting data about how they are being used. This trend is extending to product categories that are deployed to achieve mostly physical effects, which did not traditionally produce any processable data. It is facilitated by the fact that product functionality is increasingly realized with the help of information-producing and networked solution elements such as embedded software, sensors – which convert measurements from the physical world to data – and actuators, which convert data to changes in the physical world. Companies can track the movements of these products and monitor interactions with them, which inspires new business models taking advantage of these behavioral data [5]. As we will see in this survey, the changes in business models can take various forms. Knowing how customers actually use the products is said to enhance a company's ability to segment customers, customize products, set prices to better capture value, and extend them with value-added services [6].

OBJECTIVE, SCOPE AND METHOD

In this paper we have focused on how product-producing companies can extract and exploit knowledge from product usage to improve their products and product-related services. It is *not* a paper about the Internet of Things, as it has been defined as "the networked interconnection of everyday objects, which are often equipped with ubiquitous intelligence" [7] or "the pervasive presence around us of a variety of things or objects which (...) are able to interact with each other and cooperate with their neighbors to reach common goals" [8]. The IoT according to these definitions adds value through networked cooperation between products of various kinds, in applications such as domotics control in smart homes [9]. Instead, our focus has been on data and knowledge collected from multiple instances of physical, tangible products of the same kind. The IoT can be seen as a possible enabler of data collection, but we have also considered more traditional, centralized communication schemes in which each instance of the product unidirectionally sends data to the company. Our focus on data to be used by the product-producing company also implies that the survey does not cover collection of data only to be used within the product itself and its local context, as is the case with applications of control engineering and local diagnostics support – e.g., distributed control of aircraft engines [10], and user access to a vehicle's diagnostic information [11]. Furthermore, our focus on how the product in question *is being* used, implies that we have not considered data collected *by* the product about (behavior of) other entities that it is monitoring, like fitness trackers, smart alarm systems, smart energy meters, etc., do [12].

This survey aims to (i) accumulate what has been done in the field of gathering and utilizing data from products by giving an overview of the applied technologies and approaches as well as the achieved results, and (ii) identify unresolved (research) challenges and unexploited opportunities.

In reviewing the current state of the art, we have started from the following questions:

- What kinds of products have been considered in the reviewed sources?
- What have been the motivations to collect and analyze the data?
- What are the technologies, infrastructures and platforms that have been considered to generate, transmit, collect and interpret the data?

In the light of the increasing connectedness of products, we were especially interested in data collection by products that traditionally would not be expected to generate any data about their usage, i.e. products that are *not* essentially computers, but products with main functions other than receiving/collecting, processing and providing information. For those products that *are* essentially computers, the so-called information-centric products – such as smartphones and tablets – the focus has been on data other than the data handled by the product's information-processing processes. For instance, for a laptop we would be interested in how users handle it mechanically rather than in how often the laptop connects to wireless networks. This special interest was motivated by the fact that for information-handling functions the collectable data about usage comes more or less for granted, while it might be more interesting to learn about the additional efforts needed if that is not the case.

As a basis of the survey, scientific publications, commercial materials as well as technology news reports collected from the Web were used. The initial, central search has been for publications and websites where the words *collect, product, usage/use,* and *data*, with or without *sensors* appear together. Subsequently, other sources to which the results referred were consulted, and potentially meaningful terminology often mentioned in the results was also used as search terms.

The structure of the remainder of this paper is as follows: in the next section, the various products that have been provided with usage data gathering capabilities are characterized. Secondly, the various motivations for collecting the data are discussed: what did companies and other stakeholders aim to achieve with it? Third, the technologies (platforms, hardware, networking, analysis techniques, etc.) are discussed. These next three sections aim to reflect the trends in collecting and utilizing data from products. After that, issues and challenges are identified, and the paper ends with a discussion and conclusions.

CHARACTERIZATION OF PRODUCTS THAT HAVE BEEN DEPLOYED TO COLLECT DATA ABOUT THEIR USE

Many reports on products collecting data during middle-of-life (MOL, also known as mid-life or the aftermarket, i.e. the stage of the lifecycle where a product is used) focus on a particular category of products or even a specific product. From our investigation, it appears that the practice of equipping products with data collection capabilities and utilizing the data is currently more widespread among products of a certain (high) complexity. These are products that buyers generally consider investment goods. Quite a number of sources [13-24][1] report on data collection from automobiles. Other sources predominantly report on B2B applications in aircraft [5,18,25], military equipment [26], industrial equipment [27-31], and infrastructure such as bridges [32,33], street lights [34] or elevators [35]. The fact that data collection has become so widespread in cars may be due to the fact that automobiles are not only investment goods, but contrary to the other products mentioned above, also mass products.

One example where MOL data collection has made it to less capital-intensive products is Hewlett Packard's Instant Ink program for inkjet printers [36]. The few other examples of data collection from less capital-intensive goods concern studies, where researchers have collected data from one product or a small number of products to investigate aspects of possible future data collection and utilization at a larger scale – for instance fridge-freezer combinations [37], notebook computers [38], and furniture [39]. Furthermore, in 2015, Miele completed a proof-of-concept study with data collection from connected kitchen equipment [40].

In the remainder of the paper, application examples will be discussed in this same order: (i) automotive, (ii) aerospace and defense, (iii) industrial equipment and infrastructure, and (iv) non-capital-intensive products including consumer goods. For each category, first, examples will be discussed where data collection actually has become practice, followed by smaller-scale data collection experiments and novel approaches proposed by researchers.

[1] The individual contributions from the referred works will be discussed in the next subsections.

