A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

> *"If we have data, let's look at data. If all we have are opinions, let's go with mine."*
>
> Jim Barksdale
> Former Netscape CEO

This study applies topic modelling analysis on a corpus of 86 publications in the *Technology Innovation Management Review* (TIM Review) to understand how the phenomenon of living labs has been approached in the recent innovation management literature. Although the analysis is performed on a corpus collected from only one journal, the TIM Review has published the largest number of special issues on living labs to date, thus it reflects the advancement of the area in the scholarly literature. According to the analysis, research approaches to living labs can be categorized under seven broad topics: 1) Design, 2) Ecosystem, 3) City, 4) University, 5) Innovation, 6) User, and 7) Living lab. Moreover, each topic includes a set of characteristic subtopics. A trend analysis suggests that the emphasis of research on living labs is moving away from a conceptual focus on what living labs are and who is involved in their ecosystems to practical applications of how to design and manage living labs, their processes, and participants, especially users, as key stakeholders and in novel application areas such as the urban city context.

Introduction

Living labs are one of the most prominent and growing areas within the popular open innovation paradigm. The potential of living labs has been acknowledged globally because they offer a fruitful architecture for deploying open innovation through user involvement and co-creation mechanisms (Nyström et al., 2014). The European Network of Living Labs (ENoLL) has over 150 active living lab members worldwide, but the organization has recognized more than 400 living labs since its inception in 2006. While some members eventually disengage from operations for one reason or another, each call for members brings a new "wave" of applicants from around the world. Simultaneously, there has been a parallel wave of increasing scholarly research on living labs.

Along with seven special issues (2012, 2013a, 2013b, 2015, 2016, 2017a, 2017b) and numerous additional articles on living labs in other issues, "living labs" has been a prominent theme in the TIM Review (McPhee et al., 2017a). Those contributions have been of relevance to scholars and practitioners of collaborative innovation, and the journal has played a considerable role in the transformative debate on living labs (Steen & van Bueren, 2017). That said, after all these years, the literature still remains sparse in terms of guidance on how to establish a living lab and how to run and manage it to create value for its stakeholders. Innovation in living labs builds on exposing participants to real-world problems and "understanding, learning, and sharing among the involved stakeholders" (Ståhlbröst & Holst, 2017), but those launching and running living labs often have to learn their lessons the hard way, and every new study on living labs is a valuable and helpful source of information.

Despite the remaining gaps, there have been numerous scholars working on the area for over a decade, and research on living labs has become increasingly fine-grained (Leminen et al., 2017). At the same time, living labs is a conceptually challenging and multifaceted area. Some fundamental aspects of living labs remain

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

dispersed, and there is not one commonly accepted definition of "living lab". This issue is highlighted by Steen and van Bueren (2017), who reviewed over 30 articles on living labs published in the TIM Review in order to compare how different authors have defined living labs, and by the fact that the ENoLL website actually provides three definitions for living labs (enoll.org/about-us/). According to ENoLL's primary definition, living labs are "user-centred, open innovation ecosystems based on systematic user co-creation approach, integrating research and innovation processes in real life communities and settings". The view is in concordance with, for example, Leminen, Westerlund, and Nyström (2012), who define living labs as "physical regions or virtual realities in which stakeholders form public-private-people partnerships (4Ps) of firms, public agencies, universities, institutes, and users all collaborating for creation, prototyping, validating, and testing of new technologies, services, products, and systems in real-life contexts."

The increasing number of studies on living labs is beneficial to scholars and practitioners, but it also brings about challenges in terms of understanding the key research streams in the area. The more research that gets published, the more information there is to be read. Fortunately, novel technologies in the era of big data and machine learning provide opportunities to examine large corpora of text in easy and convenient ways. Text mining techniques can be used to extract knowledge from unstructured or semi-structured textual data, and they have widespread applications in analyzing and processing textual documents. Such text analytics enable the discovery of previously unknown information by automatically extracting information from various written resources (Moreno & Redondo, 2016). Further, combining textual mining techniques with bibliometric analysis helps us discover more unseen patterns in research fields than with simple bibliometric analysis alone (Nie & Sun, 2017).

One of the most efficient text mining techniques is topic modelling, and it is gaining popularity among scholars in diverse fields (Alghamdi & Alfalqi, 2015). Here, we use topic modelling to analyze a corpus of 86 publications on living labs published in the TIM Review from 2011 to 2017. The analysis provides us with key topics in living labs research and their trends over the examined period of six years, which is comparable with McLoughlin and colleagues (2018), who utilized other bibliometric analysis techniques and datasets to understand topics and trends in living labs.

The article is structured as follows. First, we discuss the essentials of topic modelling. Then, we explain the method, including the data and tool(s) used for the analysis. Thereafter, we provide the results. The article concludes with a summary and discussion of contributions of our results to research on living labs, as well as limitations of the current study and future research avenues.

Topic Modelling

In today's era of booming interest in big data analytics by scholars and businesses, topic modelling provides a convenient way to analyze big unclassified text (Alghamdi & Alfalqi, 2015). Since topic modelling was first proposed, it has received a lot of attention and gained widespread interest among researchers in many research fields (Liu et al., 2016). Put shortly, topic modelling is a text-mining technique for discovering topics in documents (Blei, 2012). A topic contains a cluster of words that frequently occur together, and topic modelling can connect words that have similar meanings and can distinguish between uses of words with multiple meanings (Alghamdi & Alfalqi, 2015). Given that text documents are composed of words, a topic covered in more than one document can be expressed by a combination of strongly related words, and any given document can be associated with more than one topic (Jeong et al., forthcoming). Thus, topic modelling is a technique that can be used to infer hidden topics in a collection of text documents (Jeong et al., forthcoming). According to McPhee and co-authors (2017a), the two key outputs from generating a topic model on a collection of documents are: 1) a list of topics (i.e., groups of words that frequently occur together) and 2) lists of the documents that are strongly associated with each of the topics. Ideally, each topic should be distinguishable from other topics.

