DAVID PUBLISHING

Sustainable Futures: Comparing Methodologies for Analyzing Citizen Visions in Europe

Petteri Repo[a], Kaisa Matschoss[a], Päivi Timonen[a]

Abstract

There are increasing calls for engaging citizens in the development of future outlooks. At the same time, large-scale public engagement activities warrant appropriate methods for analyzing their outcomes. This paper reviews how topic modeling could provide such a methodology, which both accounts for all textual data collected in public engagement activities, however large in scope, yet also allows for meaningful topical analysis. It compares topic modeling results concerning a corpus of 179 citizen visions from 30 European countries on desirable and sustainable futures to those acquired through deliberative analysis. While both methodologies contend that European citizens' outlook consists of education, sustainability in the economy, health concerns, and fairness in communities, and the particular strengths of topic modeling relate to its documentability, repeatability, cost efficiency, and scalability. Topic modeling can also be considered to support public engagement analytically from the perspective of knowledge formation rather than that of common sense.

Keywords

Topic modeling, citizen visions, deliberation, public engagement, European research and innovation programs

Citizens are increasingly engaged in coproducing sustainable futures for reasons ranging from improving the quality of insights to public accountability and empowerment (Renn and Schweizer 2009; Jasanoff 2003). In Europe, there is a growing strand of public engagement activities that target the steering of research and innovation activities. Yet, public engagement is costly to carry out in particular when it involves inviting large numbers of people and organizers to attend physical or virtual spaces to follow a common procedure. Therefore, it is of key concern to analyze the documented outcomes of participation events properly to make the engagement efforts worthwhile.

New methodologies developed in the domain of digital humanities could serve this aim. Topic modeling, and more precisely latent Dirichlet allocation, is a methodology to identify word clusters (i.e. "topics") in sets of documents (e.g. Gläser, Glänzel, and Scharnhorst 2017; Yau et al. 2014). This methodology is particularly suitable for discovering topical patterns that are spread across sets of documents as is often the case when public engagement activities provide textual documentation (Stoneman, Sturgis, and Allum 2013).

To review the particular usability of topic modeling and to examine its concurrent features, the authors apply the methodology of topic modeling to the outcomes of a large-scale European public engagement exercise of the CIMULACT (Citizen and

[a]University of Helsinki, Finland

Correspondent Author:
Petteri Repo, Unioninkatu 40, 00014 Helsingin yliopisto, Helsinki, Finland

Multi-Actor Consultation on Horizon 2020) project in which over 1,000 citizens in 30 countries[1] developed 179 visions on desirable and sustainable futures. This public engagement exercise is the latest and largest in a series of the European Commission's attempts to invite citizen contributions to the development of European research and innovation agendas.

Using the visions as data, the authors compare the results obtained with the topic modeling methodology (topics), against those obtained from deliberative qualitative analysis executed by experts in a one-day workshop (themes). In particular, the authors examine: (1) if results from topic modeling conform with those from deliberative analysis; and (2) how these two methodologies differ in terms of validity and reliability, abstraction, cost efficiency, concept of public engagement, and the implications of their differences.

The upcoming sections first describe why and how public engagement has gained a growing role in research and innovation agenda formation in Europe. Then the paper introduces the examined engagement data and describes the compared methodologies. Next, outcomes of the both methodologies are presented and matched. The concluding section of the paper discusses the insights obtained from the comparative analysis from the perspectives of the research questions.

EUROPE SEEKS CITIZEN VISIONS FOR FUTURE RESEARCH AND INNOVATION AGENDAS

The European Union actively engages citizens in the development of European research and innovation programs that focus on sustainability. Such citizen engagement has been accomplished in research activities funded by the European Commission in several domains. Valued experiences from citizens' focus group interviews in the realm of waste management (VOICES 2015) were thematically broadened with engagement that reviewed public health genomics and aging society through citizen consultations and technology assessment (PACITA 2016). Public participation has also been guided towards sustainable innovation (Bedsted et al. 2016), leading to the development of engagement methodology in which citizens first envision futures that are together with expert assistance developed into input for research and innovation programs (CIVISTI 2011). The CIMULACT project belongs to the same category of citizen engagement exercises as it invites citizens to envision desirable and sustainable futures, which are later transformed to input for research and innovation programs (Jørgensen and Schøning 2016).

The accomplished engagement projects reflect enduring discussions in Europe that call for dialogue between institutions and citizens in societal debates (Ducci 2013; Smallman 2017). It has been of particular concern to engage citizens in the agenda setting stage of policy processes. Such "upstream" engagement has also been a policy response to public distrust in science especially on issues of sustainability (Burgess and Chilvers 2006; Wilsdon and Willis 2004). Accordingly, deliberation of new technologies and scientific discoveries starts early in order to modify the relationship between scientific and public decision-making, to avoid problems with public acceptance and to provide a broader view than only that of risk avoidance (Wilsdon and Willis 2004). This is in contrast with the "downstream" approach such as citizen science, which is a form of public engagement applied especially in America, as well as open science that focuses typically on the advancement of science or the creation of new technologies (Bowser and Shanley 2013).

Public engagement is by no means unproblematic regardless of its upstream or downstream focus. Even when it addresses large-scale societal developments, it frequently relies on a collection of small-scale activities (Niemeyer 2011; Smallman 2017). Such was also the case with the studied citizen engagement

when the 179 citizen visions were developed by 1,088 citizens in 30 European countries. One event was organized in each participating country, making the engagement exercise a collection of similarly executed workshops rather than a unifying, common event. It is then a highly complex task to make sense of the joint contribution of the participants of the separate events. People want to be heard in their own terms instead of those of scientists or other elites (see Nyaga and Torres 2015; Usman 2014), yet, outcomes still need to be conceptualized to become more exact and universally understood, as well as translated to further uses.

It is in this context that an analysis of the outcomes of public engagement merits particular attention. When public engagement is of large scale, taking place in numerous locations and with various compositions of attendants, it has been considered useful to involve event organizers and invited experts in the analysis of outcomes to ensure that local contexts are properly considered. Such a deliberative analysis may indeed raise the quality of the results and, therefore, further the impacts of the engagement. Nevertheless, deliberative analysis may itself be costly, especially if it involves inviting a large number of engagement organizers anew together. Outcomes, processes, and preconditions of public engagement processes indeed merit empirical attention (Cobb 2012).