MOTIVATIONS FOR COLLECTING AND PROCESSING DATA

A company that decides to collect data from its products – or from other companies' products – does so with a particular intention. This intention, motivation or rationale is the driver behind some form of exploitation of the processed data. It can be anything from optimizing business processes to achieving the "changes in business models" mentioned in the introduction. By far the largest body of literature concerns managing maintenance of products out in the field. After the state of art in that area, other reasons why data collection has been considered or implemented will be discussed.

Maintenance management of fielded products

According to the broad definition offered by the European Federation of National Maintenance Societies, 'maintenance', is *the combination of all technical, administrative and managerial actions during the lifecycle of an item intended to retain or restore it to a state in which it can perform its required function* [41]. Going by this definition, there are several differently named but similar approaches aiming to exploit data collection for support, streamlining, or optimizing maintenance of products to reduce downtime, avoid unnecessary maintenance activities, increase customer satisfaction and extending the use phase of the product lifecycle. In this section, the following approaches have been grouped together: *condition-based, predictive, proactive and preventive maintenance, prognostics & health management* (PHM) and *through-life engineering services* (TES).

Condition-based maintenance is an established and accepted maintenance practice. It aims to derive maintenance requirements from real-time assessment of the product[2] condition obtained from embedded sensors and/or external tests and measurements. It relies on built-in diagnostic equipment or portable diagnostic equipment, such as PDAs and tablets [26]. The goal of condition-based maintenance is to perform maintenance based only upon the evidence of a need rather than any predetermined time cycle, equipment activity count, or other engineered basis.

Proactive maintenance is an approach that uses integrated, investigative and corrective practices to significantly extend machinery life with the goal to eliminate failures of equipment forever [31].

PHM aims to monitor life-cycle environmental and usage conditions of products or systems to assess on-going health, provide advance warning of failure through detection of failure precursors, and provide information to improve the design and qualification of fielded and future products [42].

TES has been defined as *"a result of the application of explicit and tacit 'service knowledge' supported by the use of monitoring, diagnostic, prognostic technologies and decision support systems whilst the product is in use, and maintenance (…) functions to mitigate degradation, restore 'as design' functionality, maximize product availability, thus reducing whole-life operation cost"* [43]. This is achieved based on five sources of knowledge, namely knowledge of (i) degradation and failure mechanisms, (ii) means of repair, (iii) diagnostics and prognostics, (iv) use, and (v) design and function.

The approaches to manage maintenance described above are often considered to underlie so-called *performance models*, which represent the transition from selling products to selling performance. They are based on the rationale that there is no inherent benefit for the customer to actually own the product [44].

In the automotive industry, health monitoring and fault tracing based on diagnostics data collected from fielded cars forms an important part of service and maintenance [16]. It increases the service technicians' ability to diagnose and remedy problems in the increasingly complex electronically controlled vehicles and thus improves customer satisfaction. The offline retrospective readouts are also uploaded to the manufacturer's database to analyze fault occurrences collected from multiple cars, (i) to monitor the quality of components and subsystems, (ii) to prioritize in which order problems should be addressed and (iii) to find correlations between different faults, or between faults and the operating environment. The recent trend of offering *real-time* connectivity in vehicles is mainly motivated by customers' demand for on-board internet and on-demand entertainment [45] rather than by the need to collect data.

The automobile industry has introduced data collection platforms offering support for maintenance management. For instance, GM's OnStar emails diagnostics reports to the dealer to facilitate scheduling of service appointments [46]. A more futuristic proposition was proposed by Amor-Segan et al. [23]: their self-healing vehicle concept collects data from connected automobiles and is supposed to support in-vehicle autonomous fault management. They claim that centralized collection of data based on wireless telematics can be used to (i) facilitate more comprehensive data analysis and diagnosis at a remote support center, (ii) receiving diagnostics patches to aid in-vehicle diagnostics, (iii) update diagnostic and prognostic guidance and (iv) enable new software versions for feature enhancements, correction of design and implementation errors. In addition, Johanson et al [16] foresee support of inspection and repair during manufacturing of automobiles based on collected data.

Performance models have been introduced in the aerospace industry, where manufacturers of jet engines nowadays retain ownership of their products while charging airlines for the amount of thrust used [5,43]. As an example of predictive maintenance in industrial capital goods, Marek et al describe how this has been put into practice for mining equipment [28]. Maintenance dates are scheduled and optimized related to the actual load on the machines. Before a maintenance job, the machine informs the crew about the tools and consumables needed, thus reducing the level of service skills required. As tools and consumables can be pre-organized and made available,

[2]The original publication [26] specifically uses "weapon system" where "product" is used in this survey.

hourly-based routine maintenance can be avoided and the time involved minimized, thus reducing downtime and improve availability.

Aspects of predictive maintenance can also be found in less capital-intensive products – as Hewlett Packard's Instant Ink program for inkjet printers shows. It enables connected inkjet printers to arrange replacement cartridges for their end users before they run out [36]. Service contracts for office equipment such as printers and computers are often based on a performance model [44].

Other uses

Other than for maintenance management, one of the uses of collected data that has been foreseen by the literature is providing feedback to product design. This feedback is used, for instance, to reduce future product failures and associated services required [6] to draft better requirements based on actual usage or to redefine the functionality of a next product design iteration based on functions and features actually used [44]. Similarly, Van Horn et al [47] have suggested that data collected from deployed products enables manufacturers to quickly identify and efficiently solve quality issues in specific components. In addition, product usage data can also be used to validate warranty claims and identify warranty agreement violations [6]. Furthermore, Främling et al. [48] have suggested to collect information from connected cars to (i) proactively optimize engine tuning based on factors such as location and time of day, and (ii) present comparative performance measures affecting behavior of drivers. Al-Taee et al. have suggested collecting data from connected cars for a completely different purpose, namely, to allow the traffic control authority to record speed limit violations [21].