There are multiple techniques and algorithms that can be used when data mining text documents. Among them, Latent Dirichlet Allocation (LDA) has gained popularity, as it is known to have the highest performance among several topic modelling algorithms when dealing with large-scale documents and interpreting identified latent topics (Jeong et al., forthcoming). LDA was introduced by Blei, Ng, and Jordan (2003) as a generative probabilistic model for collections of discrete data such as text corpora; in particular, it was described as a three-level hierarchical Bayesian model. LDA-based topic modelling is a useful and increasingly applied technique for latent topic identification from a

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

large corpus (Jeong et al., forthcoming). It can be applied to text corpora comprising scholarly papers, and because it uses machine learning and has no critical presumptions on the meanings of the words, it works with texts in any discipline. For example, Nie and Sun (2017) used LDA-based topic modelling to identify research trends in design; Amado and colleagues (2018) applied it to analyze research trends on big data in marketing; and Antons, Kleer, and Salge (2016) used the technique to identify topics published in an innovation management journal over a period of three decades. As a probabilistic method, it works particularly well with large corpora; Sehra and co-authors (2017) analyzed a corpus of 1178 articles to identify research trends in software effort estimation, and Mathew, Agrawal, and Menzies (forthcoming) analyzed over 35,000 papers from software conferences.

Method

We performed the analysis using the J-Tool application for topic modelling developed by Carleton University. The J-Tool is an in-browser application that allows researchers to quickly and easily perform LDA-based topic modelling analyses on TIM Review articles or other textual corpora. The TIM Review (timreview.ca) is a monthly scholarly publication focused on technology innovation and entrepreneurship, and the publications in each issue typically revolve around a special theme introduced by the editor or guest editor(s). With the J-Tool, the researcher can analyze textual data using topic modelling, visually explore relationships of these papers, examine topic trends over time, examine author contributions, and export citations in selected formats. The J-Tool is based on open source components and, although still in a development stage, it can help scholars and students of technology innovation management to perform text data mining analyses on topics relevant to the discipline, as well as current and prospective authors of the TIM Review to gain understanding of the published research in the journal.

The first research article focused on living labs in the TIM Review was published in the October 2011 issue of the journal. Thus, we performed the analysis using the following settings:

1. Coverage: TIM Review issues and publications from October 2011 through October 2017

2. Key term(s): "living lab"

3. Document types: all document types, including articles, editorials, TIM Lecture reports, and Q&A short communication articles

4. Number of topics: provide an output of seven topics (which is the default setting in the J-Tool)

5. Topic threshold: apply a topic threshold of 30% to maximize the number of publications to be included in the topic modelling analysis

Whereas topic threshold value sets a floor for the given topic's proportions in retrieved documents, setting a lower threshold is considered useful when maximal recall is desired (Talley et al., 2011), such as in our study. That said, we also performed the analysis using 70% topic threshold, which results in fewer works involved in the analysis. The topics seemed fairly similar to those resulting from using a 30% threshold, suggesting that a lower topic threshold would not cause significant bias in the results. Finally, we used the J-Tool's default settings for stop words (i.e., common words such as "a" or "the" and domain-specific words such as "issue" or "editorial" that are to be ignored because they do not relate to the subject matter specifically), and we opted for a visualization of the results that explicates bridging articles (articles that connect multiple topics), big topics (topics that are a compound of articles as variables), and variable link lengths (reflecting loadings of articles to topics).

In summary, we included all issues of the TIM Review since the first issue in late 2011 until late 2017, the endpoint reflecting the version of the publication database connected with the J-Tool topic modelling tool that we used. The overall data covering six years of the TIM Review comprised more than 70 journal issues with almost 450 publications. In particular, the data comprised seven special issues devoted to living labs. As a result of using "living lab" as a key term, we obtained a corpus of 86 publications that were analyzed using the J-Tool topic modelling tool. The resulting corpus included 54 research articles; the remaining publications were editorials, public lecture reports, and short communications. We decided to include all types of publications because they may put forward interesting perspectives on the topics that may be otherwise underrepresented in the data. Further, including more publications in the corpus was expected to improve the results due to the probabilistic nature of LDA-based topic modelling. According to the J-Tool development team, the corpus analyzed using the method should include a minimum of 30 publications.

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

Results

According to the J-Tool settings we specified, the topic modelling analysis on the corpus of 86 publications provided seven topics that were associated with living labs. The number of topics is a user-specified parameter where larger values will produce finer-grained, more detailed topics whereas smaller values will produce coarser-grained, more general topics. There is no single value that is appropriate in all situations and all datasets (Barua et al., 2012). Thus, researchers need to choose the desired number of topics based on a calculation of the optimal number of topics (cf. Jeong et al., forthcoming) or based on the interpretability of the solution, or both. Interpretability plays a particularly significant role, as topic modelling may also result in "junk topics", which are uninterpretable topics that pick out idiosyncratic word combinations in the corpus (AlSumait et al., 2009). Following the example of Barua, Thomas, and Hassan (2012), we tried the analysis with a various number of topics until we concluded that seven topics provided a solution that is easy to interpret and provides meaningful characterization and results. For example, an eight-topic solution provided similar results, but the additional topic was overlapping with another topic – a situation that Schmiedel, Müller, and vom Brocke (2018) recommend avoiding. Out of the 86 publications, 51 (59%) were single-topic publications and 35 (41%) were multi-topic publications; the latter we consider as "bridging articles" because they represent a link between topics. Table 1 shows the seven topics and their associated keywords; these keywords are listed in a decreasing order of relevance to each topic, and the J-Tool uses the first keywords (i.e., the words that have the highest relevance to each topic) to label each topic.

In Table 1, certain keywords appear under several topics; for example, the keyword "innovation" is listed under the topics of Innovation, Living lab, Ecosystem, and City. Given that the output of the analysis provided by the J-Tool not only provides automatically generated labels for the topics but also details of the relevance of words to topics versus other topics (i.e., relation strength), we ensured that the provided labels are representative of the topics and that there are no overlapping topics that cannot be distinguished from the others.

In the following subsections, we discuss our subjective interpretations of each topic in the light of the articles that fall under that topic. We followed the guidance of Maier and colleagues (2018), who report that researchers often read through a sample of documents associated with a given topic in addition to the interpretation and labelling of the topic based on its top word(s). Hence, we drilled into the publications associated with the topics and, in particular, read the titles and abstracts of publications associated with the topics. Whenever in doubt, we also browsed the substantive contents. Of note, we were familiar with many of the included publications because we served as guest editors in their associated special issues. Thus, we were able to obtain insight of why the topic modelling tool may have associated a given publication with the specific topic, as well as subjectively identify subtopics under the topic. Put differently, drilling into the publications associated with the seven topics allowed us to better understand what each topic is about and how the subtopics comprise the topic when put together. After discussing the seven topics, we briefly report findings related to bridging articles.