Qualitative and collaborative methodologies commonly facilitate the analysis of such engagement outcomes, but typically do not scale well to large amounts of data. For instance, the data examined in this paper consist of 179 elaborated citizen visions on desirable and sustainable futures, but the number of analyzed texts could easily be much higher in future online engagement exercises (see e.g. Coleman and Gotze 2001). This paper attempts to circumvent these tensions between contexts and scalability while also targeting methodological efficiency in the analysis of outcomes of public engagement. Doing so, it strives to

retain the richness, diversity, and multitude evident in the outcomes of large-scale engagement activities. These features are important to consider when ensuring that outcomes provide opportunities for citizens to speak out on issues they themselves define, and in ways which do not hinder their messages to come through.

Methodological developments in the analysis of natural languages allow using new approaches in the analysis of public engagement outcomes. This paper reanalyzes the outcomes of a large-scale public engagement process carried out by the CIMULACT project, which produced citizen visions on desirable and sustainable futures, and has been extensively analyzed and reported (Jørgensen and Schøning 2016; Riisgaard et al. 2017). The methodology of topic modeling (Blei, Ng, and Jordan 2003) is applied in the new analysis, identifying patterns of topics across the visions. The key task is then to review how well the results of the two analyses correspond. Thereby, this paper also contributes to issues relating to internal validity and external generalizability of topic modeling, which are still under discussion. This follows the suggested procedure of evaluating validity by comparing results from machine coding with those from manual coding when making sense of texts (Jacobi, van Atteveldt, and Welbers 2016; Gläser et al. 2017). Furthermore, the paper provides a methodological comparison of these two approaches. In order to facilitate the comparison, the next sections first describe the data formed by the citizen visions and the two approaches: topic modeling and the deliberative analysis of the CIMULACT project.

STUDIED VISIONS AND COMPARED METHODOLOGIES

This paper relies on a comparison of two analyses, which utilize the same empirical dataset of citizen visions. The studied set of visions is described in the upcoming section, followed by descriptions of the two

compared methodologies, which have been used in the analysis of the visions. In order to keep the two methodologies apart, the authors refer to the findings brought up by the topic modeling methodology as "topics" and to those emerging from the deliberative methodology utilized in the CIMULACT project as "themes".

Data: Citizen Visions on Desirable and Sustainable Futures

The two analyses presented and compared in this paper are based on a dataset that is formed by 179 citizen visions that were formulated in 30 European countries in so-called "National Citizen Vision" (NCV) workshops in the CIMULACT project between November 2015 and January 2016 (Jørgensen and Schøning 2016; Riisgaard et al. 2017). The citizen engagement process was minutely designed and strictly guided to ensure that an identical format of engagement was accomplished in all workshops to produce comparable yet freely formulated visions.

The key task of the CIMULACT engagement process was to involve citizens in the actual formulation of agendas for the European Union's Horizon 2020 research and innovation program, thereby showcasing an upstream engagement activity. It was expected of the citizen engagement process to provide deep insights on the wider societal needs of the people. To reach this goal, citizens were asked to imagine desirable and sustainable futures and to jointly form visions describing them. A vision, in the context of the CIMULACT NCV workshops, describes what the future should be like regardless of whether or not it is feasible from today's point of view. The strictly steered process through which the citizens were guided was designed to empower the citizens into thinking that the future is something that can be shaped instead of it being predetermined.

The NCV workshops lasted approximately eight hours in each country. Divided in small groups, the citizens in each national workshop jointly created six

visions reflecting their desirable and sustainable futures. The citizens were asked to describe their vision for 2050 and to give concrete examples of what that vision would mean in the daily life of people. They were further asked to explain how the respective vision differs from today, assess its desirability, and consider concerns relating to it. The visions are all of a somewhat positive character because citizens were asked to turn dystopias to visions in which troubles had been solved. The workshop agenda included several stages of elaboration on the visions, giving all participants in each workshop the possibility to deliberate on the visions being developed in their own group and on those developed by other citizen groups. The final visions are each a narrative storyline of a desirable and sustainable future in 2050. Example 1 below describes one of the 179 citizen visions.

> This vision is about self-sufficiency. In 2050 we are much less dependent on fossil fuel. All kinds of energy resources we did not use in 2016 are being used and every house has a solar power installation. Energy is being generated in diverse and creative ways, e.g. from physical exercising also at home or waste that cannot be recycled is converted into energy. Education is the basis of the community and should have a lifelong perspective in order to increase acceptance and involvement towards innovations. Education about nature and environment learns (teaches) people about sustainable energy. The public transport is attractive and cheap and mostly underground. As a consequence of this all, the health levels have risen because there is much less air pollution. Every house has its own water circulation system (with filter). Roofs are used to store water and energy. People eat more conscious: organic, less meat, seasonal products, insects, etc. Nature is preserved. Robot and sensor technology are part of this vision. (Example 1: Excerpt from citizen vision from the Netherlands: Energy)

The full description of the vision is presented in Appendix 1 (Riisgaard et al. 2017).

In addition to the joint procedure, much emphasis was given to the selection of citizens. The most important criterion in the selection of participating citizens was diversity in a national context.

Accordingly, the participants of the NCV workshops represent many kinds of people with regard to age, education, gender, and place of residence. Specific national contexts could be considered to increase variety of participants. In other words, reaching different kinds of people in a deliberative tradition was strived for rather than accomplishing strict representativity according to their background factors. The citizens, furthermore, needed to be lay people in the sense of not working professionally with science, technology, nor innovation. Altogether, 1,088 European citizens participated in the creation of the CIMULACT visions. The visions were initially created in the national languages of the 30 EU countries and later translated into English. The authors' research makes use of the translated visions.

Topic Modeling as a Methodology to Approach Citizen Visions

When engagement is of large scale and creates much qualitative data that need to be analyzed quickly and reliably, computer tools designed for natural language processing are a useful option to rely on (Blei 2012; Gläser et al. 2017). The topic modeling of the citizen's visions relies on latent Dirichlet allocation and is processed with the MALLET toolkit for statistical language processing (McCallum 2002).

In short, topics across the vision data are identified through an analysis of probability of collocation of words, i.e. word clusters. Such clusters of words jointly represent meanings distributed across the visions. The topics are extracted from the full corpus (i.e. all visions), ensuring that all data are considered during the process. In the first stage of the topic modeling exercise, the data are processed to a form that is readable by the MALLET toolkit. In addition to removing repeating headings from the vision template, punctuation and special signs were also discarded from the corpus as well as removing upper-case letters. During the process, stopwords such as a, and, the, etc. are removed from the corpus by making use of the

standard English stopword list of the MALLET toolkit expanded with person names.