Data collection schemes that have been brought to practice or have been envisaged for concrete products give evidence of some of the above and several other motivations behind data collection. The initial goal that Ford envisaged with collecting usage information from customers' automobiles in the 1990s was indeed to gain understanding of how customers actually use their vehicles and to define appropriate specifications for development and testing. This has been considered as a critical factor supporting design and development in delivering affordable, high-reliability, high-quality products [20]. Hilpert et al [22] presented a system for real-time collection of CO_2 emissions from an entire company fleet of transportation vehicles to assess the carbon footprint of the products they are transporting. Although, strictly spoken, the application is outside the scope of this survey, it could theoretically be used to collect emission data from all fielded cars of a certain type, and collect usage data based on which its manufacturer could possibly reduce emissions.

Two forms of third-party use of data collected from automobiles have been reported by Chui et al [5], namely (i) insurance companies installing location sensors in customers' cars so that they can base the price of policies on how a car is driven as well as where it travels, and (ii) rental car companies using tracking data to optimize each car's use.

In the aerospace industry, important objectives – other than maintenance management – for data collection have been (i) improving crew decision-making and response in complex situations (ii) maintaining aircraft safety between major inspections; and (iii) assuring safe and effective aircraft control under hazardous conditions [18] .

Dienst et al. [29] propose a knowledge-based feedback system to assist product developers in exploiting data collected during the use of industrial goods, e.g., centrifugal pumps. From the given examples, the impression emerges that application of the system is limited to redesign based on component selection and parameter modification, e.g., selecting a better bearing to replace a bearing that the data analysis proves to fail too often, or selecting a different material.

Coca-Cola collected data from vending machines that allowed customers to compose their own drinks, with the objective to automatically schedule refills, but also for marketing purposes: the purchased mixtures provided indications of how new drinks are performing on the market over time, and of differences in regional tastes [49]. Miele's connected kitchen equipment has been developed with the initial goal to assist end users by providing recipes on demand, but future plans include data collection for generating status report for machines or enabling remote diagnosis of problems [40].

For the EU-funded ELIMA project, data from 28 fridge-freezer combinations was collected to record events of door opening and using the fast-freeze feature per user over time, with the goal to obtain an impression how useful these data would be as input for (i) design improvements, (ii) offering improved logistics and (value-added) services and (iii) possibility of reusing components from disposed products [37]. The preliminary findings indicated that some potentially useful input could be collected for design and also for the contents of the user manual.

Gu et al. [38] collected data about handling of notebook computers to (i) get an impression of variations in use conditions between different users and in one user over a longer time span, and (ii) verify that the test conditions in lab tests reasonably reflect actual use. Some of their tests involved hundreds of users over hundreds of days. They were able to point out particular use conditions that were either more critical than assumed or were not properly reflected in lab tests.

TECHNOLOGIES, INFRASTRUCTURES AND PLATFORMS TO SUPPORT UTILIZATION OF DATA FROM PRODUCTS

All the surveyed approaches to taking advantage of data collected by products assume a *processing chain* that starts with collecting or generating the data and ends with outputting the results of data processing for utilization and storing it for possible later use. Our goal in surveying technologies, infrastructures and platforms has been to get a general overview of:

- how processing chains have been implemented:
  - To what extent are data stored and processed in the product?
  - Is it a continuous stream of data or is it a list of events?

  – Is the data transferred by wire or wirelessly?
  – Is this done continuously in real time or in batches?
- what kind of analysis is performed:
  – How has the need for collecting data affected the product, i.e., to what extent does it require additional PEIDs (product-embedded information devices, i.e., sensors, transmitters and processors)?

In the automotive industry, 'on-board diagnostics system' or ODB is the common umbrella term used for systems collecting MOL data [13]. The ODB in automobiles physically manifests in the form of the ODB-II connector which is connected to the Controller Area Network (CAN) bus. The CAN bus is in turn responsible for the communication between the electronic control units (ECUs) of the car [15]. Diagnostic trouble codes (DTCs) from ECUs are routinely being read out during service from customer vehicles using a wired connection. DTCs are stored only when a reading is out of range. Readings produced at other times are generally not recorded. This is a missed opportunity, because these could potentially be used to gain knowledge about usage and vehicle behavior, for instance to predict faults. By establishing a real-time connection to the ODB these off-line retrospective readouts can however be collected and sent to a manufacturer's database for further statistical analysis to find correlations between detected events [16].

Connecting cars to the Internet can be achieved indirectly through a smartphone [17,22], although today's connected cars usually have their own direct access to the mobile phone network [19]. The increasing demand for bandwidth requires implementation of multiple radio interfaces, which may incur a high cost and thereby impede further developments [45]. In the 1990s, Ford introduced CVDAS (customer vehicle data acquisition system), the first platform to wirelessly connect vehicles [20]. CVDAS uses the same SAE J1850 protocol for the vehicle data communication backbone that was prescribed for OBD. Its wireless data communication is based on mobile telephony standards. A recent development in that area is the ISO 13400 standard for Diagnostics over Internet Protocol (DoIP) [16]. To be able to collect the desired usage data in CVDAS-equipped cars, the existing ECUs have been extended with additional sensors such as an ambient temperature sensor and a rotary position sensor. To keep hardware requirements manageable, the data are statistically analyzed inside the car and only the results are transmitted. A drawback of this approach is that, in order to perform the right type of analysis, a-priori knowledge about system interaction effects is needed [20]. In the system for real-time collection of CO_2 emissions presented by Hilpert et al [22], OBD data were combined with GPS, transmitted wirelessly through the mobile phone network, and collected and processed by ERP systems.