Table 1. The seven identified topics and their associated keywords

Topic	Keywords
1. Design	design, system, user, work, group, technology, study, care, context, phase
2. Ecosystem	ecosystem, university, business, region, local, creation, global, innovation, tourism, research
3. City	city, urban, innovation, citizen, public, smart, collaborative, platform, development, sector
4. University	university, business, technology, innovation, entrepreneur, management, Canada, open, source, global, program
5. Innovation	innovation, process, company, user, knowledge, service, model, business, project, research
6. User	user, research, field, test, trial, living lab, panel, context, end, project
7. Living lab	living lab, innovation, project, user, network, study, case, activity, process, research

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

Topic 1: Design

The first topic focuses on the design approach to living labs with three easily identifiable subtopics: 1) design-driven approaches to living labs (e.g., Brankaert & Ouden, 2017), 2) design as a driver of innovation (e.g., Gray et al., 2014), and 3) design of living lab methodology and processes (e.g., Mulder, 2012). In general, design is a term defined and understood in different ways by different people. In our data, design-driven approach refers to adding design characteristics such as exploration and dealing with uncertainty in order to help living labs to better deal with complex problems. It also considers who the stakeholders are that should be involved in the living lab. Design as a driver is an approach that argues that design is a key factor in innovation and highlights the importance of users as co-designers. Design of living lab methodology and processes is a view that argues living labs are methodologies that need to be designed in a way that they provide experiences to participants, and thus become living and sustainable.

Topic 2: Ecosystem

The second topic focuses on the ecosystems approach to living labs, and it has three subtopics: 1) regional innovation ecosystems (e.g., Viitanen, 2016), 2) global innovation ecosystems (e.g., Seppä, 2012), and 3) open innovation ecosystems (e.g., León & Martinez, 2016). Regional innovation ecosystems view regions as areas whose innovation performance can be improved through collaborative initiatives such as living labs and the related ecosystem. Conversely, global innovation ecosystems discusses a new era of knowledge institutions building globally distributed living lab ecosystems to meet global innovation needs. Finally, the open innovation ecosystem focuses on the challenges and opportunities for the tertiary educational sector to partake in novel open innovation ecosystems such as living labs. Common to all of these subtopics is that the described ecosystems include or build around regional universities.

Topic 3: City

The third topic clearly focuses on the use of living labs in the urban city context. The topic has three subtopics: 1) cities as collaborative innovation platforms (e.g., Tukiainen et al., 2015), 2) urban living labs (e.g., Juujärvi & Lund, 2016), and 3) smart city development (e.g., Khomsi, 2016). Cities as collaborative innovation platforms focuses on the role of the city in innovation and the uses of living labs for collaborative innovation. Typically, the idea is to develop the city and improve the lives of its residents, businesses, public sector organizations, and others such as tourists. A very specific goal for a city's development through living labs is that of becoming a smart city.

Topic 4: University

This topic builds upon a large number of editorials of special issues related to living labs and other collaborative forms of innovation. Moreover, it includes other types of university-driven activities such as public lectures on innovation and entrepreneurship organized by the university. Although some editorials only briefly mention living labs, for example, because of announcing a forthcoming issue on living labs, other editorials discuss them in more detail. Three main subtopics in the corpus are: 1) research advances on living labs (e.g., McPhee et al., 2017b), 2) universities as knowledge mobilization platforms for innovation (e.g., McPhee, 2016), and 3) entrepreneurial practice and experiences using collaborative innovation (e.g., McPhee, 2014). In this respect, the topic does not introduce a specific perspective to living labs, but likely reflects the university's key role in knowledge dissemination of research related to living labs and other collaborative innovation to scholarly and practitioner communities. That said, it should be noted that TIM Review editorials commonly introduce authors and their institutions and, hence, the term "university" comes up multiple times in each editorial. While this fact supports the view of universities and their scholars as disseminators of knowledge related to living labs, it also weakens the interpretability and validity of this topic.

Topic 5: Innovation

The fifth topic addresses the use of living labs by companies and other organizations for innovation. The main subtopics are: 1) the challenges of using living labs (e.g., Westerlund & Leminen, 2011), 2) the benefits of using living labs (e.g., Niitamo et al., 2012), 3) the perceived experiences of using living labs (e.g., Ståhlbröst, 2013), and 4) the management of living labs (e.g., Katzy & Bücker, 2015). In essence, these subtopics describe the motivation and expected value of getting involved in innovation through living labs, as well as the management and coordination challenges of conventional development projects versus the open innovation model. Management is discussed in terms of activities and processes ensuring innovation performance and economic sustainability of the living lab.

Topic 6: User

This topic focuses on the quintessential role of users in living lab experiments. The subtopics within the topic are: 1) managing user involvement in living labs (e.g.,

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

Schuurman & De Marez, 2012), 2) understanding user dropouts in living lab experiments (e.g., Georges et al., 2016), and 3) the real-life environment of user as a context for innovation (e.g., Coorevits & Jacobs, 2017). The first and second subtopics discuss how to increase the likelihood that a user will remain involved in innovation throughout multiple phases of the process. In addition, real-life environment refers to the intense user–system interaction in a real-life context as a key to successful user input during the process.

Topic 7: Living lab
Finally, the seventh topic examines the living lab itself to understand what living labs and their defining characteristics are. The subtopics are focused on providing taxonomies, typologies, and categorizations: 1) the characteristics of living labs (e.g., Steen & van Bueren, 2017), 2) living labs as a type of user innovation methodology (e.g., Almirall et al., 2012), 3) living labs as open innovation networks (e.g., Leminen et al., 2012), 4) categorization of innovation tools in living labs (e.g., Leminen & Westerlund, 2017), and 5) classification of the environment in living labs (e.g., Bergvall-Kåreborn et al., 2015). Characteristics of living labs can be used to categorize true living labs from improperly labelled living labs. Living labs, as a type of user innovation methodology, categorize different user innovation methodologies and explain how living labs differ from other methodologies. The network view considers living labs as networks to categorize different living labs based on their driving actors. Categorization of innovation tools suggests a new typology of living labs based on their innovation process characteristics and usage of tools. Finally, classification of the environment views living labs as places and spaces where innovation happens.

