# From Big Data to Smart Data: Changing Behavior with Online Communication

Anna Fensel
Semantic Technology Institute (STI) Innsbruck,
Department of Computer Science, University of Innsbruck
Innsbruck, Austria
anna.fensel@sti2.at

## Abstract

The humanity is rapidly developing and persistently experiencing local and global challenges, such as global warming/climate change, disbalances in demand and supply, reaching efficiency and handling complexity. Mastering most (if not all) of them require a behavior change. At the same time on the technical side, the dramatic growth of data volumes (Big Data) in the infrastructures, the data's heterogeneity and increased power and impact on the people's daily lives are calling for new methods, practices and policies for data management. Herewith the role of semantic technology becomes even more crucial: particularly, when it comes to providing and sharing a well-defined meaning, reducing complexity, and eventually delivering Smart Data for a functional and fair data value chain. I discuss Smart Data service enablers for interoperation, interactivity and human participation and behavior change, as well as demonstrate them in applications going beyond the current state of the art in the domains of energy efficient smart buildings, and online communication and marketing. In conclusions, I outline some of the data value chain -related challenges for future work.

CEUR-WS.org/Vol-2316/invited1.pdf

## 1 Introduction

"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation." - this statement of Tim Berners-Lee has gained even more relevance since the start of this century, in particular, taking into account the new issues (e.g. fake news, inequality) that have appeared with the spread and the success of the Web and the Semantic Web[1].

The humanity per se is also rapidly developing and persistently experiencing local and global challenges, such as global warming/climate change, dis-balances in demand and supply, among many others. Mastering most (if not all) of them require a behavior change. Behavioral change is difficult to achieve per se, and it is important that technology – as a major enabler - has a positive rather than a negative impact here. Further, the dramatic growth of data volumes (Big Data, Internet of Things) and the data's increased power and impact and on the people's daily lives are calling for new types, practices and policies of behavior with data. These factors made the role of semantic technology even more crucial: in terms of providing a well-defined meaning, and eventually delivering Smart Data for a functional and fair data value chain.

1 „Father of the Web Confronts his Creation in the Era of Fake News", URI: https://www.bloomberg.com/news/articles/2017-11-13/father-of-the-web-confronts-his-creation-in-the-era-of-fake-news, November 13, 2017.

Copyright held by the authors. NOBIDS 2018

Such data value chain should be realized for a number of domains in practice, and the domain of online communication and media is often one of the forerunners, given the vast availability of the online data and content. Current research, for example, is addressing important directions such as empowerment of the end users within the recommender systems [Gul18], user privacy matters [Moh18] and semantically enabled secure access to the data [Gar11], as well as creating of scalable basic infrastructures supporting ubiquitous communication and media consumption [Opd17]. Despite extensive technology developments in the addressed area, the take up in practice still lags behind, as can be demonstrated by the studies of the actual technology use.

This paper is structured as follows. A general approach to the addressed challenge is presented in Section 2. Specific examples are provided in Section 3. Section 4 concludes the paper.

# 2 General Approach

In this section, I outline three developments employable in the implementation of the behavaioural change: namely, knowledge graphs and data management with them, personalized media consumption, as well as Explainable Artificial Intelligence (AI).

## 2.1 Data Management with Knowledge Graphs

Knowledge Graphs are becoming a key technology for large-scale information processing systems containing massive collections of interrelated facts [Pau17]. Examples include the Google's Knowledge Graph with over 70 billion facts (in 2016), dataCommons, DBPedia, YAGO, NELL, and Knowledge Vault, a very large scale probabilistic knowledge base created by information extraction methods for unstructured or semi-structured information. Specifically, Knowledge Graphs provide the means for development of the newest data methods for data management, data fusion / data merging, and graph and network optimization and modeling, serving as a source of high quality data and a base for a web-scale information integration. In particular, Enterprise Knowledge Graphs help to infer new relationships out of existing facts, giving context and meaning to the content, and can be used in applications.