It is a key task in topic modeling to determine the number of examined topics (Gläser et al. 2017; Jacobi et al. 2016). In this study, the authors sought to identify the maximum number of topics to describe the data while still being able to explain and distinguish between the topics well. Too few topics miss out on important aspects of the data while too many make it difficult to distinguish between the topics.

Accordingly, an iteration process for the search of an appropriate number of topics was initiated. From previous experience with similar data sets, the authors used seven topics as a starting point before examining five, nine, and 12 topics. Examining the data with five topics indicated that there was one large topic concerned with education in the corpus and that other topics could be interpreted and labeled well. The same large topic had existed in the original run with seven topics, and its smaller topics could readily be explained and labeled. Runs with nine and 12 topics kept the large topic, but made it increasingly difficult to distinguish between the other topics. In a final run with eight topics, the large topic remained and the seven smaller topics could be interpreted and labeled well.

Hence, the corpus was modeled with eight topics. Diagnostics measures such as appearances of words in topics and documents, and their exclusivity were considered when interpreting and labeling topics. Standard hyperparameter optimization every 10 iterations was used to allow the model to better fit the data by allowing some topics to be more prominent than others, i.e. showing their Dirichlet parameter which reflects the weight of the topic in the corpus. While topic modeling includes an element of randomness, different runs produce similar results as the methodology is probabilistic. The results of the modeling are presented in Table 1.

Topic modeling helps to discover reoccurring elements, i.e. topics, in the studied data before

introducing prior knowledge or frames of reference to the analysis. The method does not require any *a priori* interpretations as such, except for the determination of the examined number of topics. Yet, as the methodology does not examine meanings of the words studied, their interpretation becomes a part of the research task.

Deliberative Methodology Applied in the CIMULACT Project

In order to analyze the outcomes of the citizen engagement, the approach of the CIMULACT project was to organize a one-day "Interactive Synthesis Workshop" (Warnke et al. 2017) in which 27 project team members and 11 external experts identified underlying themes in the citizen visions. This deliberative method was intended to increase the validity of the result by ensuring reflection on outcomes and by counteracting possible biases of individual participants, which could emerge from their expert and cultural backgrounds. The identified themes were designed to be later used in the creation of research programs for Europe, and were considered to have strong legitimacy and power particularly due to the deliberative and participatory approach. Such a process of identifying themes in bodies of textual data is an established approach in qualitative methodologies. Corbin and Strauss (1990) as well as others in the qualitative research tradition relate such thematization to categorization and abstraction of content. Themes are to be developed in terms of the dimensions of the phenomenon they depict and in the context in which it is expressed.

Before the workshop, involved project team members and experts each read 30 citizens' visions allocated to them from the complete set of 179 visions. In the first stage of the workshop, the participants were divided in six groups, each of which had a task to identify five "social needs" in the material consisting of the allocated 30 visions (Warnke et al. 2017). The identification of social needs was first

accomplished individually in the groups. Then the suggestions of each member of the group were jointly discussed until a common understanding was reached, which resulted in a total of 29 identified social needs. In the second stage, 12 themes i.e. clusters of the identified social needs were created based on similarities, while each of the initial visions was attached to a suitable theme in small, deliberative groups. This turned out to be a difficult task, as the visions are very diverse in character. Each vision can also encompass a number of social needs and the end result was that most visions were attached to several needs while some were not attached to any need (Warnke et al. 2017). This required a phase of post-processing of visions to social needs by the project team after the workshop. The process up to the creation of 12 themes is presented in Figure 1.

The CIMULACT approach focused at the number of outcomes, finding cross-cutting issues and at raising the abstraction level of themes. Only themes relating directly to a certain sustainability issue (economy, energy, or food) ended up not to correspond to more than one social need. The final themes were named "Citizenship awareness and participation", "Equality", "Green habitats", "Harmony with nature", "Holistic health", "Life-long processes", "Personal development", "Strengths-based education and experiential learning", "Sustainable economy", "Sustainable energy", "Sustainable food", and "Unity and cohesion".

FINDINGS FROM THE TWO METHODOLOGIES

To facilitate the comparison of the two methodologies, i.e. topic modelling and deliberative analysis, it is first useful to have a look at the outcomes that they contribute. The analysis based on topic modeling has been carried out by the authors of this paper, and they did not take part in the deliberative analysis of the Interactive Synthesis Workshop, which was
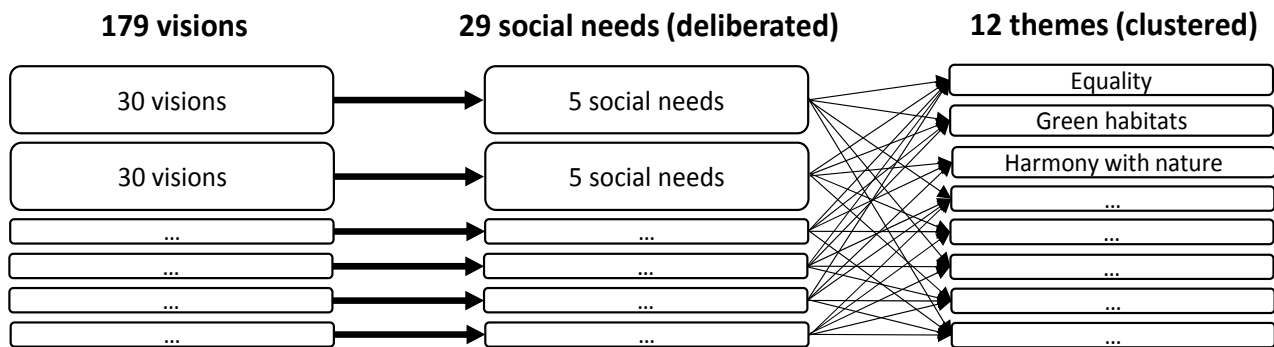
179 visions   **29 social needs (deliberated)**   **12 themes (clustered)**

Figure 1. Process of the CIMULACT Interactive Symbiosis Workshop to the Creation of 12 Themes (See Warnke et al. 2017).

accomplished in the CIMULACT project. If the results of the two analyses conform, topic modeling could be seen as a viable alternative for deliberative analysis (cf. Gläser et al. 2017; Jacobi et al. 2016). Differences, on the other hand, would highlight the particular character of each methodology.