The term OBD is also used in the aerospace industry [18], where the proactive maintenance schemes that have been introduced by manufacturers rely on networked sensors built into airframes that send continuous data on product wear and tear to the manufacturers' computers [5].

In the knowledge-based feedback system that Dienst et al. [29] proposed for industrial equipment, the collected data consists of (i) sensor data, which are collected automatically, and (ii) data that have been entered manually by maintenance engineers and customers[3]. The system prepared the data so that these could be handled by a product lifecycle management (PLM) system. This is needed because, according to the authors, conventional PLM systems cannot deal with multiple individual instances of products, and therefore the systems cannot store the collected data directly. Before further processing, the collected data require additional human intervention: a knowledge engineer aggregates the data from numerous databases and initiates Bayesian-networks based statistical analysis and visualization techniques. With the results, designers can perform what-if studies with different usage conditions, and identify weak spots in the design to be reconsidered. A decision support module guides towards the best solution from available alternatives.

The automated maintenance planning and diagnostic fault-finding for mining equipment that Marek et al. reported on, uses on-board sensors. The machines' on-board control system processes the incoming data and compares these with the machine manufacturers' database to ascertain whether the values are within the defined parameters. If not, maintenance is scheduled automatically through a wired interface with the SAP™ ERP (enterprise resource planning) system. In addition, the ERP integration facilitates automated ordering of the consumables needed for maintenance, and assessment of the machine's performance in the context of the entire mine. The sensor data themselves are stored at the mining site in an SQL database to allow further (unspecified) post-processing and visualization [28].

The fridge-freezer combinations in the ELIMA study reported in [37] were equipped with extra sensors to log energy consumption, door openings, power on/off cycles and temperatures every second. Several other parameters could be read from the embedded software without the need for adding additional sensors. Data logged by a built-in custom logger were transferred to the ELIMA database by a GSM module once every three hours. At the end of the running time of the study, the collected data were visualized in histograms, presumably by using a spreadsheet application.

The notebook computers in the experimental setup discussed in [38] were equipped with sensors capable of measuring temperature, humidity and vibration. Part of the collected data were visualized in graphs without additional prior processing, to qualitatively assess characteristic patterns of variables over time and relations between variables (e.g. between temperature and humidity inside the notebook), other part of the data was statistically analyzed by means of ANOVA tests. Since the investigators performed analysis based on recorded history of sensor data, real-time communication of data does

[3] These data that are not generated by the product and therefore outside the scope of this paper.

not seem to have played a significant role.

Wrapping up the inventory of technology that is used to realize the data-processing chain, we can state that the first step, collection of data, typically takes place inside products, and is typically done by sensors. Embedded software can also produce valuable data, thus reducing the need for additional sensors. The subsequent steps may take place anywhere between ‘inside the product’ and ‘at a central location’. If data processing is done in-product, it is typically transmitted to and stored at a central location afterwards, i.e. the product’s manufacturer or a service provider’s site. Some basic pre-processing of the bandwidth for data transfer, e.g., when the average is considered instead of the individual values. Details about the subsequent processing that is performed to produce actionable knowledge are not always given. Approaches that have been mentioned are statistical techniques such as Bayesian networks analysis, ANOVA and visualization. Some of the reported processes involve multiple steps and in some cases human interventions by, for instance, knowledge engineers. In none of the discussed implementations, continuous data transfer and real-time knowledge conversion seem to play a role.

For storage of the data and the findings, and making these accessible and manageable, PLM systems, ERP systems and databases such as MySQL are used. The name ‘product lifecycle management’ suggests that PLM includes management of MOL and tracking how products are actually being used. However, several authors have indicated that conventional PLM systems are not adequately equipped for that purpose. For a long time, these systems have focused on processes where digital systems such as CAx traditionally produce large amounts of data to be managed [48]. Conventional PLM systems are generally not equipped to keep records of any dynamical process after the product has left the factory [50].

As a follow-up to this conclusion, it is interesting to note that software vendor PTC has recently announced that the latest version of its Windchill™ PLM system was designed to to support collection of PEID data during MOL [51]. This would facilitate exploitation of usage data in predictive maintenance and MOL-information-based design, which has also been referred to as ‘closed-loop PLM’ [52].

ISSUES AND CHALLENGES

The sources that we consulted pointed out several issues and challenges related to collection of MOL data from products. In addition to these sources, we reviewed several issues and challenges that were identified in works related to the IoT [53-57] to check whether these would also apply to data collection within the scope of our survey. The following issues and challenges were brought forward by two or more sources:

- limitations of the current internet [e.g., 45,53];
- privacy, trust and security [e.g., 4,37,44,53,54];
- conversion of data to knowledge [e.g., 53,54,55,56];
- achieving standardization and overcoming heterogeneity [e.g., 4,44,53];
- energy efficiency [e.g., 56,57].

Below these issues will be addressed more specifically in the context of data collection and utilization during MOL.

Limitations of the current internet

The current Internet architecture is limited in terms of mobility, availability, manageability and scalability [53]. This may give rise to problems if data are collected to provoke immediate action on critical events [45] or if the quality and/or completeness is crucial for achieving the objectives of collecting and processing the data.

Privacy, trust and security

The data collected by products during MOL hold information or knowledge about product usage, and thus also about the users. Social acceptance of data collection and utilization is expected to strongly depend on the respect for privacy that is being observed, and the protection of personal data. [4].