Creation and population of such Knowledge Graphs from the data, that is often of inferior quality and lacks sufficient context information, comprises a number of challenges. These challenges, for instance, require resolution of the needs such as duplication elimination, error correction, range prediction. They can be addressed with data analytics and machine learning techniques, as well as the human engagement, to ensure the presence of the semantics in the resulting outcome.

Further, intelligent data value chain production and consumption ecosystems require new methods for automated exchange of and reasoning about the information across systems. For example, the data generated by an image analysis system could be semantically represented, shared and employed across numerous systems, taking into account their aims and requirements, as well as the specifics and provenance of the generated data. These methods are to employ semantic technologies, which provide standards for the data production and consumption, and comprise facilitating solutions such as for semantic licensing of data and content.

## 2.2 Personalized Media Consumption

There are different types of consumers/ personality profiles, and the currently sometimes still deployed in media and online communication "one-size-fits-all" methods, as well as just one "typical" combination of methods are below the potential in effectively addressing certain media consumer groups on every media channel. In the worst case scenario, the "wrong" mix can cause most of the media consumers to disengage and eventually unfollow the channel. While user segmentation and personalized addressing is very common in such fields as marketing or gaming (e.g. see existing categories for user gamification types [Ton16]), the studies for similar directions in eLearning domain are appearing (see [Gil15]), the media production and consumption also is moving in this direction in practice. Also, in addition to the personality types, profiling the media consumers by the prior types and quantities of the consumed media, and the media consumption habits / speed that is the most efficient for them is important for finding the mix of methods that would produce the maximum engagement.

Personalization is also crucial in addressing behavioral change. Apart from understanding the profiles of the consumers, it is also essential to understand the conditions under which people change their minds and behaviors [Gar06]. Employing semantic analysis to identify the conditions which can be most effectively appealed to in the case of a specific consumer or a consumer group would make online communication and media more effective.

### 2.3 Explainable AI

Last but definitely not least, the data management needs to be explainable and actionable for the end users, thus involving data visualization and communication methods. The need for transparent and explainable data sharing has been highly emphasized in the last years. In particular, Big Data research and policy roadmaps have been produced for the European Union, and making the Big Data more transparent for the end users has been appearing as a requirement among the conclusions of the extensive roadmapping studies [Cuq16].

Designing concepts and prototypes for transparent and explainable sharing of the data becomes an essential and often legally required part in the project development. The solutions typically should facilitate the understanding of the data sharing obligations and permissions, both for the data owners as well as for the data users. The actual data sharing workflows and usages can be also to be made traceable and displayable for the data owners (e.g. employing the blockchain facility), giving them a better understanding of the actual usage and thus value of their data.

Such development have to do with the area of explainable AI, which is currently an active field. Approaches and solutions are being developed for explaining machine learning (including deep learning) to the users, but less is made in explaining the data sharing, display of knowledge graphs to humans in a clearer or more perceivable manner, particularly, within specific projects' contexts and domains.

## 3 Examples

In this part of the paper, I showcase the solutions implementing a knowledge graph -based data management lifecycle, i.e. demonstrate them in advanced applications from the domains of energy efficiency, and digital transformation. I list technology solutions investigating behavioral change in the domain of smart home appliances (OpenFridge project), behavioral change towards reaching higher levels of energy efficiency (Entropy project), online communication and marketing in the area of tourism (TourPack project), and finally, the legal aspects of the data and content licensing (DALICC project).

### 3.1 OpenFridge – Internet of Things Data Publishing and Service Ecosystem

Addressing the behavioral change with Smart Data, innovative directions include potential ICT solutions investigating the domain of energy efficient buildings. Particularly, our completed OpenFridge experiment [Fen17] belongs to this category. It comprised design and development of the Internet of Things data system with semantic and data analytics enablers for building new services on a top of typical home appliance data — in particular, refrigerators. The system has been evaluated with real life end-user pilots, and the constructed knowledge graph and the experiment data has been published openly in a semantic format.