Modeled Topics in Citizen Visions

This section presents results from topic modeling carried out on the data of the citizen visions. The eight topics produced are presented in Table 1 below. The topics were named i.e. labeled to reflect their collocated keywords. Diagnostic data such as word frequencies, exclusivity, and distribution characteristics were used in addition to the keyword list in Table 1. The procedure of labeling involves qualitative interpretation although it is based on results from modeling. A good label depicts the topic well and differentiates it from the other topics, and is meaningful for others to understand.

The weights of the topics (Dirichlet parameter) are presented in the second column of the table, showing the relative prevalence of each topic in the corpus of citizen visions. All topics except for the largest one gain fairly balanced weights depicted by the Dirichlet parameter (range from .247 to .125), which implies that the visions corpus can be well

presented with eight topics.

The largest topic with the most relative weight in the vision corpus concerns "Education for future society" (Dirichlet parameter 2.504). The weight of the topic indicates that the corpus of visions sees education to support the emergence of future society while taking comprehensively into account a great variety of aspects that contribute to desirable life and sustainability. Due to its prevalence in the visions, the topic of education should be considered differently than other topics. It indeed represents more an overall approach to reach to the future than an education concern to be addressed.

The topic "Responsible consumption" relates to taking responsibility of the environment in terms of products, energy, and water. Making ecological choices and applying a longer perspective when doing so relate to this topic.

The topic labeled "Justice for all" addresses citizens' wishes to live in just communities, where differences in cultures are accepted. Transparency and accountability then relate to success and wellbeing on the one hand and trade and carbon neutrality on the other.

"Culture of disease prevention" presents a view of a world were diseases are not only treated but also actively prevented. Health is strived for systematically

Table 1. Eight Modeled Topics in Citizen Visions

Topic	Dirichlet parameter	Keywords in topic
Education for future society	2.504	People education life society work energy vision social development health future system community environment food resources technology citizens children time
Responsible consumption	.247	Energy instance ecological products longer perspective smart produced water consumption locally home choices solar efficient power car responsible focus freedom
Justice for all	.191	Community power sustainability culture valued trade cleaner justice fulfilment accountability international feel success transparency equitable successful wellbeing leading carbon land
Culture of disease prevention	.183	Medical treatment diseases pension prevention data school cancer pharmaceutical european health disease monitoring industry member proper application lifestyle methods increasing
Planet with scientific futures	.151	Science research support treatment transportation future long technology replaced planet improved vision scientific taxes story challenge back space possibilities early
Local citizen collectives	.138	School city parents shared time collective citizen education spaces ecological takes exchange accessible open minimum loss public energies transports introduction
Participatory community	.137	Participatory shared communities community colleagues multinational lifelong grandpa knowledge region mobility farms beings chances conflicts women approach peace freely living
Active families	.125	Family healthy talents creative harmony develop functioning remote things families reduce helps replaced person medicine accordance waiting physical high risk

with medical treatment before people get ill especially in old age. Health monitoring systems are well developed to support healthy lifestyles.

"Planet with scientific futures" corresponds to the idea of technological solutions emerging from scientific research, which globally guides the development of societies. New technologies are expected to challenge and replace old technologies and open new opportunities that are currently unimaginable. These could include new modes of transportation or even space travel.

The topic of "Local citizen collectives" focuses on the development within cities. Cities gather collectives of people and parents save time when they share the responsibilities of taking their children to school in turns, thus contributing also to sustainability. Public places are accessible and open.

"Participatory community" differs from the topic of local citizen collectives in its focus on the community aspect. It is based on a vision of communities participating in the development of their regions. It involves the participatory support of knowledge creation and mobility. Inclusive approaches reduce the danger of conflicts and ensure secure and free living.

"Active families" presents a functional view of society in which families take the responsibility of their lives in relation to health issues and personal development of talents. Families are then also in a position to introduce creative harmony in the society.

These eight topics provide an encompassing view on citizen visions on desirable and sustainable futures. They describe the hopes and dreams of involved citizens and present various levels of abstraction. On the one hand, there are topics that arch over the whole humankind, such as "Planet with scientific futures" or "Justice for all", implying deep changes in the current structures and the functioning of the society. On the other, some topics focus on individuals and their role in the future, such as "Local citizen collectives" or "Active families". Moreover, the results of the topic modeling show one common and a very clear priority

of the European citizens: education. For the emergence of a future society that is desirable and sustainable, education emerges as a key supporting issue across all European countries.

Themes Emerging From Deliberative Analysis

Results from the deliberative analysis are used as a comparative reference point for those of topic modeling, which were presented in the previous sections. In the deliberative analysis, the identification of visions' themes was carried out in an interactive process consisting of several stages. Key observations were made first, then these were elaborated into social needs, which in turn were clustered to the 12 themes. The themes were labeled accordingly, but with words that did not necessarily emerge from the citizens' visions. Instead, the labels are punchy and use descriptive catchwords such as "harmony", "holistic", "life-long", and "strengths-based".

Table 2 presents the results of the deliberative analysis in more detail. The table provides a short description of each of the 12 themes in addition to their names. The short descriptions list the underlying social needs which have been assigned to them and brief descriptions of these needs (Warnke et al. 2017).

The theme labeled "Citizenship awareness and participation" corresponds to education that empowers citizens, provides personal freedom and responsibility, and promotes participation in governance. "Equality" supports such social goals by targeting social justice, accessibility of education, and equal and open access to health care system. "Unity and cohesion" further portrays a society with room for all, community building and development, and education by the community.

"Green habitats", in turn, reviews transport and contrasts rural and urban areas while also accounting for city development. The theme labeled "Harmony with nature" takes the sustainability dimension one step further and considers balance with nature and is accompanied with education for green living.

"Holistic health" pays attention to preventive health care and sees technology to provide healthier living. "Life-long processes" continues with the theme of healthy life and combines it with life-long learning. "Personal development" provides a more individualistic approach to balancing life and meaningful work, while encouraging personal and spiritual development. It also follows up on technology and sees it serve human beings and society. "Strengths-based education and experiential learning" highlights enhancement of quality of education while it acknowledges hands-on applied learning and actionable knowledge learning.

Three themes are labeled directly sustainable: economy, energy, and food. These three themes are further united by that they all correspond to only one social need, as identified in the deliberative process. "Sustainable economy" values economy for wellbeing and sustainability while "Sustainable energy" accentuates the need for sustainable production and consumption of energy. Sustainable food, in turn, highlights equal access as well as a sustainable production and dissemination of healthy food. Altogether, the 12 themes identified in the citizen visions through deliberation provide a good reference point for a comparative review of the two methodologies.