The privacy concerns arising from the collection of usage data from tangible products are perhaps best illustrated by what is known from the car industry, which is obviously a prominent domain where data collection has become common practice.

In 2015, researchers from the General German Automobile Club ADAC were commissioned by the International Automobile Federation FIA to investigate data collection by cars with wireless connection capabilities. They examined two cars from one manufacturer – one with combustion engine and one electrical car. The goal was to uncover (i) what data these cars collect and make available to the manufacturer and/or the workshop, (ii) how long these are stored inside the car and (iii) on what occasions the data are transferred [14]. Since data collection and transfer is based on closed-source mechanisms devised by the manufacturer, the investigators had to reverse-engineer ODB information and signals transferred by the built-in wireless communication means. For the same reason, the manufacturer’s motivation behind collection of the data could not be determined.

Of the dozens of information items that they identified to be stored and/or retrieved during workshop visits and/or wirelessly transmitted, several were labeled potentially privacy-sensitive. Among these are preferred seat positions, telephone contacts and numbers of drives covering particular distance ranges. In the electrical car they even found that, each time the ignition is turned off and the car is locked, it transmits data such as GPS location of the parking spot, previous charging stations, recent destinations entered in the navigation system as well as at least 25 other items. Based on these and similar results, the FIA has demanded new legislation to ensure that car manufacturers (i) reveal what they collect, (ii) give customers access to the collected data and (iii) offer an opt-out from data collection.

Findings from [37] suggest that, especially if it is used to improve service or recycling processes, most consumers (~70%) would accept recording of technical data, provided that not too much of actual usage is revealed. Furthermore, they appeared to accept data collection at end-of-life more easily than continuous collection over the Internet. Besides, it has been suggested that the IoT and other recent ICT developments are affecting the way privacy is understood, particularly among younger generations [4]. In that context, future users might be more willing to accept collection of data by products.

There are strategies that can be applied to reduce the privacy sensitivity of transmitted data – for instance, limiting the data transmission or reducing the quality or fidelity of the transmitted data. However, there is a trade-off in applying these: it is considered unavoidable that these approaches compromise the quality of the extracted knowledge and thus the user's trust in it [58,59].

Besides privacy, security of information is considered a major concern when data are collected that can reveal insights on users [4,60]. Industrial espionage can be a threat for business data [61], and hackers can be a threat for both business data and potentially privacy-sensitive data of users [62]. Since this is a whole field of research in itself, it will not be elaborated here; the reader is invited to refer to the many surveys on this specific topic, for instance [63-66].

Converting data to knowledge

The whole point of collecting data is to transform these into actionable knowledge [55]. In the context of this survey, 'actionable' means that knowledge satisfies the motivation behind the data gathering (e.g., service management or design improvement). It is however somewhat disappointing that our sources hardly provide details on how the data were analyzed. In most cases, sources state that 'statistical analysis' or even just 'data analysis' was performed. Only in a few cases, more specifics were given, such as Bayesian Networks [29] and ANOVA [38].

Statistical analysis is just one of the more traditional forms of data analysis, and there is a large collection of other techniques available, including various data mining and discovery techniques, prediction techniques and simulation techniques using real-time acquired data [43,54,67]. Developing methods to select the best out of many analysis techniques given the characteristics of the available data and the motivation that is to be satisfied, still seems to be a challenge. A possible reason why the industry does not seem to explore potentially more advanced analysis techniques is given in the next subsection.

Standardization and homogeneity

Most implementations of data collection and utilization have been developed in closed innovation processes [44], which gives rise to the problem that components (including networks and software) from different companies have to work together, yet cannot be integrated or run on a common operating system [68]. Managing heterogeneous applications, environments and devices constitute a major challenge [53]. Consequently, in practice, the collected data are mainly used for anomaly detection and control, but not for more sophisticated forms of analysis such as optimization, prediction or discovery [67].

Energy efficiency

Collecting, processing and transferring data consumes electrical power. Especially, the power required by 3G and Wi-Fi connectivity is relatively high. Problems may arise when a user is responsible for maintaining the battery and other connectivity aspects of the product [59]. Energy supply is also an issue for products that are traditionally not powered and need to be powered just for data collection, such as furniture [39].

DISCUSSION AND CONCLUSIONS

From the inventory, the impression emerges that, apart from products that are essentially computers, MOL data collection and analysis has mainly been deployed in the context of capital-intensive goods, such as automobiles, airplanes, professional equipment and manufacturing equipment. With the exception of automobiles and a few other products where data collection has been studied in small-scale experimental setups, these products are typically deployed in a B2B context. However, data collection from consumer products seems to be on the rise, as most examples in that area appear to be recent. This trend can perhaps be accounted to the fact that the contact between companies and consumers is more anonymous than between companies and corporate customers in a B2B context. Collecting data about these previously anonymous consumers would offer a good opportunity to get to know them better.

In the majority of the cases, the rationale behind collecting and processing usage data is to manage maintenance activities. Other purposes to which data collection has been exploited most prominently include feedback to design, for future products, and providing advice to end users. Furthermore, analysis results were used for diverse purposes such as marketing, tracking and classifying users and environmental impact assessment.

Currently, most data are collected at intervals and analyzed retrospectively. Real-time monitoring does not seem to be much needed, except for condition-based maintenance. In some cases, products such as smartphones play an intermediary role in collecting the data.