Bridging articles
As mentioned previously, more than 40% of included publications were multi-topic publications, or "bridging articles". The analysis revealed 11 combinations of topics; most of them were a combination of two topics, and one was a combination of three. It turned out that 7 out of these 11 combinations included Innovation as one of the topics. In fact, Innovation was connected with all the other topics and such combinations covered 77% of the bridging articles. This is not surprising given that the topic of Innovation was found to discuss fundamentals of using living labs for collaborative innovation. Another non-surprising topic that showed up in various combinations was that of University. Again, this is expected given that the majority of publications falling under this topic were editorials or other non-research article publications emphasizing the role of university in disseminating knowledge. However, we consider the possibility that most of the authors mentioned in the editorials were academics and therefore the word "university" often appeared in the editorials. That said, the role of university is visible even in the basic definition of living labs, which emphasizes them as public–private–people partnerships (cf. Leminen et al., 2012); in fact, universities typically represent a key public sector participant.

Figure 1 visualizes how the seven topics and their associated publications are constructed and interlinked. The large nodes in the illustration are topics and each small node is a publication; those small nodes that are connected to only one topic are single topic publications reflective of that specific topic, whereas those that connect two or more topics are bridging articles. The J-Tool allows the researcher to easily identify any of the articles by simply hovering the mouse pointer on nodes. Further details are then given in an output table that helps to report the results. The TIM Review is represented by the centre of the illustration because it is what connects the topics. As said previously, we also opted for variable length links to reflect article to topic loadings in the illustration; however, a brief investigation did not reveal anything interesting in terms of those nodes that have short links versus those that have long links. Thus, we decided to leave a more detailed analysis of variable link lengths outside of the scope of this article.

Trend Analysis

Similar to McPhee and co-authors (2017a), we use the degree of association of documents to a topic over time to reflect overall trends in topics. Figure 2 shows the overall trends of the seven identified living lab topics in the TIM Review from October 2011 through October 2017. Of note, although the vertical axis (i.e., relative strength of association) does not show the count of words or articles, it does reflect the popularity of the topics in the journal issues and publications over the examined six years period of time. Further, although an analysis of statistical significance would add credibility to the trend analysis (cf. Choi et al., 2017), the output did not provide accurate, usable data for such calculations. However, the output enabled us to estimate values to calculate an increase index as suggested by Sun and Yin (2017). Drawing on their idea of increase index formulation, we used estimates of the document-to-topic relevance from the first two years (2011–2012) and the last two years (2016–2017) of the examined period

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

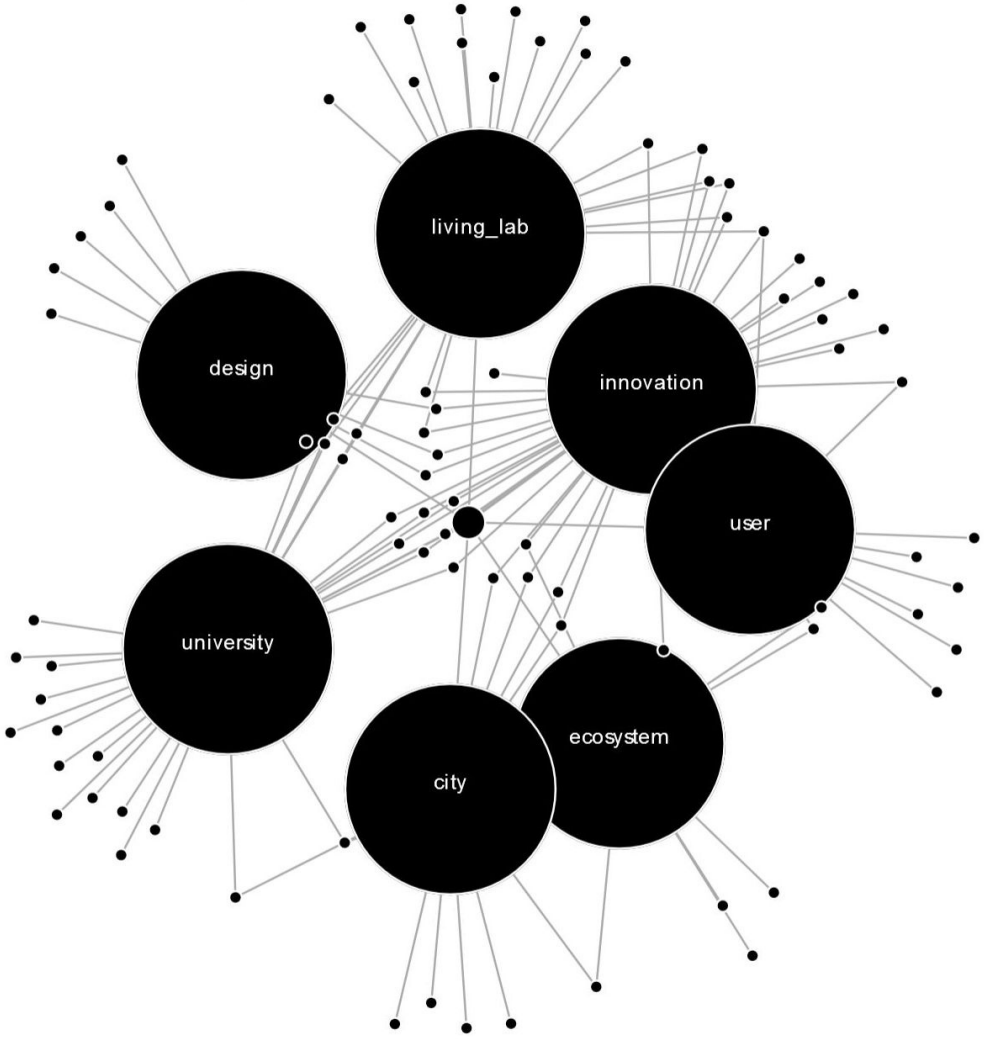

Figure 1. Visualization of topics and their connections

in order to create an index. In short, our increase index r^k shows the difference in the document-to-topic relevance for each topic between two time windows, and it provides numerical support for visual interpretation of the trends shown in Figure 2. An index value in excess of 1.00 reflects an upward trend and lower than 1.00 indicates a downward trend. Further, values of 1.00 ± 0.05 suggest a stagnant trend; this accepted margin of error is especially due to potential inaccuracies in estimated values.