COMPARISON OF FINDINGS AND METHODOLOGIES

A key task of this paper is to compare if the results from topic modeling conform with those from deliberative analysis. As matches between results cannot be complete due to the probabilistic and qualitative characters of the methodologies as well as the richness of the data on citizen visions, the authors first observe parallel topics and themes. Close parallels would indicate that topic modeling is a valid approach to analyze such data on citizen involvement. Then the authors discuss how these two methodologies differ in terms of validity and reliability, abstraction of outcomes,

Table 2. Twelve Themes Emerging From Deliberative Analysis of Citizen Visions

Name of theme	Short description of theme
Citizenship awareness and participation	Education for citizenship (education that supports citizenship, empowers people to fulfill the role of citizen), personal freedom and responsibility (balance between technology and personal values, maximize your happiness), control and influence through participation in governance (effective engagement in decision making, citizen taking responsibility)
Equality	Social justice (basic income, social cohesion, gender equality, equal access to healthcare), accessible education (free education, fair territorial equity), equal and open access to health care system (securing public funding, non-discrimination)
Green habitats	Fast, flexible, affordable, smart, reliable, and clean transport, balanced and fair development of both rural and urban areas (unfair balance of developmental chances leaving the countryside abandoned), enabling systems for green and smart city development (stopping pollution, developing healthier cities)
Harmony with nature	Balance with nature (less harmful activities to preserve life on a livable planet, nature as part of daily life even in cities), education for green living (sustainability education as a transversal concept including life-cycle thinking and global perspective)
Holistic health	Holistic and preventive health care system (to improve the quality of life and life expectancy, help people to take care of themselves), healthier living provided by technology (health by technology, research and develop for health technology, convenience self-treatment and autonomy)
Life-long processes	Healthy life from childhood to old age (long living rely on healthy lifestyles and taking responsibility for one's own health), life-long learning (as a human right aiming to achieve the individual and societal full-potential)
Personal development	Balanced life and meaningful work (realignment between work and personal and community life), personal and spiritual development (encouraging and enabling personal development to grow more skilled, balanced, and happy individual), technology serving human beings and society (technologies which enhance human capabilities)
Strengths-based education and experiential learning	Enhance quality of education (restructuring content and learning context, personalisation), hands-on applied learning and actionable knowledge (learning which develops personal interests and skills, develops awareness about everyday issues, acquires a profession)
Sustainable economy	Economy for wellbeing and sustainability (consumers/modesty and responsibility, work for money/status improve in society)
Sustainable energy	Need for sustainable production and consumption of energy (different energy mixes, self-sustaining regions)
Sustainable food	Sustainably produced, healthy, clean, and responsible food (accessible for all, locally and ethically produced food is safe, healthy, and nutritious)
Unity and cohesion	A society with room for all (diversity with increased interaction and integration), community building and development (solidarity and cooperation, responsibility, social cohesion), education by the community/shared learning (education powered by the community)

Note: Source: Warnke et al. 2017.

cost efficiency and approach to public engagement, and the implications of these differences for the contributions of the two methodologies.

Topics and Themes Match

It is meaningful to compare the results from the topic modeling carried out in this paper with the outcomes of the deliberative analysis in the CIMULACT project

because they both attempt to represent comprehensively the same set of data. At the same time, this comparison functions as a validation test of the topic modeling. If topics and themes find their parallels, the methodology of topic modeling can be considered validated in this context.

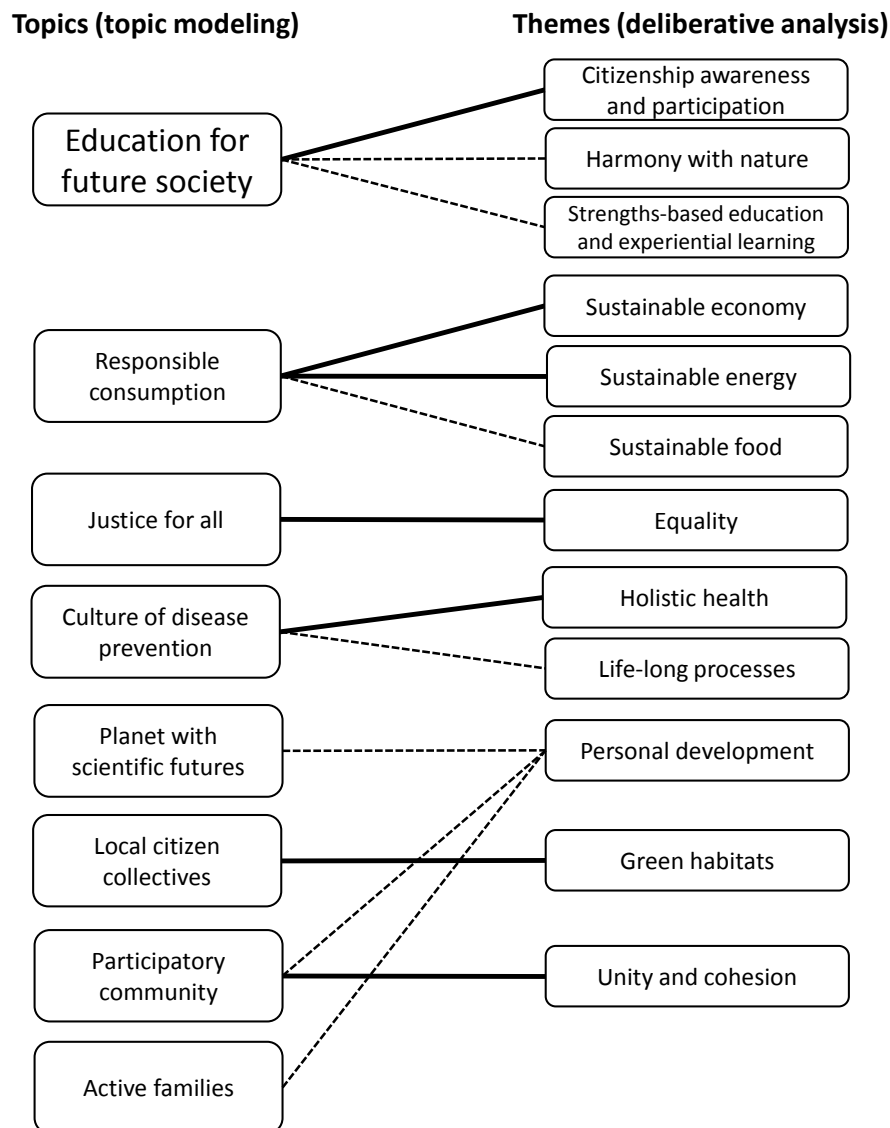Parallels between the topics and the themes are matched in Figure 2. The labels of the topics and their

Topics (topic modeling)   **Themes (deliberative analysis)**

Figure 2. Correspondence of Topics (Topic Modeling) and Themes (Deliberative Analysis) of the Two Methodologies.

