Knowledge gaps in evaluating the effectiveness and impacts of Living Labs focused on environmental and agricultural sustainability

Albana Berberi, Carleton University
Christine Beaudoin, University of Ottawa
Vivian Nguyen, Carleton University
Kelly Bronson, University of Ottawa
Chris McPhee, Agriculture and Agri-Food Canada
Steve Joncoux, University of Quebec
Jean-François Jasmin, Living Lab en innovation ouverte (LLio)
R. Sandra Schillo, University of Ottawa

June 30, 2021

This report draws on research supported by the Social Sciences and Humanities Research Council.

Table of Content

Executive Summary

Background: The issue

Living Labs (LLs) bring users of potential innovations together with stakeholders from the private and public sectors, and they show great potential for bringing forward user-centric solutions to complex environmental issues. In Canada, there are notable examples of agricultural sustainability LLs at the regional AcadieLab (https://www.rang3.org/le-labo) and national scale. In 2018, Agriculture and Agri-Food Canada (AAFC) launched its five-year Living Laboratories Initiative (http://www.agr.gc.ca/livinglab) which is a large-scale application of the LL approach within an agroecosystem context and in April 2021 launched its 10-year Agricultural Climate Solutions program (https://www.agr.gc.ca/agriculturalclimatesolutions), which will further expand this national network using the same living labs approach.

Despite the promise of the LL approach for addressing complex challenges like socio-ecological change, there is a gap in practitioner and academic community knowledge surrounding how to measure and evaluate both the performance of a given LL process and its wider impacts. Notably, this gap appears particularly acute in LLs designed to address environmental or agricultural sustainability.

Objectives

This study addresses the need for measures and evaluation frameworks for LLs with a specific focus on LLs that aim to generate environmental or agricultural sustainability. The knowledge synthesis project uses a mixed-methods approach including scoping literature review (peer-reviewed and grey literature) as well as expert workshops targeting a specific community of LL practitioners focused on sustainability. The project objectives were to:

1. Synthesize best practices for evaluating impacts and effectiveness of LLs, drawing in insights from related multi-stakeholder initiatives as appropriate;
2. Identify knowledge gaps and strengths of LLs focused on environmental or agricultural sustainability in order to develop an agenda for future research and action; and,
3. Build and engage a network of cross-sectoral LL experts, stakeholders and partners interested in LL for sustainability to solidify Canada as a leader in this field.

Results

Our scoping literature review used a combination of human and automated textual analysis. 41 articles were found relevant for this study. The findings reveal limited studies putting forward generalizable approaches or frameworks for evaluating the impact of LLs and even fewer in the agricultural or sustainability sector. Additionally, our expert workshops produced a research agenda on evaluating effectiveness and impacts in LLs. The main research themes include diversity of actors and roles, need to better define and identify methods to measure effectiveness/success, process and different types of impacts at different scales and timelines, challenges in having an overarching method between sectors/contexts and knowing which methods to use for what goal, when to use the methods, and at which scale.

Key messages

This knowledge synthesis project uncovers a potential tension regarding LL work: the specificity of LL studies works against the development of evaluation indicators and a universal framework to guide the impact assessment of LLs across jurisdictions and studies in order to move toward generalizability. By working with experts, the hypothesis that more detailed empirical work is necessary on the evaluation of LLs was validated and knowledge gaps were prioritized to guide future research: better understanding diversity and empowerment of actors in evaluation, lack of clear definitions of the elements of evaluation, better understanding alignment, mismatch and integration across scales and timelines, among others.

Methodology

Our project takes a mixed methods approach composed of two parts: (1) scoping review and (2) Delphi survey and workshop.

The scoping review determined the scope or coverage of the peer-reviewed and grey-literature on evaluating and measuring the impacts of LLs. Article searches were conducted on Web of Science, Scopus, Google Scholar, and Google using the following search string: *("living lab*" AND (evaluat* OR performance OR effective* OR impact OR assess* OR metric OR measure* OR indicator)*. LL specialists were also asked to provide relevant articles. Search results were exported to Covidence (www.covidence.org) where duplicates were merged, and the total remaining set of search results were screened for relevance. Search results were screened with eligibility criteria at two subsequent phases: (1) Title and Abstract and (2) Full Text. Researchers underwent a consistency check during each screening phase. Exclusion criteria included: exclude on irrelevant LL definition, exclude on evaluation (i.e., article does not discuss LL evaluation), and exclude on effectiveness (i.e., article does not discuss LL effectiveness). After screening, data was extracted using Voyant Tools online website. Texts from articles were extracted verbatim into a summary table. After extraction, quantitative data such as frequency and percentage were calculated using Excel while qualitative data was categorized under prominent themes which emerged across the dataset after reading the full text.

Surveys and workshops were conducted in English and French, and mobilized knowledge from the scoping review. Online surveys prioritized LL research themes that were later used to structure workshop discussions. The workshop primarily focused on the Canadian LL context, while inviting select world-renowned experts in LLs. Within this workshop, we borrowed ideas from Delphi methodologies, scaled down to appropriateness for this project. In the initial research stage, participants outlined their current thinking about strengths and knowledge gaps of LLs that address environmental and agricultural sustainability. During the middle stage, participants underwent exercises to think and reflect on how they understood the processes and impacts of these LLs and what was generally known about evaluation. In the final stage, participants identified research questions and key themes. A finalized LL research agenda with compiled research themes was distributed to workshop participants for feedback and validation.

Background

The magnitude of environmental challenges requires the involvement of a diversity of stakeholders from public and private sectors as well as civil society in order to develop socially and economically effective sustainability practices. LLs bring users together with stakeholders from the private and public sectors, and they show great potential for bringing forward user-centric solutions and innovations for solving complex environmental issues, for example generating climate change adaptation, and more sustainable natural resource management (Hossain et al. 2019). LLs build off of the insight, now well accepted in environmental studies, that transdisciplinary and participatory approaches are most effective for addressing complex problems, specifically when dealing with transitions to sustainable, resilient, and adaptive societies (Pohl 2005, Boon et al. 2014). Transdisciplinary research is especially useful in studying mechanisms and institutions that work towards people's participation in environmental decision-making (Lam et al. 2014).

LLs involve a variety of stakeholders – including users and sometimes citizens – in the exploration, co-creation, testing and evaluation of innovations within real-world environments (Almirall et al. 2012, Ballon et al. 2015, Ballon et al. 2018, Leminen & Westerlund 2019). LLs promote alignment of process, social, technological innovation with user practices and preferences. LL arrangements stimulate cooperation among stakeholders on difficult-to-implement practices such as environmentally friendly processes for reducing social and economic impact or collective governance and experimentation to address sustainability (Voytenko et al. 2016, Hossain et al. 2019). LLs help elucidate for their participants the connection between environmental and human systems, including potential for great potential for academic, public, private and non-for-profit sectors to apply the LLs approach to environmental and agricultural sustainability issues such as generating behavioural innovation in environmental practices (Hossain et al. 2019).