As shown in Figure 2, there are three kinds of trends related to the identified topics: decreasing, increasing, and constant. We report these trends together with their increase index values (r^k). To start with, University (r^k=0.47) and Innovation (r^k=0.73) are decreasing trends in terms of popularity. In particular, the decreasing trend of University is evident as its relative strength (reflecting popularity) halved in six years. Innovation faced a slightly smaller decrease, losing a third of its relative strength during the years. Then again, User (r^k=5.80), City (r^k=2.25), and Design (r^k=1.88) are increasing trends. Interestingly, the trend reflecting the popularity of User in connection with living labs shows the highest growth. Whereas the relative strength for the topic User was almost non-existent and clearly lowest of the seven in late 2011, it had reached the third-highest rank by late 2017. Both City and Design doubled in terms of relative strength during the period. Finally, Living lab (r^k=0.97) and Ecosystem (r^k=1.00) seem to be constant trends showing little to no changes over the six-year period. That said, the relative strength of the Living lab topic is high throughout the period, which is not surprising given the topic's foundational nature.

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

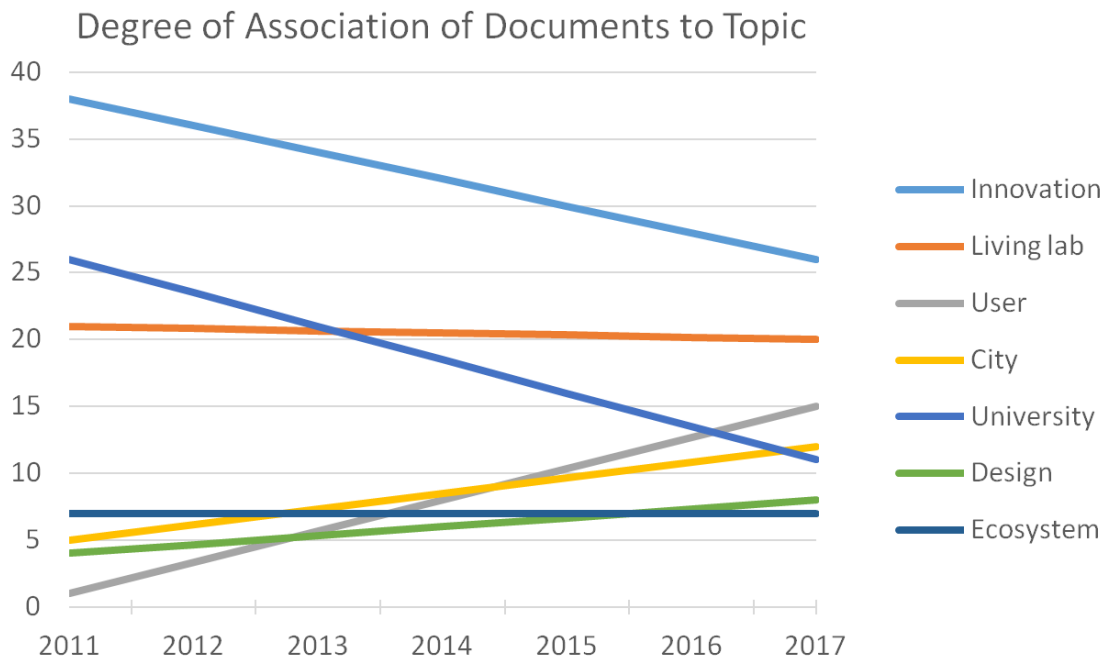

Figure 2. Overall trends of living lab topics in the TIM Review during 2011–2017

Discussion

This study applied topic modelling analysis on a corpus of publications in the TIM Review published from 2011 to 2017 to examine how authors have approached living labs in the recent innovation management literature. To our knowledge, the TIM Review has published the largest number of special issues focused on living labs to date; thus, we considered that it can reflect the progress of the scholarly research in regard to living labs. Further, we drilled into the topics and examined the titles and contents of the articles that were associated with each topic. In this vein, our study combined textual mining techniques and bibliometric analysis to discover unseen patterns in a specific research field, as suggested by Nie and Sun (2017).

We found that research related to living labs in the TIM Review can be categorized under seven broad topics: 1) Design, 2) Ecosystem, 3) City, 4) University, 5) Innovation, 6) User, and 7) Living lab. In addition, each topic includes various subtopics that, when put together, reflect the topic in a comprehensive way. Out of the seven identified topics, Ecosystem, University, Innovation, and Living lab reflect broader, more conceptual approaches to the phenomenon of living labs. These topics are essentially focusing on what living labs are by definition, who are being involved in the operations, and what the benefits of living labs are both in a broad sense and in specific. Conversely, Design, City, and User represent a more applied approach to living labs. In other words, these topics discuss how living labs can be designed and managed to overcome various challenges, how users as key participants should be handled, and how living labs can be applied to urban contexts in order to create value to stakeholders. Of note, our results are in concordance with those of McLoughlin and co-authors (2018), who performed various bibliometric analyses on datasets comprising publications with a conceptual or methodological focus on living labs. Although their study included articles from various disciplines and outlets (mostly computer and information science as well as engineering publications), their analysis revealed fairly similar topics, with "smart city" emerging among the most prevalent topics in terms of popularity and maturity. Our analysis using a different method and dataset also put the urban/smart city context among the top topics. In this vein, our results gain support from recent research applying bibliometric approaches in order to understand the scholarly field of living labs.

Interestingly, the trend analysis we performed on the topics suggested that the research emphasis in living

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

labs seems to be moving away from conceptual understanding of what living labs are, what types of living labs there are, and who are being involved in living lab ecosystems to practical applications of how to design and manage living labs and their participants, especially users as key stakeholders, in novel application areas such as the urban city context. We believe that, despite the yet emerging literature on living labs, there is nowadays a sufficient number of conceptually oriented studies on the fundamentals of living labs in order for scholars to move on to a deeper, more practically oriented level of research. In addition, there may be more empirical data for practice-oriented research available as the phenomenon of living labs matures. An aspect supporting the view that the fundamentals of living labs are quite well discussed in the literature is our finding of a large number of bridging articles that connect two or more topics. Almost all of the bridging articles combined either Innovation or University with another topic; the two being fundamental topics when we think of what living labs are and how they are defined. Our results add another dimension to the findings by McLoughlin and co-authors (2018), who investigated trends in living lab research, including a number of pre-2011 publications, and found that there has been a shift from a technology focus to a social focus in the application and context of living labs.