Note: Solid lines represent direct parallels and dotted lines represent weaker connections.

keywords as well as the names of the themes and their short descriptions have been used in the comparison. The comparison shows that all topics and themes can be matched with parallels of which many are direct and the remaining are existing albeit weaker. Education, sustainability in the economy, health concerns, and fairness in communities are highlighted in European citizen visions on desirable and sustainable futures.

On the whole, both the topic modeling and the deliberative theme approach provide well-grounded interpretations of the citizen visions. One particular observation needs, nevertheless, to be made. One topic is of quite a different magnitude than the others, which

is not reflected in the deliberative analysis. "Education for future society" indeed dominates the results of topic modeling, and parallels directly to several deliberatively constructed themes of "Citizenship awareness and participation". Weaker parallels can be observed to the themes "Strengths-based education and experiential learning" and "Harmony with nature". "Education for future society" nevertheless most clearly challenges the results of the deliberative analysis due to its prevalence across the corpus. The topic contrasts its approach to the future against several deliberated themes that implicitly include education as an idea but target specific solutions.

In addition:

(1) "Responsible consumption" finds its direct parallels in "Sustainable energy" and "Sustainable economy", while also addressing "Sustainable food" at a weaker level. "Justice for all" finds a direct parallel in the theme of "Equality", highlighting that citizen calls for justice, wellbeing, and accountability.

(2) The topic concerning "Culture of disease prevention" parallels directly with the theme of "Holistic health" and to a lesser degree with "Life-long processes". Disease prevention relates to advances in medical treatment and technologies, although the deliberated themes focus more on people's responsibilities.

(3) Perhaps surprisingly, the topic labeled "Planet with scientific futures" finds only a weak parallel with one theme: "Personal development". The parallel is weak because this theme is not coherent, and only one of its subthemes corresponds to the topic (Technology serving human beings and society). The lack of coherence of the theme "Personal development" becomes clear when considering its weak parallels also to the topics of "Participatory community" and to "Active families", both accounting for human development.

(4) "Local citizen collectives", on the other hand, has a direct parallel in the curiously named theme "Green habitats". This theme relates to transport, rural and urban areas, and smart cities. Naming the theme "green" appears to involve an agenda or a direction towards which to proceed.

(5) "Participatory community" also has a direct parallel in "Unity and cohesion", both accounting for community as an asset in actively contributing to sought targets such as peace, freedom, solidarity, and cooperation.

The comparative analysis shows that six of the eight modeled topics have direct parallel counterparts in deliberated themes. The remaining two modeled topics also have thematic counterparts, but their connection is weaker. Additionally, weaker parallels could be observed also concerning those topics which had direct parallels to deliberated themes.

The observed parallels between the modeled topics and the deliberated themes indicate that both approaches relate well to the data studied, i.e. the citizen visions, and to each other. This observation provides validation of that topic modeling can be used as a methodology to analyze textual corpuses which are produced in public engagement activities of this kind. These parallels also indicate that topic modeling is a valid approach to analyze such a collection of textual data, while accomplishing it even with a smaller number of topics than in the compared deliberative method.

Methodological Implications Vary

Despite similar outcomes between topic modeling and deliberate analysis, there are indeed differences between the two methodologies. Differences can be identified e.g. in terms of processes, approach, and interpretation of results and as they stem from two distinctly different kinds of methodologies, it is worthwhile to consider their particular characters. Table 3 reviews the epistemological characters and practical usability of examined methodologies. The criteria for methodological comparison have been developed from works on public engagement by Kahane et al. (2013), Renn and Schweizer (2009), as well as basic criteria for scientific work.

Table 3. Comparison of Methodological Approaches

	Topic modeling	Deliberative analysis
Validity and reliability	- Results represent analysis of texts with documented methodology - All data are included in easily repeatable analysis - Probabilistic approach leads to similar results	- Results represent contextual interpretation based on expertise of interpreters - Risk to omit relevant data via selection based on evaluation of importance - Results may vary depending on the expertise areas of the participants and interaction in the deliberation
Abstraction of outcomes	- Outcomes reflect public interest - Topical abstractions vary in their character	- Outcomes promote strategic purposes of particular interests - The abstraction level of the themes is similar
Cost efficiency	- One-time moderate investment in methodological competences - Additional narratives need to be developed - Excellent scalability and easy to repeat with low cost	- Significant allocation of resources - Analysis itself contributes to narratives - Low scalability, costly and burdensome to repeat
Approach to public engagement	- Post-modern focus on knowledge formation - Analysis of outcomes non-normative	- Anthropological focus on common sense and consequences - Analysis of outcomes programmatic and agenda-based

Concerning validity and reliability, topic modeling is rooted in the quantitative analysis of texts, and its methodology is readily documented and analyses are easily repeatable. Deliberative analysis as it was carried out in the CIMULACT workshop, in contrast, relies on the expertises of the participants and can further make use of the contextual setting of the public engagement. The contextual understanding as a difference in approach also has effects on how the examined texts are considered and on the outcomes of replications of analyses. Topic modeling itself is carried out on the full corpus and its outcomes thus reflect this corpus in full. Deliberative analysis, in contrast, relies on the expertises of the analysts and thereby carries a risk of omitting relevant data from the analysis. A replicated analysis is unlikely to lead to identical outcomes concerning either methodology, but the outcomes of topic modeling will be similar due to its probabilistic character. Deliberative analysis appears more prone to rely on invited expertises and situational settings. This helps to explain why the dichotomy of urban and rural issues emerges in deliberative analysis while it remains unnoticed in topic modeling.

The methodologies also vary in their abstraction of outcomes. Topic modeling reflects on common, public interests while those of deliberative analysis may be more strategically focused and thereby varied. Yet concerning abstraction, topic modeling provides varying outcomes which range from concrete to very abstract, from actors to targets, and from values to actions. Deliberative analysis, instead, provides similar kinds of abstractions, which highlight values and targets.