In various forms, LLs have gained traction since the early 2000s. For example, the European Network of Living Labs (ENoLL; https://enoll.org/) has more than 150 active LL member organizations worldwide, while recognizing more than 475 members historically, including active members in 20 of the 28 European Union Member States, and it is present in five continents in addition to Europe. Branching out from their original focus on the information management and technology sector, living labs are now being used to improve practices and technologies in a broad range of sectors including health care, urban planning, application design, service delivery and information management and technology. ENoLL's network reflects this trend, including an increasing proportion of members engaged in rural and urban activities (notably including a trend toward agriculture-focused living labs and sustainability). With an increase of 250% since 2016, these types of living labs now make up more than 10% of the network's membership (ENoLL, pers. comm.). In identifying the defining characteristics of "agroecosystem living labs", McPhee et al. (2021) have highlighted the evolution of the living lab approach in the agricultural context, including case studies from France and Canada.

Indeed, Canada has been a leader in applying the LL approach to sustainability, specifically in the domains of agriculture and regional innovation, with notable examples being AcadieLab

(https://www.rang3.org/le- labo) and LLio (http://llio.quebec/) in Quebec. The approach is now applied at a nationwide scale through Agriculture and Agri-Food Canada's (AAFC) Living Laboratories Initiative (www.agr.gc.ca/livinglab) and the new Agricultural Climate Solutions program (https://www.agr.gc.ca/agriculturalclimatesolutions), which was launched in 2021. Both of these programs use a common "agroecosystem living labs" approach and network (McPhee et al., 2021). However, despite extensive scientific coverage and capacity for biophysical performance indicators in these types of living labs, there is a need to develop and implement methods for evaluating the impacts and performance of this approach from a social sciences perspective. As such, LLs provide a practical and timely context for studying and developing methods to measure environmental behavioural change.

Objectives

Despite the growing literature on and implementation of LLs, what makes an effective LL vis-à-vis policy or sustainability practice goals remains underexplored (Ballon et al. 2018, Hossain et al. 2019). A gap exists in our understanding of how to evaluate and measure the social processes that influence the effective functioning of a LL as well as their longer-term impacts (e.g., social capital, effectiveness of collaborations, transdisciplinary governance structure, barriers to adoption, pro-environmental behaviours). Given AAFC's new network of agroecosystem LL sites across Canada, there is an urgent need to understand how to evaluate the effectiveness of LLs and, ideally, identify the best practice drivers of "effective and impactful LLs" in the form of procedural or structural characteristics. A second gap relates to the fact that LLs on environmental and agricultural sustainability are underexplored; thus, understanding and identifying research questions related to these LLs remains important.

This study addresses the need for measures and evaluation frameworks for LLs with a specific focus on LLs that aim to generate environmental or agricultural sustainability. The knowledge synthesis project used a mixed-methods approach including scoping literature review (peer-reviewed and grey literature) as well as expert surveys and workshops targeting a specific community of LL practitioners focused on sustainability.

The project objectives were to:

1. Synthesize best practices for evaluating impacts and effectiveness of LLs, drawing in insights from related multi-stakeholder initiatives as appropriate;
2. Identify knowledge gaps and strengths of LLs focused on environmental or agricultural sustainability in order to develop an agenda for future research and action; and,
3. Build and engage a network of cross-sectoral LL experts, stakeholders and partners interested in LL for sustainability to solidify Canada as a leader in this field.

The guiding research questions were:

- How are the value, performance and success of LLs conceptualized?
- What general evaluation metrics are most appropriate to ensure the effectiveness of LLs?
- What research and knowledge are required to evaluate LLs, generally, and what ones are specific to environmental and agricultural sustainability?

The project aimed to increase visibility and leverage LL approaches for environmental and agricultural sustainability, in Canada and globally. This is important as LLs for sustainability remain underexplored (Hossain et al. 2019). The fact that Canada is leading nation-wide agroecosystem LLs initiatives serves as a platform and opportunity to understand its effectiveness on sustainable agricultural practices, and how it may serve as a model for other environmental and natural resource sectors. Our project provides insight on how to evaluate the social processes underlying the functioning of LLs and the long-term social impacts of both the processes and outcomes of LL innovations. Our unique collaborators and workshops increase the visibility of LLs for sustainability and gather a network of LL experts and partners interested in using LL approaches for sustainability.

Methods

Our project takes a mixed methods approach composed of two parts: (1) scoping review and (2) Delphi surveys and workshops. The scoping review determined the scope or coverage of the peer-reviewed and grey-literature on evaluating and measuring the impacts of LLs. The Delphi surveys and workshops aimed to set a future research agenda for evaluating and understanding impacts of LLs for sustainability, as well as mobilized knowledge from the scoping review.

Scoping Review Methods

To answer our research question, we used an adapted scoping review method of available peer-reviewed literature (Tricco et al. 2018). We developed an a-priori protocol to outline the objective of the scoping review, article search process, article screening process, eligibility criteria, and data extraction. It is important to note that the study did not undergo a critical appraisal or meta-analysis of included articles as these steps are not mandatory for a scoping review (Munn et. al 2018).

The PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist was used for the initial screening. Additionally, we conducted text mining using an automated computer tool called Voyant (https://voyant-tools.org/). Voyant Tools is an open-source, web-based application designed for text mining (Miller 2018), which was developed by Stéfan Sinclair (McGill) and Geoffrey Rockwell (University of Alberta) (Sampsel 2018). Voyant tools is considered one of the costs and time effective ways of analyzing qualitative data quantitatively because it provides a quick interpretation and visualization patterns in the data which then demand for further qualitative analysis (Miller 2018, Hetenyi 2019).