The summary of the OpenFridge project is as follows:

"*While the mass consumers' demand and expectations in the energy efficiency field grow, the providers and manufacturers of electrical appliances are searching for the approaches and infrastructures enabling them to build new kind of added-value services, based on the large volumes of data available from the appliances. Thus, the goal of the project OpenFridge is to design and develop a pilot simple and scalable Internet of Things data infrastructure, empowering building new services based on the typical home appliance data, e.g. data on the energy consumption of the fridge. The infrastructure is comprised of semantic domain models for opening of the appliances data, data analytics module for aggregation of the raw Internet of Things energy data in adding value energy efficiency information, as well as provisioning of this information to interested stakeholders (appliance manufacturers, end users, utilities, municipalities, etc.) under the new access mechanisms and business models.*"

The experiment and the final end user survey has indicated that most people see the potential in the behavior change, and also would be sharing their consumption data in an anonymized manner. So here is a clear development potential for steering the behavioral change with the data sharing, which would be pursued further in data value chain infrastructures.

### 3.2 Personalized Energy Efficiency Services in Buildings

To extend the topic of the intelligent energy efficient buildings that have an impact on the behavior of the end users, the project ENTROPY - Design of an innovativE eNergy-aware it ecosysTem for motivating behavioRal changes towards the adOption of energy efficient lifestYles[2] has been recently completed. The technical development is advancing the field of intelligent building systems, where the knowledge graphs and rules have already been applied in real world

2ENTROPY project: http://entropy-project.eu settings [Fot17]. Further, the work here is conducted together with a team of psychologists that take part into designing of energy efficiency recommendations for the users, taking into account their personality profiles.

The summary of the ENTROPY project is as follows:

*"Taking into account the fact that buildings constitute the largest end-use energy consuming sector, the design and development of solutions targeted at reducing their energy consumption based on the adoption of energy efficient techniques and the active engagement of citizens/occupants is considered crucial. Innovative solutions have to be implemented upon properly understanding the main energy consuming factors and trends, as well as properly modeling and understanding the citizens' behavior and the potential for lifestyle changes.*

*The ENTROPY project addresses this challenge by building upon the integration of technologies that facilitate the deployment of innovative energy aware IT ecosystems for motivating end-users' behavioral changes and namely: (1) the Internet of Things that provides the capacity for interconnecting numerous devices and applying energy-efficient communication protocols, (2) the evolution of advanced Data Modelling and Analysis techniques that support the realization of semantic models and knowledge extraction mechanisms and (3) the Recommendation and Gamification eras that can trigger interaction with relevant users in social networks, increase end users' awareness with regards to ways to achieve energy consumption savings in their daily activities and adopt energy efficient lifestyles as well as provide a set of energy efficient recommendations and motives."*

A number of the ENTROPY results have been already published, e.g. the design of the semantic part of the system and the designed ontologies [Sim16], as well as design and implementation details of the whole system [Fot17]. The recommendations towards energy saving have been delivered in a personalized manner to the users via mobile phones, according to the pre-derived psychological user profiling as defined with the gamification types [Ton16]. The project has been implemented in real buildings in three countries (Italy, Spain, Switzerland), and the experiments have demonstrated a ca. additional 13% savings in the energy consumption, directly attributed to the behavioral change of the users.

### 3.3 TourPack – Touristic Service Packaging

As in many service-oriented businesses nowadays, the touristic service consumers want individualized experiences and no longer want the "one-size fits all" touristic packages, as, for example, produced in a generic way by travel agencies. Thus, the aim of the TourPack project[3], in the settings of which we develop our approach, is "to design and a prototype a production system that creates "on-demand" touristic packages catering to the individual touristic service consumer needs and preferences – applying the smart usage of the open and proprietary data for the information integration and service composition, and eventually, improving the multi-stakeholder data-driven production processes of touristic service offer" [Fen15], or, more specifically *"While the touristic service offers become present and bookable in abundance on the Information and Communication Technology (ICT) communication channels, TourPack aims to build a linked data -empowered system for touristic service packaging. Integrating information from multiple sources and systems employing linked data as a global information integration platform, and mining from the depths of the "closed" data, the touristic service package production system will be able to cater to creating the most optimal travel experience for the traveler. Further, the service packages will be efficiently published and made bookable to the end consumers via intelligently selected most suitable communication and booking channels: especially the ICT channels with rapidly growing user audiences, such as the social media and the mobile apps."*.