There are also differences in the cost efficiency between the two methodologies, which are important to consider when accomplishing public engagement. Topic modeling requires a moderate investment in methodological competences, and does not provide narratives of results during analysis. Deliberative analysis, on the other hand, does not require significant investment in special methodological competences, yet directly provides sought narratives during analysis. However, gathering a large number of

experts to provide deliberative analysis itself requires a significant allocation of resources, and reduces flexibility. Scalability, i.e. how large a corpus can be analyzed, is much greater for topic modeling as computing resources are abundant while resources of experts are costly and scarce.

Finally, both methodologies represent very different approaches to the analysis of outcomes in public engagement. Topic modeling is post-modern in that it focuses on knowledge formation and it is principle non-normative and neutral towards expression of values, for instance. Deliberative analysis as it was carried out in the CIMULACT project, on the other hand, relied on common sense and focused on the consequences of outcomes (see Renn and Schweizer 2009), and is therefore programmatic and agenda-based. Indeed, topic modeling brought forth the roles of individuals in the societies of the future, whereas deliberative analysis accentuated the responsibilities of individuals, thereby providing a more normative interpretation of citizen visions.

In conclusion, although the two sets of topics and themes both relate well to the studied data and to each other, they do indeed represent two distinct approaches to examining data developed in public engagement. Yet, the differences can be considered to represent complementarity rather than contradiction, and thereby do not have significant impact on the next steps in public engagement, i.e. proceeding from analysis of citizen contributions (visions) to early formulations of messages or priorities. Further analysis is necessary to determine which suits better the studied data, and the outcome of this analysis may also relate to preferences on the usability or contextual focus of results from the analysis. Topic modeling could be used to provide descriptive data for succeeding deliberative analysis, if not even replacing deliberative analysis, or through savings to provide better opportunities to enlarge the scope of participants in the deliberative analysis. In the

deliberative analysis as it was carried out in the CIMULACT project, relying on topic modeling could have replaced the need for deliberative analysis altogether, replaced project workers with a greater number of invited experts in the deliberation, or added citizens to the deliberation.

CONCLUSIONS

Large-scale public engagement activities may provide much data which need to be analyzed to be translated to policy action. This paper has compared two methodologies for an analysis of citizen engagement outcomes and reviewed how well the methodology of topic modeling suits this aim. It compared the methodology of topic modeling with the approach of deliberative analysis, and examined the same data on citizen visions on desirable and sustainable futures (Jørgensen and Schøning 2016; Riisgaard et al. 2017). On both accounts, the European prospects on the future are formed by education, sustainability in the economy, health concerns, and fairness in communities.

Topic modeling provided results which rather well adhere to those results that emerged in deliberative analysis. The analysis also confirms that large amounts of citizen generated data can be reviewed quickly and reliably with computer tools (Blei 2012; Gläser et al. 2017; Jacobi et al. 2016). Such analysis further follows citizen contributions closely as it eliminates the need to rely on separate interpretive frameworks. Yet, although results appear parallel, and follow-up steps in public engagement may further unify outcomes, it should be noted that there are notable differences in the epistemological base and practical arrangements between topic modeling and deliberative analysis.

Particular attention should be paid to what the outcomes represent, and to what kinds of narratives or policy messages the outcomes contribute to (Cobb 2012; Smallman 2017). Topic modeling appears to

suit a more neutral and descriptive approach than deliberative analysis, which enables to translate analysis of outcomes towards programmatic agendas. The latter balances to its part the significant allocation of resources, which easily relates to gathering large numbers of experts for deliberation. When the outcomes of public engagement are used for policy aims, such as the steering of research and innovation programs, the translation nevertheless needs to be made carefully.

Further work needs to be carried out to determine how the methodology of topic modeling could best support public engagement activities. If successfully implemented, it could strengthen or replace deliberative analysis as well as save economic resources, which could be used to extend the scope or depth of engagement itself. Some types of large-scale engagement activities may even depend on methodologies such as topic modeling. The CIMULACT project was successful in engaging over 1,000 citizens in 30 countries, but if massive open online public engagement is to become commonplace, it requires appropriate methodologies for data analysis.

Appendix 1

Full description of citizen vision from the Netherlands (Riisgaard et al. 2017) is as follows:

Title of vision: Energy

Description of our vision for 2050

This vision is about self-sufficiency. In 2050 we are much less dependent on fossil fuel. All kinds of energy resources we did not use in 2016 are being used and every house has a solar power installation. Energy is being generated in diverse and creative ways, e.g. from physical exercising also at home or waste that cannot be recycled is converted into energy. Education is the basis of the community and should have a lifelong perspective in order to increase acceptance and involvement towards innovations.

Education about nature and environment learns (teaches) people about sustainable energy. The public transport is attractive and cheap and mostly underground. As a consequence of this all, the health levels have risen because there is much less air pollution. Every house has its own water circulation system (with filter). Roofs are used to store water and energy. People eat more conscious: organic, less meat, seasonal products, insects, etc. Nature is preserved. Robot and sensor technology are part of this vision.

In your desirable future, what is different from today?

• Self-supporting.

• Solar energy.

• Use of smart sources, resources, substances and materials.

• Use of nature: subterranean storage, for instance for water, toilets, energy.

• Every house has its own energy supply and storage (with the chance to exchange energy) and water recycling system.

• Energy storage is legally arranged.

Is your vision desirable for everyone? Are there any concerns?

The wish is that it should be desirable for everyone (within Europe).

What is your vision in brief?

Energy independence.

Exchange of self-supplying energy.

Stimulate innovation.

Acknowledgements

Note

1. Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom.

References

Bedsted, B., J. Ibsen-Jensen, E. Kloppenborg, B. Kyhn, M. Kaarakainen, K. Matschoss, and P. Repo. 2016. *European Citizens' Visions for a Sustainable EU Future. Research Priorities and Policy Advice*. Deliverable 3.3 of the CASI project. Retrieved September 12, 2017 (http://www.casi2020.eu/app/web1/files/download/casi-d3-3-european-citizens-visions-for-sustainable-eu-future.pdf).

Blei, D. M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55(4):77-84.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3(Jan):993-1022.

Bowser, A. and L. Shanley. 2013. *New Visions in Citizen Science*. Case study series, vol. 3. Woodrow Wilson International Center for Scholars, Washington, D.C.

Burgess, J. and J. Chilvers. 2006. "Upping the Ante: A Conceptual Framework for Designing and Evaluating Participatory Technology Assessments." *Science and Public Policy* 33(10):713-728.