Search terms and database

First, based upon an initial scan of the living lab literature, search terms and search strings were identified for the three following concepts relevant to our research questions: “Living Lab”; “evaluation”; “Agroecosystem/Environment”. Web of Science (WoS) and Scopus were used as a database. Our aim was to search broadly rather than within specific disciplines and these two databases are two of the world leading databases for multidisciplinary academic articles. Additionally, they contain both natural and social science articles and they are known to pull a subset of highly ranked social science journals. Trial searches were performed on Web of Science (Core Collection) and Scopus databases. These searches were continued through an iterative process until a comprehensive search string was developed (Table 1). There are two notable components of the finalized search string. First, Concept 1 only included “living lab*” as a search term; this is because other synonymous Concept 1 search terms yielded a wider range of irrelevant articles when searched independently from “living lab*”. Moreover, at closer inspection, these related Concept 1 search terms were found in relevant articles when searching only with “living lab*”. Therefore, related Concept 1 search terms were redundant and not included in the finalized search strategy. Second, Concept 3 was omitted from the finalized search strategy because it yielded few numbers of articles at the Title and Abstract screening

phase. We employed this term later in the Full Text screening process. The final search string use for the first phase of literature review was: *("living lab*" AND (evaluat* OR performance OR effective* OR impact OR assess* OR metric OR measure* OR indicator)*. This string was run on the 2nd of June 2020 on both Scopus and WoS. Scopus generated 946 references and WoS generated 591 references. The total is 1101 references including 5 articles from snowball search, of which 411 were duplicates and excluded.

Table 1

Finalized Search String Using Boolean Operators

Concepts	Search Terms
(1) LLs	"living lab*"
	AND
(2) Evaluation	(evaluat* OR performance OR effectiv* OR impact OR assess* OR metric OR measure* OR indicator)

Note. The asterisk (*) represents a wildcard that allows for any character(s) to replace it (e.g., evaluat* includes evaluate, evaluates, evaluation, etc.).

There were no restrictions placed on publication year, and only search results in English and French were considered. Details on the sources and total numbers of articles included in this study are provided in Figure 1. A team of four research assistants were involved in the search process while two academic researchers were involved in verification of the articles.

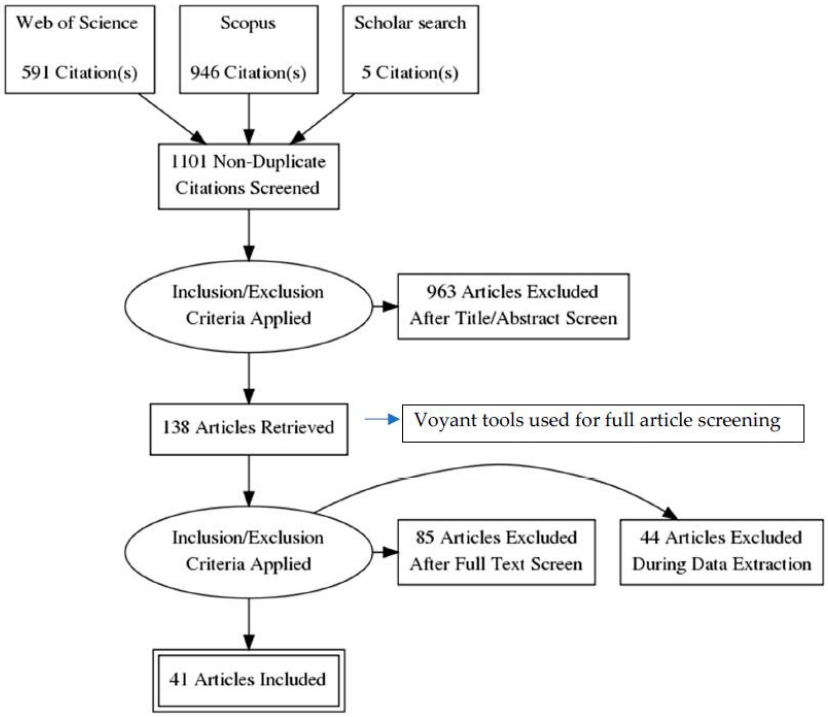

Figure 1. Flow diagram of scoping review stages using PRIMSA-ScR checklist in this study.

Comprehensiveness of searches

Deliberation within our team of researchers and benchmark papers were used to test the comprehensiveness and validity of the search strategy. Selected benchmark papers include academic literature that were mostly published in Journals, books, and Proceedings (see Appendix A).

Article screening and study eligibility criteria

Screening process

Search results were exported to Covidence (www.covidence.org) where duplicates were merged, and the total remaining set of search results were screened for relevance. Search results were screened with eligibility criteria at two subsequent phases: (1) Title and Abstract and (2) Full Text. Articles that posed uncertainty were categorized as “Include for Second Opinion” and were assessed by the research team until a final decision was made on inclusion/exclusion. In total, 138 articles were screened in their full text version.

Consistency check

Before both the screening phases, we carried out a consistency check on 5% of the total articles, selected at random to ensure the consistency of screening across reviewers. Article selection for the consistency check was done through a double-blind method and each article was screened by each reviewer. A Kappa test was used to assess the inter-reliability of

screening outcomes, and inconsistencies were reconciled by the research team (McHugh, 2012).

Eligibility criteria

Articles were screened for inclusion or exclusion using eligibility criteria at each phase, as outlined in Appendix B. At full text screening, the following specific exclusion criteria were introduced: exclude on LL definition, exclude on evaluation (i.e., article does not discuss LL evaluation), and exclude on effectiveness (i.e., article does not discuss LL effectiveness) (see Appendix B). Articles screened for Include Second Opinion were further deliberated by the research team for either inclusion/exclusion.

Data extraction and analysis

After Full Text screening, 138 articles were extracted as relevant, meaning they focus on measurement and evaluation within LLs. These articles were further screened using Voyant Tools to get to a reasonable number of articles for data extraction. A corpus was created in Voyant Tools online website by uploading the 138 full text articles and using unique search terms for this text mining. The major search terms used in this process were: agri*, sustainabl*, evaluat*, impact*. Thus, screening the articles we initially deemed relevant for those which use terms relevant to our research question. The Voyant Tools analysis resulted in 88 articles which were manually screened according to the above eligibility criteria (Appendix B). All of these articles were reviewed in full. After applying the eligibility criteria, 44 articles were excluded from data extraction. Finally, only 41 articles were found highly relevant for this study and useful for extraction—these are articles which specifically focus on evaluation and impact assessment. Texts from articles were extracted verbatim into a summary table. After extraction, quantitative data such as frequency and percentage were calculated by using excel while qualitative data were categorized under prominent themes which emerged across the dataset after reading the full text.

Delphi Survey and Workshops Methods

In parallel with the scoping review, a survey followed by workshop discussions inspired by the Delphi method aimed to co-construct a research agenda. This research agenda addressed knowledge gaps from the following question: How can we measure the effectiveness and social impacts of "Living Labs" in agriculture and the environment? This whole process was conducted online and took place between November 2020 and July 2021.

What is the Delphi method?