Furthermore, we believe that another reason driving the shift from conceptually oriented studies to more practically oriented studies on living labs is that there is a growing demand for practically oriented studies that can help newcomers in the field, namely inexperienced innovation and living lab managers seeking to build new living labs that are arguably a complex form of collaborative innovation. There are few practical guidelines on how to create and grow a living lab or how to manage its crucial processes. Newcomers to the field need advice, best practices, and lessons learned from others about what to do in order to maximize the success of their initiative and speed up innovation. That said, there is an increasing pool of knowledge and expertise accumulating, and this knowledge can be turned into research outputs. In particular, interesting research opportunities are arising as some living labs that have been operating for a long time are ending their operations, merging into new forms of collaborative innovation (cf. Claudel, 2018; Leminen et al., 2017; Steen & van Bueren, 2017), or establishing financial mechanisms to support their operations after the initial funding dries out.

Limitations and Future Research

Every research project has limitations. In this study, there are several limitations that may affect the generalizability of the results. First, the articles used in the analysis may represent a specific perspective to living labs. We only analyzed works published in one journal, namely the TIM Review. Although the journal has earned a strong reputation as a scholarly peer-reviewed journal that has published the largest number of special issues focused on living labs as of 2018, its focus on innovation management and the fact that many of the special issues were developed from papers published in innovation management conferences may have affected the approaches and views taken by the authors of the studied publications. It is possible that certain scholars of livings labs who represent another academic discipline beyond innovation management, such as sociology, may have produced interesting insights on the topic that are not represented in the TIM Review.

Second, the examined timeframe does not include foundational studies from the early period of living lab research, because the first article on living labs was only published in the TIM Review in late 2011. Until late 2011, the journal operated under another name and focused on the business aspects of open source software. However, a number of foundational articles were published on living labs prior to 2011 in other scholarly outlets, and including them in the study would enrich the results. In comparison, the bibliometric analysis of living lab literature by McLoughlin and co-authors (2018) comprised numerous pre-2011 publications, including some foundational papers. However, due to the utilized filtering criteria, they ended up with a narrowed-down dataset comprising mainly studies in computer and information science as well as engineering.

Third, the examined data included a relatively small number of papers. LDA-based topic modelling is considered an effective method for analyzing textual corpora. However, due to the probabilistic nature of the method, results from the analysis are likely to be better and more reflective of the data when applied to larger corpora. Hence, previous research providing bibliometric analyses on given disciplines has applied topic modelling to textual corpora consisting of publications in the range of a thousand to tens of thousands. Our data only included 86 papers. Reviewing and analyzing 86 papers is a major effort for a human researcher, and certainly exceeds the minimum requirement for the topic modelling tool; nonetheless, it is still a small number in

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

machine learning based data analytics that was essentially designed for big data.

Fourth, there are various limitations related to the topic modelling tool that we used. For instance, topic models are likely to vary based on the random seed that is required for the analysis. In the J-Tool, the seed is hard-coded in order for it to remain the same over multiple runs, thus allowing for replication of the analysis with similar results. In this vein, it can be considered a strength, because we could run multiple experiments without the fear of facing varied results due to the changes in the seed. On the other hand, if the seed were set differently in the first place, the results might be somewhat different. Moreover, the early version of the tool that we used did not provide the outputs that would add further credibility and accuracy to the analysis. For example, we were unable to calculate statistical significance for the identified trends.

Future research should apply topic modelling over a larger corpus of studies on living labs, potentially including the early (pre-2011) notions of the concept and reaching out to the most recent publications on the phenomenon. Although the early version of the J-Tool that we used was limited to the TIM Review publication database, the version under development can handle almost any article and many other types of data entry inputted into the analysis. In addition, the added features of the newer version provide the researchers with additional output and reporting tools, such as data required for statistical significance calculations, as well as word clouds and various illustrations showing document counts that may illuminate the contents and development of the identified topics in a richer and more accurate manner, and allow for a more descriptive discussion.

Acknowledgements

Funding from FedDev Ontario, through the Innovation Centre at Bayview Yards in Ottawa, Canada, enabled the development of the J-Tool used to perform the topic modelling analysis presented in this article.

About the Authors

Mika Westerlund, DSc (Econ), is an Associate Professor at Carleton University in Ottawa, Canada. He previously held positions as a Postdoctoral Scholar in the Haas School of Business at the University of California Berkeley and in the School of Economics at Aalto University in Helsinki, Finland. Mika earned his doctoral degree in Marketing from the Helsinki School of Economics in Finland. His research interests include open and user innovation, the Internet of Things, business strategy, and management models in high-tech and service-intensive industries.

Seppo Leminen is an Adjunct Professor of Business Development at Aalto University in Helsinki, Finland, and an Adjunct Research Professor at Carleton University in Ottawa, Canada. He holds a doctoral degree in Marketing from the Hanken School of Economics in Finland and a doctoral degree in Industrial Engineering and Management from the School of Science at Aalto University. His research and consulting interests include living labs, open innovation, innovation ecosystems, robotics, the Internet of Things (IoT), as well as management models in high-tech and service-intensive industries. He is serving as an associate editor in the BRQ Business Research Quarterly, on the editorial board of the Journal of Small Business Management, as a member of the Review Board for the Technology Innovation Management Review, and on the Scientific Panel of the International Society for Professional Innovation Management (ISPIM). Prior to his appointment at Aalto University, he worked in the ICT and pulp and paper industries.

Mervi Rajahonka, DSc (Econ), works as an RDI Advisor at the Small Business Center (SBC), currently a part of the South-Eastern Finland University of Applied Sciences XAMK, Finland, and she is an Adjunct Research Professor at Carleton University in Ottawa, Canada. She has been working at the SBC for about 10 years. She earned her doctoral degree in Logistics from the Department of Information and Service Economy at Aalto University School of Business in Helsinki, Finland. She also holds a Master's degree in Technology from the Helsinki University of Technology and a Master's degree in Law from the University of Helsinki. Her research interests include sustainable logistics and supply chain management, business models, service modularity, and service innovations. Her research has been published in a number of journals in the areas of logistics, services, and operations management.