Within the project's technical development a vast amount of data and services to be used for such personalized online communication has discovered, and also much of that online communication has been shown to be made automatic with the use of semantic technologies and ontologies [Akb17]. However, there has also been many challenges encountered, the most complex and time consuming having to do with the data quality and heterogeneity, as well as with the services heterogeneity and their integration aspects, including the legal conditions imposed by practice.

### 3.4 Semantic Data Licensing

Last, but not least, for the efficient data and content reuse, enabling semantically annotated data licenses is important, as they can facilitate the correct legal (re-)use of the data and content. Currently, such languages and tools are being developed the Permissions and Obligations Expression (POE) Working Group at W3C[4], and in particular within the project DALICC – Data Licenses Clearance Center[5]. The summary of the DALICC project is as follows:

"*The creation of derivative data works, i.e. for purposes like content creation, service delivery or process automation, is often accompanied by legal uncertainty about usage rights and high costs in the clearance of licensing issues. The*

3 TourPack project: http://tourpack.sti2.at

4 POE group at W3C: https://www.w3.org/2016/poe/wiki/Main_Page

5 DALICC project: https://www.dalicc.net

*DALICC project aims to develop a software framework that supports the automated clearance of rights issues in the creation of derivative data works. In essence, DALICC helps to determine which information can be shared with whom to what extent under which conditions, thus lowering the costs of rights clearance and stimulating the data economy*."

The project has developed a library of semantic data and content licenses [Pan18], and its infrastructure [Pel18] can be freely used to select, create, compose licenses, and use their semantic export in the software solutions. Further steps would include the direct integration of the solution in the data and content production systems, which are relevant for numerous domains, including media in particular. The systems can be also technically implemented employing new types of distributed systems such as blockchain and smart contracting, essentially implementing the vision of the semantic web services.

## 4 Conclusions and Future Work

As a conclusion, it is possible to see semantic technology and particularly knowledge graphs as a mature and powerful instrument, which works in practice and has many instantiations. The technology is particularly useful for:

- Communication between a computer and a human, and between computers,
- Information integration, serving as a global reference,
- Bringing over parts that cannot be "learned" (when too few data are available, in matters that cannot be machine-learned, such as ethics or privacy).

A lot of innovation currently comes from new combinations of various existing systems, and managing behavioral change is one of the fields that can be addressed in this manner. Also, combining machine learning and semantics brings new results at a technical level, while interdisciplinary research with fields like psychology, sociology and law play increasing in importance roles.

Integrating humans in the loop will be made in a more advanced manner, as well as of the communication between the information technology infrastructures. For example, while now the human machine interaction over voice assistants / chatbots is becoming to be in the mainstream, the human machine interaction of the future will becoming even more simplified for the humans and will increasingly rely on human sensing. Simpler human sensing solutions (e.g. eye tracking) are already getting close to the broad practical application, and are being implemented with typical solutions for the semantic Internet of Things. The trend is in the progress, and now also the first attempts towards more complex developments, such as a semantic standard for brain computer interfaces, are appearing [Jos18]. With semantic technologies, the knowledge graph basing methods will be also capturing the information about the trends and will be trying to envision and represent more of the data from the future, and less of the data from the past. Finally, such societal aims as raising efficiency, bringing transparency, inclusion and user empowerment, thus building the "internet for people" should be always on the top of the agenda.

**Acknowledgements**

This work has been supported by Horizon 2020 project ENTROPY and FFG project DALICC.

## References

[Akb17] Z. Akbar, A. Fensel, & D. Fensel (2017). An ontology-based coordination and integration of multi-channel online communication. International Journal of Metadata, Semantics and Ontologies, 12(4), 219-231.

[Cuq16] M. Cuquet, & A. Fensel, A. (2018). The societal impact of big data: A research roadmap for Europe. *Technology in Society*.