According to Linstone and Turoff (2002), the Delphi method is a predictive, systematic and interactive method that relies on a panel of experts. The experts answer surveys in two or more rounds. After each round, a facilitator leads an exchange of opinions by providing an anonymous summary of the thoughts and projections of the experts from the previous round. The experts are then encouraged to consider different points of view and revise their previous responses in light of the responses from other panel members. It is anticipated that during this

process, the diversity of responses will decrease, making the converging and diverging positions more clearly evident.

Participants

Participants were selected on the basis of their expertise in 3 main areas: LLs, agro-environment and collaborative social science, and evaluation. Researchers and practitioners of LLs in the agro-environment field who spoke English or French fluently were particularly targeted.

Participants were recruited in a “call for participation” action through direct and indirect contact in the Canadian and European Networks of Living Labs (ENOLL). Participants met the following main criteria and guidelines for the surveys, workshops, and research agenda validation stages, respectively.

Survey:

- Experts and/or researchers in both agriculture and environment Living Labs and Evaluation processes.
- Experts and/or researchers in agriculture and environment Living labs process with strong Knowledge on Evaluation processes.
- Experts and/or researchers in Evaluation practices with a strong knowledge of agriculture and environment Living Labs processes.

Workshop:

- Participants from the survey.
- Researchers, practitioners, and experts evaluating LL processes in the field of agricultural and environmental Living Labs.

Research agenda validation:

- Participants from the survey and/or workshop.

Delphi survey

An initial list of 89 participants were identified, of which the online surveys (hosted on Survey Monkey in both French and English) were sent to 34 participants who were willing to engage in the process . Of these, we received responses from 23 participants after sending 3 reminders. The survey closed on February 21st, 2021.

The development of the survey was informed by preliminary data from the Bronson et al. (2021) scoping review. The analysis themes of the scoping review have been taken up and completed from one of the authors' previous work (Joncoux & Lewis, 2019) in order to cover all the building blocks of an evaluation process (temporality, methods, objects, actors, etc.). We asked mixed closed- and open-ended questions about areas of expertise and involvement in LLs, effectiveness and social impacts of LLs, and questions about specific themes related to the evaluation of LLs, such as how participants would rank the themes in relation to each other as

well as if references and/or knowledge gaps exist for each theme (see Appendix C for survey questions).

Open questions about knowledge gaps were imported into Nvivo for inductive, thematic coding. Nodes were grouped to identify major themes. Participants' ranking of themes was used to generate a list of top, medium, and low priority themes. Responses to a closed question to identify if references exist in the literature were counted. Participants also provided references for each theme, they were counted and used to generate a list of references. The surveys prepared participants with background thinking and reflections before identifying research questions at the workshop.

Delphi workshop

Two 3-hour virtual workshops were held, one in French and one in English, simultaneously using Zoom (video conference) and Mural.co (digital collaborative board) platforms. Mural.co allows workshop participants to see movements, post-it notes, and typing in real-time and thus contributes to an advanced process of co-creation through the cross-pollination between participants. The workshop limited participants to 25 to ensure optimal and manageable performance according to the number of facilitators mobilized that day. The workshop was designed with three main activities: convergent prioritization activity, divergent deepening of the prioritized themes, and a fictional case study. The exercises were designed to allow participants to first come to consensus on priority research gaps and questions, and second, to offer opportunities to diverge and explore with relatively few boundaries.

For the first exercise of convergent prioritization of research gaps and questions, participants were divided based on diversity of backgrounds and expertise into breakout rooms of 5-6 people. Participants were presented with ranked research themes that emerged from the survey and given three votes each. The objective was to test the strength of the prioritization established in the previous stage against a larger and more diverse panel. Once the votes were casted on what they felt were most important (importance in relation to the need for further study of the theme), the teams discussed and agreed on the top six themes featuring the most important research gaps. The next step within this activity was a divergent individual exercise where participants were asked to develop aspects that remain to be explored for each of the top six themes and to formulate as many proposals as possible in the form of research questions. This research question brainstorm culminated in a dialogue elaborating and exploring these questions. The first stage concluded in a plenary speed presentation of key discussion points to the entire group.

The second stage of the workshop consisted of discussing a fictional case study grounded in reality. A short text was presented to the participants, outlining the key elements of a fictional living lab inspired by real cases from the AAFC living lab network. The fictional case was built around the objective of co-developing innovative and beneficial management practices that would help maintain year-round plant cover to reduce erosion and maintain the physical, chemical and biological health of the soil. The same groups of 5 to 6 people were then asked: "If

you were in charge of evaluating the effectiveness and social impacts of this particular Living Lab what would you do?" The main objective was to bring out new research questions from a different perspective (more operational than theoretical) and to explore the evaluation tools and methods of reference for the participants.

Research agenda

Research questions formulated by participants in the workshops were extracted per research theme from the Mural.co platform. Emergent priority themes were identified by comparing top rated themes in the workshop breakout groups and by rewording themes to account for participant views of interconnections between themes. A research agenda was developed that included research questions and themes generated in the workshops, and further synthesized by the research team.

The research agenda was distributed to workshop participants for feedback and validation in the third and final Delphi stage. Participants were invited to contribute further to the conversation in an anonymous way by entering comments for each of the questions and their descriptions. The main objective was to validate the wording of the questions and make general or transversal comments. Participants also had the opportunity to see how synthesized questions for the research agenda and their descriptions were formulated from raw questions shared by participants in the workshops.

Results

Scoping Review Results

Overview of results

This scoping review resulted in 41 articles that are relevant to measuring the impact of LLs (see Appendix D for a reference list for included articles). The majority of these articles are journal articles including peer reviewed articles (56%) followed by the proceedings/conference papers (34%) with very few from book/book chapters (1%) (Figure 2). A sectoral analysis further shows that the publications come from LLs focused on diverse sectors and they are largely studies based on more than one LL project (Figure 2). However, the analysis also shows that social innovation is the major focus of the LLs studies which assess the impact of LLs (Figure 2).

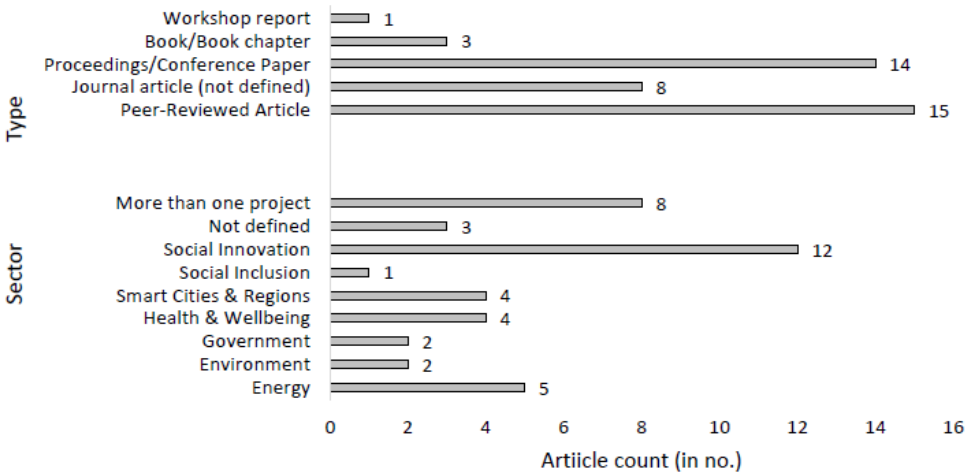

Figure 2. Type and sector details on articles in Living Labs (LLs) literature (in no.).