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

References

Alghamdi, R., & Alfalqi, K. 2015. A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications,* 6(1): 147–153.
https://doi.org/10.14569/IJACSA.2015.060121

Almirall, E., Lee, M., & Wareham, J. 2012. Mapping Living Labs in the Landscape of Innovation Methodologies. *Technology Innovation Management Review,* 2(9): 12–18.
http://doi.org/10.22215/timreview/603

AlSumait, L., Barbará, D., Gentle, J., & Domeniconi, C. 2009. Topic Significance Ranking of LDA Generative Models. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD):* 67–82.
https://doi.org/10.1007/978-3-642-04180-8_22

Amado, A., Cortez, P., Rita, P., & Moro, S. 2018. Research Trends on Big Data in Marketing: A Text Mining and Topic Modeling Based Literature Analysis. *European Research on Management and Business Economics,* 24(1): 1–7.
https://doi.org/10.1016/j.iedeen.2017.06.002

Antons, D., Kleer, R., & Salge, T. O. 2016. Mapping the Topic Landscape of JPIM, 1984–2013: In Search of Hidden Structures and Development Trajectories. *Journal of Product Innovation Management,* 33(6): 726–749.
https://doi.org/10.1111/jpim.12300

Barua, A., Thomas, S. W., & Hassan, A. E. 2012. What Are Developers Talking About? An Analysis of Topics and Trends in Stack Overflow. *Empirical Software Engineering,* 19(3): 619–654.
https://doi.org/10.1007/s10664-012-9231-y

Bergvall-Kåreborn, B., Ihlström Eriksson, C., & Ståhlbröst, A. 2015. Places and Spaces within Living Labs. *Technology Innovation Management Review,* 5(12): 37–47.
http://doi.org/10.22215/timreview/951

Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research, 3: 993–1022.

Blei, D. M. 2012. Probabilistic Topic Models. *Communications of the ACM,* 55(4): 77–84.
http://dx.doi.org/10.1145/2133806.2133826

Brankaert, R., & Ouden, E. 2017. The Design-Driven Living Lab: A New Approach to Exploring Solutions to Complex Societal Challenges. *Technology Innovation Management Review,* 7(1): 44–51.
http://doi.org/10.22215/timreview/1049

Choi, H. S., Lee, W. S., & Sohn, S. Y. 2017. Analyzing Research Trends in Personal Information Privacy Using Topic Modeling. *Computers & Security,* 67: 244–253.
https://doi.org/10.1016/j.cose.2017.03.007

Claudel, M. 2018. From Organizations to Organizational Fields: The Evolution of Civic Innovation Ecosystems. *Technology Innovation Management Review,* 8(6): 34–47.
http://doi.org/10.22215/timreview/1163

Coorevits, L., & Jacobs, A. 2017. Taking Real-Life Seriously: An Approach to Decomposing Context Beyond "Environment" in Living Labs. *Technology Innovation Management Review,* 7(1): 26–36.
http://doi.org/10.22215/timreview/1047

Georges, A., Schuurman, D., & Vervoort, K. 2016. Factors Affecting the Attrition of Test Users During Living Lab Field Trials. *Technology Innovation Management Review,* 6(1): 35–44.
http://doi.org/10.22215/timreview/959

Gray, M., Mangyoku, M., Serra, A., Sánchez, L., & Aragall, F. 2014. Integrating Design for All in Living Labs. *Technology Innovation Management Review,* 4(5): 50–59.
http://doi.org/10.22215/timreview/793

Jeong, B., Yoon, J., & Lee, J.-M. Forthcoming. Social Media Mining for Product Planning: A Product Opportunity Mining Approach Based on Topic Modeling and Sentiment Analysis. *International Journal of Information Management.*
https://doi.org/10.1016/j.ijinfomgt.2017.09.009

Juujärvi, S., & Lund, V. 2016. Enhancing Early Innovation in an Urban Living Lab: Lessons from Espoo, Finland. *Technology Innovation Management Review,* 6(1): 17–26.
http://doi.org/10.22215/timreview/957

Katzy, B. R., & Bücker, C. 2015. The Organization of Living Labs: Coordinating Activities for Regional Innovation. *Technology Innovation Management Review,* 5(9): 23–28.
http://doi.org/10.22215/timreview/927

Khomsi, M. 2016. The Smart City Ecosystem as an Innovation Model: Lessons from Montreal. *Technology Innovation Management Review,* 6(11): 26–31.
http://doi.org/10.22215/timreview/1032

Leminen, S., Westerlund, M., & Nyström, A. - G. 2012. Living Labs as Open-Innovation Networks. *Technology Innovation Management Review,* 2(9): 6–11.
http://doi.org/10.22215/timreview/602

Leminen, S., & Westerlund, M. 2017. Categorization of Innovation Tools in Living Labs. *Technology Innovation Management Review,* 7(1): 15–25.
http://doi.org/10.22215/timreview/1046

Leminen, S., Rajahonka, M., & Westerlund, M. 2017. Towards Third-Generation Living Lab Networks in Cities. *Technology Innovation Management Review,* 7(11): 21–35.
http://doi.org/10.22215/timreview/1118

León, G., & Martínez, R. 2016. Q&A. How Can a University Drive an Open Innovation Ecosystem? *Technology Innovation Management Review,* 6(7): 48–51.
http://doi.org/10.22215/timreview/1004

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. 2016. An Overview of Topic Modeling and Its Current Applications in Bioinformatics. *SpringerPlus,* 5(1): 1608.
http://doi.org/10.1186/s40064-016-3252-8

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, A. 2018. Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures,* 12(2-3): 93–118.
https://doi.org/10.1080/19312458.2018.1430754

Mathew, G., Agrawal, A., & Menzies, T. Forthcoming. A Method for Finding Trends in Software Research. *IEEE Transactions in Software Engineering.*
https://arxiv.org/pdf/1608.08100

A Topic Modelling Analysis of Living Labs Research

Mika Westerlund, Seppo Leminen, and Mervi Rajahonka

McLoughlin, S., Maccani, G., Prendergast, D., & Donnellan, B. 2018. *Living Labs: A Bibliometric Analysis.* Paper presented at the 51st Hawaii International Conference on System Sciences (HICSS), January 2–6, 2018, Hawaii, USA.