For the manufacturer, the collected data can generally be characterized as a contribution to management of the product lifecycle. Recently, vendors of product lifecycle management software appear to have recognized the potential, and have started offering functionality to collect data from fielded products. Among the parties taking advantage of the data are not only the manufacturers of the products, but also resellers and third parties such as maintenance providers and insurance companies. In addition, non-commercial parties such as law enforcement authorities have shown interest in how particular products are being used. It is not surprising that all this interest in usage data might cause privacy concerns among end users – especially in cases where they do not seem to benefit from it (e.g., validation of warranty claims). Offering the possibility to opt out from data collection seems to be a good solution to this problem. Data security is a related issue; offering solutions for secure data handling is, however, a discipline of its own.

For the actual analysis of the data, a wide range of techniques are available including solutions from machine learning, statistics, pattern recognition, simulation and combinations thereof. However, in most cases of actual data collection, no further analysis tools than basic statistics are being applied. One of the biggest unresolved challenges is to match the characteristics of the available data to those analytics tools that best support the extraction of the sought-after knowledge. A first step towards achieving this would be to conduct further research to characterize and classify (i) all types of data that can possibly collected from products on how they are used, (ii) motivations of stakeholders for collecting the data in terms of possible analysis results (i.e., the sought-after knowledge: answers to questions/queries about the data), and (iii) data analytics techniques, their data requirements and their knowledge extraction capabilities. In addition, the development of knowledge extraction approaches would benefit from standardization among the involved applications, environments and devices.

Another important issue, especially when it comes to societal acceptance of data collection practices, is finding a way to deal with the trade-offs that arise between respecting privacy of individual end users and striving to get the highest-quality knowledge out of the collected data. This, however, is a problem that is also being dealt with in related other fields – in particular analysis of website statistics.

ACKNOWLEDGMENTS

Part of this research has been funded under the EC Horizon 2020 Programme, in the context of the FALCON project (http://www.falcon-h2020.eu/). The author wishes to acknowledge the Commission and all the FALCON project partners for the fruitful collaboration. In particular, he would like to thank Karl A. Hribernik of BIBA - Bremer Institut für Produktion und Logistik GmbH for his valuable suggestions and corrections.

REFERENCES

[1] Atterer, R., Wnuk, M., and Schmidt, A., (2006), "Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction", Proceedings International Conference on World Wide Web, ACM, pp. 203-212.

[2] Kanai, S., Higuchi, T., and Kikuta, Y., (2009), "3D digital prototyping and usability enhancement of information appliances based on UsiXML", International Journal on Interactive Design and Manufacturing (IJIDeM), Vol. 3 (3), pp. 201-222.

[3] Bollen, L., Giemza, A., and Hoppe, H.U., (2008), "Flexible analysis of user actions in heterogeneous distributed learning environments", in: Times of convergence. Technologies across learning contexts, Springer, pp. 62-73.

[4] Commission Of The European Communities, (2009), "Internet of Things - An action plan for Europe", Brussels, 2009.

[5] Chui, M., Löffler, M., and Roberts, R., (2010), "The internet of things", McKinsey Quarterly, Vol. 2 (2010), pp. 1-9.

[6] Porter, M.E., and Heppelmann, J.E., (2014), "How smart, connected products are transforming competition", Harvard Business Review, Vol. 92 (11), pp. 11-64.

[7] Xia, F., Yang, L.T., Wang, L., and Vinel, A., (2012), "Internet of things", International Journal of Communication Systems, Vol. 25 (9), p. 1101.

[8] Atzori, L., Iera, A., and Morabito, G., (2010), "The internet of things: A survey", Computer networks, Vol. 54 (15), pp. 2787-2805.

[9] Jimeno, R., Salvador, Z., Lafuente, A., Larrea, M., and Uribarren, A., (2004), "An architecture for the personalized control of domotic resources", Proceedings Proceedings of the 2nd European Union symposium on Ambient intelligence, ACM, pp. 51-54.

[10] Tulpule, B., Behbahani, A., and Millar, R., (2007), "Vision for next generation modular adaptive generic integrated controls (MAGIC) for military/commercial turbine engines", American Institute of Aeronautics and Astronautics, JPC.

[11] Yun, D.S., Lee, J.W., Lee, S.K., and Kwon, O.-C., (2011), "Development of Mobile Common Component for providing vehicle information on mobile device", Proceedings Computer Sciences and Convergence Information Technology (ICCIT), 2011 6th International Conference on, IEEE, pp. 809-812.

[12] Koreshoff, T.L., Robertson, T., and Leong, T.W., (2013), "Internet of things: a review of literature and products", Proceedings Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration, ACM, pp. 335-344.

[13] Society of Automotive Engineers, (1988), SAE On-board Diagnostics for Light and Medium Duty Vehicles Standards Manual, Society of Automotive Engineers, Warrendale, PA.

[14] FIA, (2015), My car, my data - technical study, Fédération Internationale de l'Automobile, Region I, Brussels.

[15] Johanson, M., (2011), Information and Communication Support for Automotive Testing and Validation, INTECH Open Access Publisher.

[16] Johanson, M., Dahle, P., and Söderberg, A., (2011), "Remote Vehicle Diagnostics over the Internet using the DoIP Protocol", Proceedings Systems and Networks Communications (ICSNC), pp. 226-231.

[17] Tahat, A., Said, A., Jaouni, F., and Qadamani, W., (2012), "Android-based universal vehicle diagnostic and tracking system", Proceedings Consumer Electronics (ISCE), 2012 IEEE 16th International Symposium on, IEEE, pp. 137-143.

[18] Hunter, G.W., Ross, R.W., Berger, D.E., Lekki, J.D., Mah, R.W., Perey, D.F., Schuet, S.R., Simon, D.L., and Smith, S.W., (2013), A Concept of Operations for an Integrated Vehicle Health Assurance System, NASA Center for Aerospace Information, Hanover, MD.