It is evident from our review that there is limited published work discussing evaluations of the impact of LLs, and relevant articles included here only emerge from 2009 and after. Furthermore, most of our articles are from Europe or focused on LL based out of Europe (51%) (Figure 3). The most comprehensive LL project is arguably ENoLL (https://enoll.org/), a network of LLs which has expanded across Europe, rising from 20 to over 440 LLs in between 2007 to 2020. We feel this may explain the dominant presence of European publications in this sector (Hossain et al. 2019, Leminen & Westerlund 2019, Gamache et al. 2020).

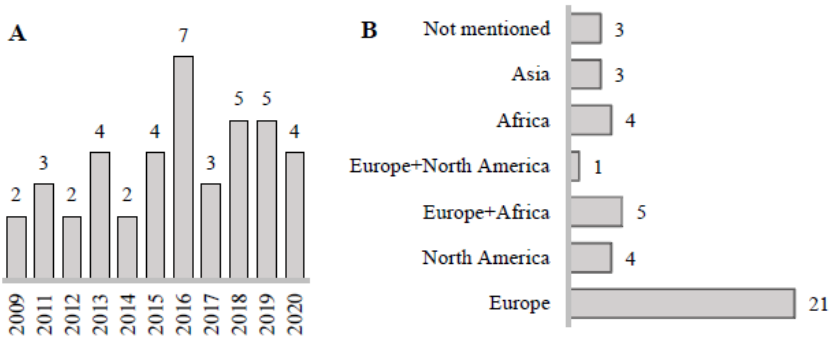

Figure 3. The total article included in the scoping review process. Figure (A) shows the total articles published (in no.) by year and (B) shows the countries involved in publishing articles on Living Labs.

Evaluation assessment methods adopted within the LL literature

Our study confirms that there is currently a gap in the academic literature on how to measure LL efficacy across contexts; additionally, there are limited existing studies on measuring efficacy among those LLs focused on agriculture or sustainability. Among our dataset environmental issues are discussed almost entirely in relation to new technology (digital technology) that furthers environmental goals—for example “green” energy technologies (Gamache 2020).

Few studies on LLs and sustainability are available (Gamache 2020, Bulkeley et al. 2016, Voytenko et al. 2016, Keyson et al. 2013, Burbridge 2017, Puerari et al. 2017, Schliwa et al. 2015) and even fewer of these focused on measurement/impact (Voytenko et al. 2016, Ondiek & Moturi, 2019). In our scoping review, 30 percent (41 out of 138 full text screening) articles were found relevant to measuring impact. However, only four of 41 articles (screened as relevant for evaluation of LL) focused on agriculture and sustainability and yet these studies did not focus on measuring impact. It is obvious from the literature that agroecosystem LLs are a recent phenomenon which does appear to connect with what is happening outside the academy. For instance, the international Agroecosystems Living Laboratories (ALL) working group was formed at the 2018 G20 Meeting of Agricultural Chief Scientists (MACS) in Argentina, Co-chaired by Canada (Agriculture and Agri-Food Canada, AAFC) and the United States (U.S. Department of Agriculture, USDA) (see https://www.macs-g20.org/). A major and recent initiative in Europe is Agrilink, which established six living laboratories (in Italy, Norway, Latvia, Spain, Romania, The Netherlands and Belgium) supported by Horizon 2020 research and innovation programme (https://www.agrilink2020.eu/). The screening processes-both word count and term search of the full article databases/corpus (138 articles) also showed that few articles contained the term “agri*, eval*, sustainab*, and impact*” (Figures 4 and 5).

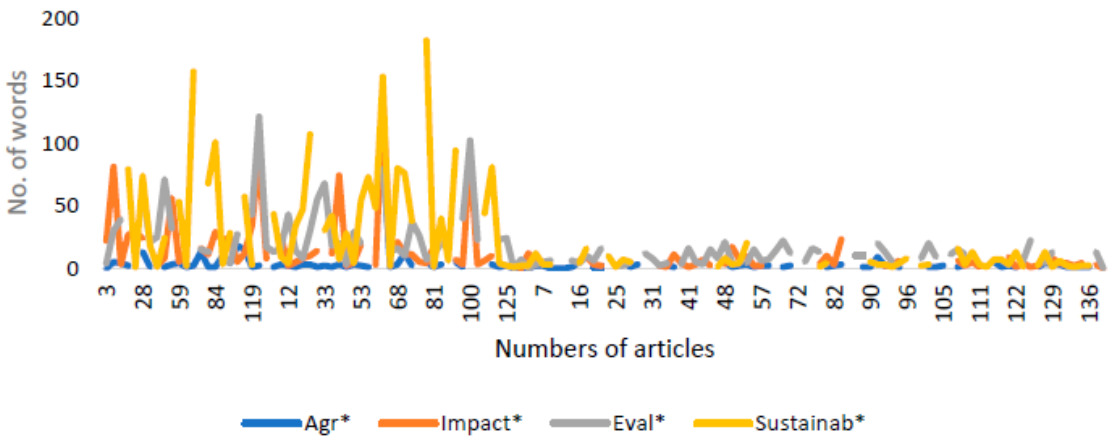

Figure 4. Frequency of terms in articles included in this review. Figure generated from Excel with manual counting.

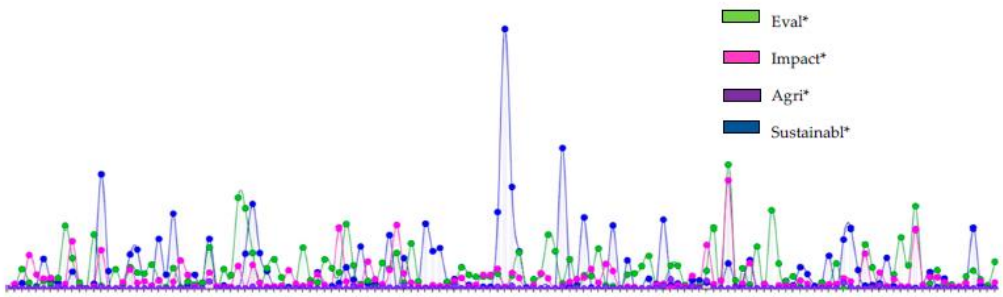

Figure 5. Frequency of words (Eval*, Agri*, Impact* and Sustainabl* in 138 articles included in this study. Figure generated using Voyant.