CIVISTI (Citizen Visions on Science, Technology and Innovation). 2011. *Periodic Report Summary*. Retrieved September 12, 2017 (http://cordis.europa.eu/result/rcn/45954_en.html).

Cobb, M. D. 2012. "Deliberative Fears Citizen Deliberation About Science in a National Consensus Conference." Pp. 115-122 in *Public Engagement and Emerging Technologies*, edited by K. O'Doherty and E. Einsiedel. Vancouver, Toronto: UBC Press.

Coleman, S., and J. Gotze. 2001. *Bowling Together: Online Public Engagement in Policy Deliberation*. London: Hansard Society.

Corbin, J. and Strauss, A. 1990. "Grounded Theory Research: Procedures, Canons, and Evaluative Criteria." *Qualitative Sociology* 13(1):3-21.

Ducci, G. 2013. "Digital Public Communication in Europe for a European Public Sphere." *Sociology Study* 3(6):425-436.

Gläser, J., W. Glänzel, and A. Scharnhorst. 2017. "Same Data—Different Results? Towards a Comparative Approach to the Identification of Thematic Structures in Science." *Scientometrics* 111(2):981-998.

Graham, S., S. Weingart, and I. Milligan. 2012. *Getting Started With Topic Modeling and MALLET*. Retrieved September 12, 2017 (http://programminghistorian.org/lessons/topic-modeling-and-mallet).

Gudowsky, N., W. Peissl, M. Sotoudeh, and U. Bechtold. 2012. "Forward-Looking Activities: Incorporating Citizens' Visions. A Critical Analysis of the CIVISTI Method." *Poiesis & Praxis* 9:101-123.

Jacobi, C., W. van Atteveldt, and K. Welbers. 2016. "Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling." *Digital Journalism* 4(1):89-106.

Janasik, N., T. Honkela, and H. Bruun. 2009. "Text Mining in Qualitative Research: Application of an Unsupervised Learning Method." *Organizational Research Methods* 12(3):436-460.

Jasanoff, S. 2003. "Technologies of Humility: Citizens Participation in Governing Science." *Minerva* 41:223-244.

Jørgensen, M.-L. and S. Schøning. 2016. *Vision Catalogue. Encompassing the Visions From All 30 Countries*. CIMULACT project, deliverable 1.3. Retrieved September 12, 2017 (http://www.cimulact.eu/wp-content/uploads/2016/06/D1.3final.pdf).

Kahane, D., K. Loptson, J. Herriman, and M. Hardy. 2013. "Stakeholder and Citizen Roles in Public Deliberation." *Journal of Public Deliberation* 9(2), Article 2.

McCallum, A. K. 2002. *MALLET: A Machine Learning for Language Toolkit*. Retrieved September 12, 2017 (http://mallet.cs.umass.edu).

Niemeyer, S. 2011. "The Emancipatory Effect of Deliberation: Empirical Lessons From Mini-Publics." *Politics & Society* 39(1):103-140.

Nyaga, D. and R. A. Torres. 2015. "The Politics of Cultural Representation." *Sociology Study* 5(9):744-758.

PACITA (Parliaments and Civil Society in Technology Assessment). 2016. *Final Report Summary*. Retrieved September 12, 2017 (http://cordis.europa.eu/result/rcn/177948_fr.html).

Rehurek, R. and P. Sojka. 2010. "Software Framework for Topic Modeling With Large Corpora." In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

Renn, O. and P. J. Schweizer. 2009. "Inclusive Risk Governance: Concepts and Application to Environmental Policy Making." *Environmental Policy and Governance* 19(3):174-185.

Riisgaard, K., S. Schøning, and CIMULACT Consortium Partners. 2017. *Vision Catalogue—Encompassing the Visions From All 30 Countries*. Document originated from D1.3. CIMULACT project. Retrieved (http://www.cimulact.eu).

Smallman, M. 2017. "Science to the Rescue or Contingent Progress? Comparing 10 Years of Public, Expert and Policy Discourses on New and Emerging Science and Technology in the United Kingdom." *Public Understanding of Science*, May 1, pp. 1-19.

Steyvers, M. and T. Griffiths. 2007. "Probabilistic Topic Models." *Handbook of Latent Semantic Analysis* 427(7):424-440.

Stoneman, P., P. Sturgis, and N. Allum. 2013. "Exploring Public Discourses About Emerging Technologies Through Statistical Clustering of Open-Ended Survey Questions." *Public Understanding of Science* 22(7):850-868.

Usman, A. K. 2014. "Analysis of the Story-Line Behind Selected Hausa Proverbs." *Sociological Study* 4(8):673-680.

VOICES. 2015. *Final Report Summary*. Retrieved September 12, 2017 (http://cordis.europa.eu/result/rcn/158173_en. html).

Wallach, H. M., I. Murray, R. Salakhutdinov, and D. Mimno. 2009. "Evaluation Methods for Topic Models." Pp. 1105-1112 in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM.

Warnke, P., A. Meroni, M. Rossi, D. Selloni, and A. M. Ospina Medina. 2017. *First Draft of Social Needs Based Research Scenarios*. Deliverable 2.1 of the CIMULACT project. Retrieved September 12, 2017 (http://www.cimulact.eu/wp-content/uploads/2017/03/CIMULACT-D2.1_final.pdf).

Wilsdon, J. and R. Willis. 2004. *See-Through Science: Why Public Engagement Needs to Move Upstream*. London: Demos.

Yau, C. K., A. Porter, N. Newman, and A. Suominen. 2014. "Clustering Scientific Documents With Topic Modeling." *Scientometrics* 100(3):767-786.

Bios

Petteri Repo, Ph.D. (Econ.), adjunct professor, principal investigator and university researcher, Consumer Society Research Centre, Faculty of Social Sciences, University of Helsinki, Finland; research fields: user driven innovation, consumer studies, policy analysis, textual analysis.

Kaisa Matschoss, Ph.D. (Econ.), adjunct professor, university researcher, Consumer Society Research Centre, Faculty of Social Sciences, University of Helsinki, Finland; research fields: energy issues, innovation and experimentation, sustainability, public engagement.

Päivi Timonen, Ph.D. (Econ.), adjunct professor, director, principal investigator and university researcher, Consumer Society Research Centre, Faculty of Social Sciences, University of Helsinki, Finland; research fields: user driven innovation, everyday studies, consumer values, qualitative analysis.