[19] Phalake, M., and Bhalerao, D., (2011), "Vehicle telematics system using GPRS", International Journal of Computer Technology and Applications, Vol. 2 (1).

[20] Hung, S.T., (1998), "CVDAS-a data acquisition/retrieval architecture for statistical characterization of vehicle usage", Proceedings AUTOTESTCON Systems Readiness Technology Conference, IEEE, pp. 239-246.

[21] Al-Taee, M., Khader, O.B., and Al-Saber, N., (2007), "Remote monitoring of vehicle diagnostics and location using a smart box with Global Positioning System and General Packet Radio Service", Proceedings ACS International Conference on Computer Systems and Applications (AICCSA), IEEE, pp. 385-388.

[22] Hilpert, H., Thoroe, L., and Schumann, M., (2011), "Real-time data collection for product carbon footprints in transportation processes based on OBD2 and smartphones", Proceedings 44th Hawaii International Conference on System Sciences (HICSS), IEEE, pp. 1-10.

[23] Amor-Segan, M., McMurran, R., Dhadyalla, G., and Jones, R., (2007), "Towards the Self Healing Vehicle", Proceedings Institution of Engineering and Technology Conference on Automotive Electronics, IET, pp. 1-7.

[24] Barabba, V., Huber, C., Cooke, F., Pudar, N., Smith, J., and Paich, M., (2002), "A multimethod approach for creating new business models: The General Motors OnStar project", Interfaces, Vol. 32 (1), pp. 20-34.

[25] Benedettini, O., Baines, T., Lightfoot, H., and Greenough, R., (2009), "State-of-the-art in integrated vehicle health management", Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, Vol. 223 (2), pp. 157-170.

[26] Cutter, D.M., and Thompson, O.R., (2005), "Condition-Based Maintenance Plus Select Program Survey", http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA430668, Accessed February 15, 2016.

[27] Koç, M., and Lee, J., (2002), "E-manufacturing-fundamentals, requirements and expected impacts", Proceedings International Conference on Responsive Manufacturing (ICRM), Turkey.

[28] Marek, A., Thorley, S., and Haubmann, H., (2012), "The age of intelligence: online diagnostic and automated maintenance planning", Proceedings International Platinum Conference, The Southern African Institute of Mining and Metallurgy, Sun City, pp. 91-100.

[29] Dienst, S., Fathi, M., Abramovici, M., and Lindner, A., (2014), "Development of a knowledge-based feedback assistance system of product use information for product improvement", International Journal of Product Development, Vol. 19 (4), pp. 191-210.

[30] Ellen Macarthur Foundation, (2016), "Enit Systems: The Enit Agent", in: Intelligent Assets: Unlocking the Circular Economy Potential. Appendix: Selected Case Studies, Ellen Macarthur Foundation, Cowes, p. 15.

[31] Lee, J., (1995), "Machine performance monitoring and proactive maintenance in computer-integrated manufacturing: review and perspective", International Journal of Computer Integrated Manufacturing, Vol. 8 (5), pp. 370-380.

[32] Ellen Macarthur Foundation, (2016), "Arup: IoT in Construction and Infrastructure", in: Intelligent Assets: Unlocking the Circular Economy Potential. Appendix: Selected Case Studies, Ellen Macarthur Foundation, Cowes, p. 9.

[33] Ellen Macarthur Foundation, (2016), "Cisco: Europe's first SmartROAD", in: Intelligent Assets: Unlocking the Circular Economy Potential. Appendix: Selected Case Studies, Ellen Macarthur Foundation, Cowes, p. 17.

[34] Ellen Macarthur Foundation, (2016), "Philips City Touch: Public Lighting Services", in: Intelligent Assets: Unlocking the Circular Economy Potential. Appendix: Selected Case Studies, Ellen Macarthur Foundation, Cowes, p. 19.

[35] Ellen Macarthur Foundation, (2016), "Delta Development: Park 20|20 and Schiphol Trade Park", in: Intelligent Assets: Unlocking the Circular Economy Potential. Appendix: Selected Case Studies, Ellen Macarthur Foundation, Cowes, p. 14.

[36] Ellen Macarthur Foundation, (2016), "HP: Instant Ink", in: Intelligent Assets: Unlocking the Circular Economy Potential. Appendix: Selected Case Studies, Ellen Macarthur Foundation, Cowes, p. 16.

[37] Bodenhoefer, K., Schneider, A., Cock, T., Brooks, A., Sands, G., Allman, L., Simon, M., Chong, S., Yang, X., and Delannoy, O., (2004), "Environmental life cycle information management and acquisition–first experiences and results from field trials", Proceedings Electronics Goes Green, pp. 5-8.

[38] Gu, J., Vichare, N.M., Tinsley, E.C., and Pecht, M.G., (2009), "Computer usage monitoring for design and reliability tests", Components and Packaging Technologies, IEEE Transactions on, Vol. 32 (3), pp. 550-556.

[39] Bleda, A.L., Maestre, R., Santa, G., Jara, A.J., and Skarmeta, A.G., (2012), "Web of Things as a product improvement tool: Furniture as case study", Proceedings Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), , IEEE, pp. 846-851.

[40] Wilson, M., (2015), "Microsoft and Miele team up to cook up an IoT revolution" Betanews, http://betanews.com/2015/04/12/microsoft-and-miele-team-up-to-cook-up-an-iot-revolution/, Accessed February 11, 2016.