Our scoping review shows that case studies and qualitative methods of data collection (and among that semi-structured interviews and workshops) were more common methods used in the evaluation of LLs. This might be due to the fact that the LL is considered a novel approach to innovation, and qualitative methods are found to be more relevant for this kind of emergent research. Quantitative methods did appear in the literature but were more common in assessing LL focused on technology development and the technology adoption. The figure below (Figure 7) shows the snapshot of methods used in evaluating LLs among our datasets. The most common methods used in LL evaluation are discussed below.

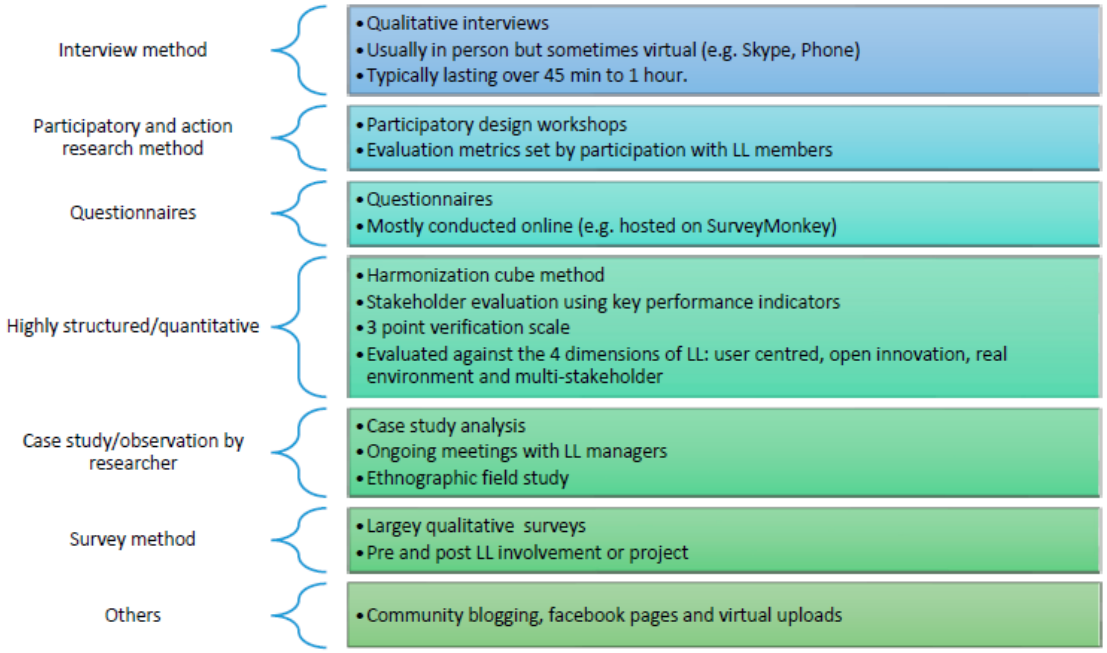

Figure 7. Summary of different evaluation approaches used in LLs studies which discuss evaluation. studies included in this scoping review.

Case study analysis and action research for LL evaluation

Our scoping review reveals that using a case study approach is most common in measuring the impact of LLs. Furthermore, most of the studies we reviewed used more than one case of LL (multi cases) to compare and contrast its implementation approaches and its outcomes. Yin (1984, p. 16) defines the case study research method as “an empirical inquiry that investigates a contemporary phenomenon within its real-life context that should be used when the boundaries between phenomenon and context are not clearly evident, and in which multiple sources of evidence are used”. This description of case study design makes evident why it suits the LL context. Schuurman et al. (2016) also suggest that the case study approach is appropriate to study LLs due to their complexity and specificity regarding particular innovation systems. The particularities of the case study method varied depending on the nature of the evaluation. Approximately 83 percent of the studies included in our review used the case study approach with the number of LLs included in each study ranging from one up to 135. A second dominant method from across the articles is “action research” which is commonly used as a general approach or entry point for evaluation of LLs (Van Geenhuizen, 2018), wherein participants develop the evaluation metrics and even, in some studies, conduct the evaluation themselves. One way the literature we reviewed could be categorized vis-à-vis evaluation is into person-oriented LLs, where implicit evaluation was adopted, versus organization-oriented LLs. The latter are evaluated by comparing expected results with actual results, often using satisfaction among the participating actors and their perceptions of the results assessed after the LL has ended.

One seemingly emergent method for evaluation is the use of digital technologies like smart phones and specific evaluation applications or “apps.” Hofte et al. (2009) argued that “user

experience can be evaluated with lab experiments, interviews, focus groups and/or surveys, many other aspects are harder to investigate if taken out of the natural context of use. Instead of focusing solely on bringing people to the lab, researchers who want to evaluate mobile devices and services are increasingly doing the opposite: bringing the lab to the people." These researchers recommend using a mobile tool—notably smartphones—for data collection. ContextPhone, MyExperience, Xensor, RECON and BeTelGeuse are some of the recently introduced tools used for LL evaluations (Hofte et al. 2009).

Another key insight from our review is that employing a diversity of tools is an important attribute of LL evaluation, for example studies use self-reported methods (for e.g., diaries, experience sampling) alongside researcher measurement (e.g., observation, ethnography) (Hofte et al. 2009, Schuurman et al. 2013).

Qualitative evaluation tools

Structured and systematic evaluation methods exist in the literature, but they are under-represented (Van Geenhuizen, 2018). As such, our review shows that qualitative research methods are most commonly used for evaluating LLs. Some noteworthy studies which detail their qualitative methods used in LLs assessment are Callari et al. (2019), Cech & Wagner (2019), and Georges et al. (2016). Within those qualitative methods used, participatory design but also workshops and open-ended qualitative interviews were the most common methods deployed (Figure 8). Figure 8 was generated from the Full Text screening and indicates the high number of word counts returned from our corpus (138 articles) for the terms "participatory", "workshop", and "qualitative". 74 percent of articles reviewed in this study used participatory action research, workshops, email surveys, phone surveys and semi-structured surveys. 46 percent of the fully screened articles (41 articles) used semi-structured interviews which allows for studying how research participants evaluate the LL process and outcomes themselves (Callari et al. 2019). Our review showed that the length of a single interview varies from 30 min to one hour. Key informants' interviews with stakeholders were also widely used in LLs evaluation, mostly as a measure of validity. Our review also indicated that data generated from qualitative evaluation methods were mostly analyzed by using inductive content analysis (for e.g., Holappa & Sirkka, 2017). Very few studies (see Callari et al. 2019) used a deductive, concept-driven coding frame to analyze interview transcripts.