[Fen15] A. Fensel, E. Kärle, I. Toma (2015). "TourPack: Packaging and Disseminating Touristic Services with Linked Data and Semantics". In Proceedings of the 1st International Workshop on Semantic Technologies (IWOST), CEUR Workshop Proceedings, Vol-1339, ISSN 1613-0073, pp. 43-54, 11-12 March 2015, Changchun, China.

[Fen17] A. Fensel, D. K. Tomic, & A. Koller (2017). Contributing to appliances' energy efficiency with Internet of Things, smart data and user engagement. *Future Generation Computer Systems*, *76*, 329-338.

[Fot17] E. Fotopoulou, A. Zafeiropoulos, F. Terroso-Sáenz, U. Şimşek, A. González-Vidal, G. Tsiolis, P. Gouvas, P. Liapis, A. Fensel, & A. Skarmeta (2017). Providing Personalized Energy Management and Awareness Services for Energy Efficiency in Smart Buildings. *Sensors*, *17*(9), 2054.

[Gar11] A. García-Crespo, J. M. Gómez-Berbís, R. Colomo-Palacios, & G., Alor-Hernández (2011). SecurOntology: A semantic web access control framework. *Computer Standards & Interfaces*, *33*(1), 42-49.

[Gar06] H. Gardner (2006). *Changing minds: The art and science of changing our own and other people's minds*. Harvard Business Review Press.

[Gil15] B. Gil, I. Cantador, & A. Marczewski (2015). Validating gamification mechanics and player types in an E-learning environment. In *Design for Teaching and Learning in a Networked World* (pp. 568-572). Springer, Cham.

[Gul18] J. A. Gulla, Ö. Özgöbek, & X. Su (2018). Exploratory News Recommendation. In *Developments and Trends in Intelligent Technologies and Smart Systems* (pp. 1-15). IGI Global.

[Jos18] S. José, & R. Méndez (2018, October). Modeling actuations in BCI-O: a context-based integration of SOSA and IoT-O. In *Proceedings of the 8th International Conference on the Internet of Things* (p. 46). ACM.

[Moh18] I. Mohallick, K. De Moor, Ö. Özgöbek, & J. A. Gulla (2018, December). Towards New Privacy Regulations in Europe: Users' Privacy Perception in Recommender Systems. In *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage* (pp. 319-330). Springer, Cham.

[Opd17] A. L. Opdahl (2017). Load-time reduction techniques for device-agnostic web sites. *Journal of Web Engineering*, *16*(3&4), 311-346.

[Pan18] O. Panasiuk, S. Steyskal, G. Havur, A. Fensel, & S. Kirrane (2018, June). Modeling and Reasoning over Data Licenses. In *European Semantic Web Conference* (pp. 218-222). Springer, Cham.

[Pel18] T. Pellegrini, V. Mireles, S. Steyskal, O. Panasiuk, A. Fensel, & S. Kirrane (2018). Automated Rights Clearance Using Semantic Web Technologies: The DALICC Framework. In *Semantic Applications* (pp. 203-218). Springer Vieweg, Berlin, Heidelberg.

[Pau17] H. Paulheim (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, *8*(3), 489-508.

[Ton16] G.F. Tondello, R.R. Wehbe, L. Diamond, M. Busch, A. Marczewski, & L. E. Nacke (2016, October). The gamification user types hexad scale. In *Proceedings of the 2016 annual symposium on computer-human interaction in play* (pp. 229-243). ACM.

[Sim16] U. Simsek, A. Fensel, A. Zafeiropoulos, E. Fotopoulou, P. Liapis, T. Bouras, F.T. Saenz, A.F. Skarmeta Gómez (2016). A Semantic Approach towards Implementing Energy Efficient Lifestyles through Behavioural Change. In *Proceedings of the 12th International Conference on Semantic Systems, SEMANTICS'16*, Leipzig, Germany, pp 173-176, ACM.

[Sta13] I. Stavrakantonakis, I. Toma, & D. Fensel (2013). Hotel websites, Web 2.0, Web 3.0 and online direct marketing: The case of Austria. In *Information and communication technologies in tourism 2014* (pp. 665-677). Springer, Cham.