McPhee, C. 2014. Editorial: Innovation and Entrepreneurship. *Technology Innovation Management Review,* 4(12): 3–5. http://doi.org/10.22215/timreview/852

McPhee, C. 2016. Editorial: Knowledge Mobilization. *Technology Innovation Management Review,* 6(9): 3–3. http://doi.org/10.22215/timreview/1013

McPhee, C., Santonen, T., Shah, A., & Nazari, A. 2017. Reflecting on 10 Years of the TIM Review. *Technology Innovation Management Review,* 7(7): 5–20. http://doi.org/10.22215/timreview/1087

McPhee, C., Schuurman, D., Ballon, P., Leminen, S., & Westerlund, M. 2017b. Editorial: Innovation in Living Labs. *Technology Innovation Management Review,* 7(1): 3–6. http://doi.org/10.22215/timreview/1044

Moreno, A., & Redondo, T. 2016. Text Analytics: The Convergence of Big Data and Artificial Intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence,* 3(6): 57–64. https://doi.org/10.9781/ijimai.2016.369

Mulder, I. 2012. Living Labbing the Rotterdam Way: Co-Creation as an Enabler for Urban Innovation. *Technology Innovation Management Review,* 2(9): 39–43. http://doi.org/10.22215/timreview/607

Nie, B., & Sun, S. 2017. Using Text Mining Techniques to Identify Research Trends: A Case Study of Design Research. *Applied Sciences,* 7(4): 401. https://doi.org/10.3390/app7040401

Niitamo, V.-P., Westerlund, M., & Leminen, S. 2012. A Small-Firm Perspective on the Benefits of Living Labs. *Technology Innovation Management Review,* 2(9): 44–49. http://doi.org/10.22215/timreview/608

Nyström, A.-G., Leminen, S., Westerlund, M., & Kortelainen, M. 2014. Actor Roles and Role Patterns Influencing Innovation in Living Labs. *Industrial Marketing Management,* 43(3): 483–495. https://doi.org/10.1016/j.indmarman.2013.12.016

Schmiedel, T., Müller, O., & vom Brocke, J. Forthcoming. Topic Modeling as a Strategy of Inquiry in Organizational Research: A Tutorial With an Application Example on Organizational Culture. *Organizational Research Methods.* https://doi.org/10.1177/1094428118773858

Schuurman, D., & De Marez, L. 2012. Structuring User Involvement in Panel-Based Living Labs. *Technology Innovation Management Review,* 2(9): 31–38. http://doi.org/10.22215/timreview/606

Sehra, S.K., Brar, Y.S., Kaur, N., & Sehra, S.S. 2017. Research Patterns and Trends in Software Effort Estimation. *Information and Software Technology,* 91: 1–21. https://doi.org/10.1016/j.infsof.2017.06.002

Seppä, M. 2012. From Business Administration to Business Creation: The Case of the Kalevala Global Business Creation School. *Technology Innovation Management Review,* 2(6): 6–11. http://doi.org/10.22215/timreview/562

Steen, K., & van Bueren, E. 2017. The Defining Characteristics of Urban Living Labs. *Technology Innovation Management Review,* 7(7): 21–33. http://doi.org/10.22215/timreview/1088

Ståhlbröst, A. 2013. A Living Lab as a Service: Creating Value for Micro-enterprises through Collaboration and Innovation. *Technology Innovation Management Review,* 3(11): 37–42. http://doi.org/10.22215/timreview/744

Ståhlbröst, A., & Holst, M. 2017. Reflecting on Actions in Living Lab Research. *Technology Innovation Management Review,* 7(2): 27–34. http://doi.org/10.22215/timreview/1055

Sun, L., & Yin, Y. 2017. Discovering Themes and Trends in Transportation Research Using Topic Modeling. *Transportation Research Part C,* 77: 49–66. https://doi.org/10.1016/j.trc.2017.01.013

Talley, E. M., Newman, D., Mimno, D., Herr, B. W., Wallach, H. M., Burns, G. A. P. C., Leenders, A. G. M., & McCallum, A. 2011. Database of NIH Grants Using Machine-Learned Categories and Graphical Clustering. Supplementary Data. *Nature Methods,* 8(6): 443–444. http://doi.org/10.1038/nmeth.1619

Tukiainen, T., Leminen, S., & Westerlund, M. 2015. Cities as Collaborative Innovation Platforms. *Technology Innovation Management Review,* 5(10): 16–23. http://doi.org/10.22215/timreview/933

Viitanen, J. 2016. Profiling Regional Innovation Ecosystems as Functional Collaborative Systems: The Case of Cambridge. *Technology Innovation Management Review,* 6(12): 6–25. http://doi.org/10.22215/timreview/1038

Westerlund, M., & Leminen, S. 2011. Managing the Challenges of Becoming an Open Innovation Company: Experiences from Living Labs. *Technology Innovation Management Review,* 1(1): 19–25. http://doi.org/10.22215/timreview/489

Citation: Westerlund, M., Leminen, S., & Rajahonka, M. 2018. A Topic Modelling Analysis of Living Labs Research. *Technology Innovation Management Review,* 8(7): 40–51. http://doi.org/10.22215/timreview/1170

Keywords: living lab, living laboratory, innovation, topic modeling, topic modelling, big data, text analytics, data mining, research trends

Technology Innovation Management Review

timreview.ca

Academic Affiliations and Funding Acknowledgements

The Federal Economic Development Agency for Southern Ontario (FedDev Ontario; feddevontario.gc.ca) is part of the Innovation, Science and Economic Development portfolio and one of six regional development agencies, each of which helps to address key economic challenges by providing regionally-tailored programs, services, knowledge and expertise.

- *The TIM Review receives partial funding from FedDev Ontario's Investing in Regional Diversification initiative.*

Technology Innovation Management (TIM; timprogram.ca) is an international master's level program at Carleton University in Ottawa, Canada. It leads to a Master of Applied Science (M.A.Sc.) degree, a Master of Engineering (M.Eng.) degree, or a Master of Entrepreneurship (M.Ent.) degree. The objective of this program is to train aspiring entrepreneurs on creating wealth at the early stages of company or opportunity lifecycles.

- *The TIM Review is published in association with and receives partial funding from the TIM program.*