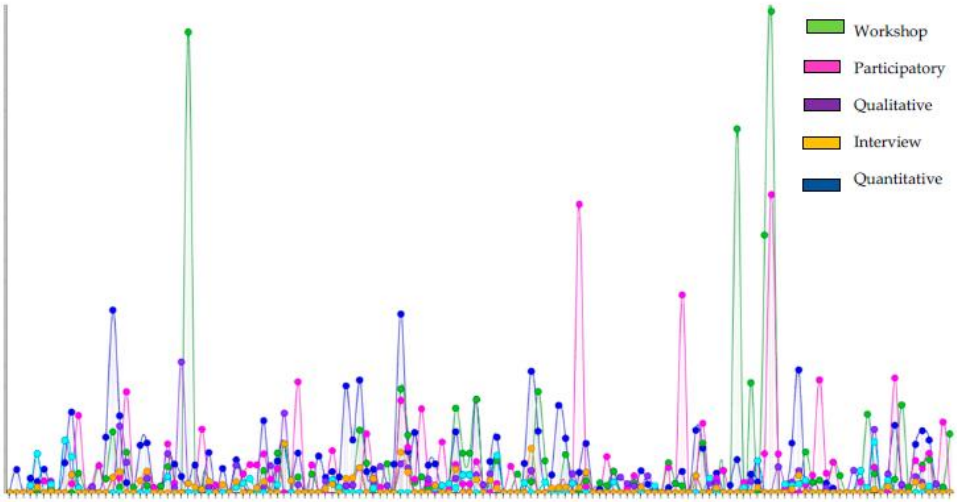

Figure 8. Word counts using Voyant ["interview (dark blue color), workshop (green), participatory (pink), qualitative (purple), quantitative (light blue), and interview (yellow)].

Quantitative evaluation methods and tools

Our review indicated that a minority of studies (26%) used only a quantitative method of data collection for evaluation of LLs and this was used primarily for measuring the impact of technology or ICTs introduced or developed by a LL approach (see Chen & Chou 2016, Hagy et al. 2017). Moreover, these quantitative methods are mostly combined with qualitative methods (as a mixed method of data collection).

One study by Dell'Era et al. (2019) is helpful to understand how study parameters across several LL sites have been quantified in the literature. This study focused on investigating the innovation impacts of user-centered and participatory strategies adopted by European Living Labs (Dell'Era et al. 2019). The adoption frequency of practices was measured using a Likert scale ranging from 1 to 5. In order to capture the strategic approach adopted by each Living Lab, researchers looked at the adoption of different practices. Leveraging the conceptualization "What people say, do and make", this study used user-centered and participatory strategies as binary variables. User-centered (participatory strategy) is equal to 1 if the LL implements at least two out of three related practices in a systematic way and otherwise the score is 0. In this way, both the quantity (breadth) and the frequency (depth) in the adoption of the two sets of practices were assessed.

Our dataset contains several models that have been used across the LL literature for analyzing data that are collected on LL function and impacts. Chen and Chou (2016) developed a Living Lab Analysis Model (LLAM) based on the concept of engineering analysis which includes three module units i.e., principle, process, and signposts. They considered principles and processes as two factors for constructing an analysis model. They developed an interoperability "cube" for harmonizing Living Lab data. Ballon et al. (2018) recommended a logit model to measure the effectiveness of involving users in digital innovation process. A model similar to this logit model is called "Reference Model" which is

recommended by Guzmán et al. (2008) for user-driven innovation assessment that is highly structured. Kovacs (2016) used an “Alcotra and Harmonization cube method” to evaluate the interactive value production coming from LL. Maciuliene and Skaržauskiene (2020) applied a newly developed digital co-creation monitoring technique called Digital Co-Creation Index (DCCI). This methodology provides a systemic understanding of the basic factors shaping the co-creative processes in LLs. Further, Vontas and Protogeros (2009) recommended a PACE
(Project Assets, Core competencies and Exploitable items) evaluation toolkit which is more elaborated than but similar to the DCCI. Overall, our review shows that several scholars are recommending different types of models for structuring evaluation, specifically for those using quantitative data, but these are also found to be rather case specific and not widely applied across contexts. Said differently, there appear to be no studies which demonstrate a robust set of approaches, metrics, analysis methods or an overarching framework for evaluation across LL contexts.

Evaluation methods for LLs specifically related to agroecosystems

Out of 41 final articles which we found relevant to LL and evaluation, only two study were found relevant to agroecosystems and sustainability (Ondiek & Moturi 2019, and Hagy et al. 2017). For instant, Hagy et al. (2017, p. 18) did a study on an innovation ecosystem-specific to the agricultural context and found that “no generic Innovation Ecosystem model was found that could be used to incorporate both Living Lab infrastructures and the built environment, yet a simplified generic model to use for mapping the case studies was still needed.” According to the author (Hagy et al. 2017), to produce an accurate representation of the Innovation Ecosystem for Living Lab infrastructures, a series of tools/methods that should be included in this kind of agroecosystem study include interviews with various actors working within LLs; a workshop with end-users and actors both internal and external to the LL ecosystem; and the authors’ own experiences working within the LL ecosystem.

Evaluation frameworks for LLs

There are several approaches to LL evaluation which have been studied for decades (Thamhain 2013, Patton 2002). For example, the World Bank and United Nations have their own project and program evaluation guidelines to follow while evaluating the outcome of the technological processes or programs. A quasi-experimental design (i.e., pre-test—real-life intervention—post-test) and SWOT (Strengths, Weaknesses, Opportunities and Threats) analyses were used in more than one study for assessing the LLs (for e.g., Schuurman et al. 2013, Schuurman et al. 2016). However, our review of 138 articles on LLs indicated that there is a lack of universally or even widely accepted evaluation methods that exist in practice and which have been established as rigorous across LL contexts.

Our dataset revealed 24 distinct frameworks used for LL evaluation (for e.g., Ballon et al. 2018, Guzmán et al. 2008, Kovacs 2016, Mačiulienė & Skaržauskienė 2020, Osorio et al. 2019, Schuurman et al. 2013, Schuurman et al. 2016). Evaluation frameworks are described in these studies as important to guide the overall assessment process and summarize the final outcome

of the evaluation; scholars argue that frameworks help to bring uniformity in the research process/study. However, among those articles in our dataset, the “harmonization cube” was the only repeated framework.