[41] European Federation of National Maintenance Societies vzw, What does EFNMS stand for?, http://www.efnms.org/What-EFNMS-stands-for/m13l2/What-EFNMS-stands-for.html, Accessed February 16, 2016.

[42] Vichare, N., Rodgers, P., Eveloy, V., and Pecht, M., (2007), "Environment and usage monitoring of

electronic products for health assessment and product design", Quality Technology & Quantitative Management, Vol. 4 (2), pp. 235-250.
[43] Redding, L., (2015), "Through-Life Engineering Services: Definition and Scope: A Perspective from the Literature", in: Through-life Engineering Services, Springer, pp. 13-28.
[44] Ellen Macarthur Foundation, (2016), Intelligent Assets: Unlocking the Circular Economy Potential, Ellen Macarthur Foundation, Cowes.
[45] Lu, N., Cheng, N., Zhang, N., Shen, X., and Mark, J.W., (2014), "Connected vehicles: solutions and challenges", Internet of Things Journal, IEEE, Vol. 1 (4), pp. 289-299.
[46] General Motors Company, OnStar website, https://www.onstar.com, Accessed February 15, 2016.
[47] Van Horn, D., Olewnik, A., and Lewis, K., (2012), "Design analytics: capturing, understanding, and meeting Customer needs using big data", Proceedings International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (IDETC), ASME, pp. 863-875.
[48] Främling, K., Holmström, J., Loukkola, J., Nyman, J., and Kaustell, A., (2013), "Sustainable PLM through intelligent products", Engineering Applications of Artificial Intelligence, Vol. 26 (2), pp. 789-799.
[49] Elliott, T., (2009), Drink Dispenser Analytics: Coca-Cola Goes Freestyle, With Help from SAP B, http://timoelliott.com/blog/2009/06/drink-dispenser-analytics-coca-cola-goes-freestyle-with-help-from-sap-bi.html, Accessed February 11, 2016.
[50] Heppelmann, J.E., (2015),"PTC Windchill 11 telecast", http://www.ptc.com/product-lifecycle-management/windchill/11/telecast, Accessed December 28, 2015.
[51] Frost & Sullivan, (2015), "PTC - 2105 Global IoT PLM Technology leadership Award", New York, http://www.ptc.com/~/media/Files/PDFs/PLM/Windchill/PTC-Inc-Award-Write-Up.ashx?la=en, Accessed April 22, 2016.
[52] Kiritsis, D., (2011), "Closed-loop PLM for intelligent products in the era of the Internet of things", Computer-Aided Design, Vol. 43 (5), pp. 479-501.
[53] Bandyopadhyay, D., and Sen, J., (2011), "Internet of things: Applications and challenges in technology and standardization", Wireless Personal Communications, Vol. 58 (1), pp. 49-69.
[54] Chen, M., Mao, S., and Liu, Y., (2014), "Big data: A survey", Mobile Networks and Applications, Vol. 19 (2), pp. 171-209.
[55] Barnaghi, P., Sheth, A., and Henson, C., (2013), "From Data to Actionable Knowledge: Big Data Challenges in the Web of Things [Guest Editors' Introduction]", Intelligent Systems, IEEE, Vol. 28 (6), pp. 6-11.
[56] Chen, Y.-K., (2012), "Challenges and opportunities of internet of things", Proceedings Asia and South PacificDesign Automation Conference (ASP-DAC), IEEE, pp. 383-388.
[57] Miorandi, D., Sicari, S., De Pellegrini, F., and Chlamtac, I., (2012), "Internet of things: Vision, applications and research challenges", Ad Hoc Networks, Vol. 10 (7), pp. 1497-1516.
[58] Aggarwal, C.C., Ashish, N., and Sheth, A.P., (2013), "The Internet of Things: A Survey from the Data-Centric Perspective", 2013.
[59] Aggarwal, C.C., (2008), "On unifying privacy and uncertain data models", Proceedings Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on, IEEE, pp. 386-395.
[60] Hoffman, D.L., Novak, T.P., and Peralta, M., (1999), "Building consumer trust online", Communications of the ACM, Vol. 42 (4), pp. 80-85.
[61] Podbregar, I., (2006), "Some patterns of industrial espionage", Journal of Criminal Justice and Security, Vol. 8 (3,4), pp. 323-331.
[62] Matwyshyn, A.M., (2005), "Material Vulnerabilities: Data privacy, corporate information security, and securities regulation", Berkeley Bus. LJ, Vol. 3, p. 129.
[63] Zhao, K., and Ge, L., (2013), "A survey on the internet of things security", Proceedings Computational Intelligence and Security (CIS), 2013 9th International Conference on, IEEE, pp. 663-667.
[64] Botta, A., de Donato, W., Persico, V., and Pescapé, A., (2016), "Integration of cloud computing and internet of things: a survey", Future Generation Computer Systems, Vol. 56, pp. 684-700.
[65] Xiaohui, X., (2013), "Study on security problems and key technologies of the internet of things", Proceedings Computational and Information Sciences (ICCIS), 2013 Fifth International Conference on, IEEE, pp. 407-410.
[66] Gang, G., Zeyong, L., and Jun, J., (2011), "Internet of things security analysis", Proceedings International Conference on Internet Technology and Applications (iTAP), IEEE, pp. 1-4.
[67] Groopman, J., (2014), "Interoperability: the challenge facing the Internet of Things", http://www.altimetergroup.com/2014/02/interoperability-the-challenge-facing-the-internet-of-things/, Accessed February 16, 2016.
[68] IBM Institute for Business Value, (2015), "Device Democracy - Saving the future of the Internet of Things", http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=XB&htmlfid=GBE03620USEN, Accessed February 16, 2016.