Table 2 provides a summary of the LL evaluation frameworks included in our review. We found that the most common element among all of these frameworks was assessment of engagement and diversity of stakeholders/partners/users within the innovation system (approach of LLs) as an important indicator of success of LL function. Ondiek and Moturi (2019) used the needs of the users, objectives of the LLs, inputs (financial indicators including budget), operations (within the LLs) and output of the project as independent variables and results (direct and immediate effect of project) and impacts were used as dependent variables to show the relationship. The relevance of LLs in targeting the needs of the users, and LLs’ efficiency, effectiveness, utility and sustainability aspect of the LLs are some of the important factors to be accounted for while evaluating any LLs (Ondiek & Moturi 2019, Ekins et al. 2008).

Table 2

Summary of relevant evaluation frameworks, principles or models found in the LL evaluation literature.

Evaluation Framework/Principles/ Models	Key Focus	Key Elements	Authors
Digital Co-Creation Index (DCCI) framework for evaluation in the EU.	A systemic understanding of the basic factors shaping the co-creative processes in LLs.	Emphasize the interplay between places, technology, and people within LLs.	Mačiulienė & Skaržauskienė (2020).
The four-capital method of sustainable development evaluation, originally developed by Ekins et al. (2008).	Relationship between the needs, objectives, inputs, operations, and output.	Consists of four capitals: human, financial, environmental, and manufactured.	Ondiek & Moturi (2019).
Conceptual framework: mixing user-centred strategy and participatory strategy.	Conceptualise the impacts of the user-centred and participatory strategies on innovation performance outcomes by assessing the project performance and transfer performance.	In user-centred strategy, observing user's behaviours, capturing users' insights, and receiving users' feedback are considered. Co-designing and collaborating with users and enabling users' experience through prototypes are the major elements of participatory strategy.	Dell'Era et al. (2019).
Logical effect model for LL projects.	For the evaluation of small and medium sized enterprises, potential effects of LL projects are categorized as short-term, mid-term and long-term.	Key elements are use, usefulness and value of LL project, initial objectives and achieved effects, effects on investments, revenues, and employment because of LL project results.	Ballon et al. (2018).
A maturity grid-based assessment tool.	Framework developed by reviewing eight frameworks that focus	Guidance tool to evaluate the maturity degree of an innovation	Osorio et al. (2019).

	specifically on innovation laboratories.	laboratory or to adapt an existing LL project.	
Harmonization cube.	LL Harmonization Cube created, in alignment with the structure of the “Rubik” cube.	The columns of the cube describe the organizational, contextual, and technological issues, the rows represent the maturity level of LLs, as: setup, sustainability, and scalability.	Kovács (2016).
Business Model Canvas (BMC) originally developed by Osterwalder & Pigneur (2010).	Strategic management tool.	Draws from the BMC, to assess whether the existing evaluation criteria proposed by ENoLL could be augmented.	Mastelic et al. (2016).
Living Lab triangle-Conceptual framework.	Linking living lab environments, approaches and innovation outcomes to each other.	For data coding, parameters were divided under LL environments (technical infrastructure, ecosystem approach, level of openness—property rights and partnerships, community, real-world context, lifespan, and scale) and LL approach (evaluation, context research, co-creation and user role).	Veeckman et al. (2013).
Process Reference Model (PRM) for Living Lab.	PRM is a catalogue of effective practices organized by maturity level and process, focusing on IT services and rural development.	Grouped into five categories: Innovation initiatives management, Organizational management, Technical development, Monitoring and evaluation and Deployment and operation.	Guzmán et al. (2013).
Key 5 principles to guide evaluation process in LL.	Guide design and assessment of LL impact.	Includes five principles: value, sustainability, influence, realism, openness.	Ståhlbröst (2012).

Monitoring framework of C@R rural living labs.	Focuses on C@R rural living labs results and impacts on value for users, innovation environment and rural development.	Focuses on three main elements: drivers and conditioners of the innovation activity; processes and decisions related to implementing and operating the innovation initiatives; results and impacts of the living lab innovation initiatives.	Guzmán et al. (2008).
Living Lab Analysis Model (LLAM).	Based on the concept of engineering analysis.	Includes three module units i.e., principle, process, and signposts (combination of conditions and actions).	Chen & Chou (2016).
SNA (Social Network Analysis), MASAI® (Marketing Strategies and Business Intelligence Model), and PACE (Project Assets, Core competencies and Exploitable items) evaluation toolkit.	Identifies the key attributes and features of European Living Labs to evaluate the core competences and assets.	Explores the current connections and interrelations among LLs and research projects; MASAI® model: to assess the market success; PACE: to evaluate the intangible assets created during a Living Lab's life and operation.	Vontas & Protogeros (2009).
The Sustainable Livelihood model.	Link between characteristics of the living labs and their effects on the outcome.	Includes three pillars, namely Innovation Outcome, Living Lab Environment, and Living Lab Approach.	Parkinson & Ramirez (2007).

Another common element across the frameworks was the aspect of time used to evaluate the LL function itself from a pre-project to post-project time period (Hyysalo & Hakkarainen 2014, von Wirth et al. 2019). For example, von Wirth et al. (2019) assessed the initial strengths and weaknesses of the LLs in their study and proposed a set of practices which were believed to support the LLs through their creation and initial setup, which were developed by the research support team in a workshop for LLs managers. In the first year of the project, the initial set of practices was used to guide the LLs in managing the participant community and shared infrastructure as well as to support the implementation of innovation initiatives led by user communities. Later in the second and third year, the adoption of these practices was assessed every three months. LL managers provided written reports on the LL's activities and the practices adopted by the end of each period.

Similarly, the long-term financing/budget is an element of LL success which is considered in more than one evaluation framework. For example, Ondiek and Moturi (2019) employed the four-capital method of sustainable development evaluation framework recommended by Ekins et al. (2008) to assess the long-term viability of living labs in Kenya. Different forms of capital—human (productive potential of individuals), financial (funding), environmental (natural resources), and manufactured (infrastructure)—were considered. Ekins et al. (2008) argued that this model is helpful in showing the relationships between key elements of projects in describing how sustainable development can be realized. Stahlbrost (2012) recommends potentially useful principles to guide and design the evaluation of LLs. These five key principles are: value, sustainability, influence, realism and openness. These key principles emphasize value creation for their partners and users as well as the LL's response towards the community within which it operates, which is thought to influence the long-term viability of the LL membership and activities.

Additionally, van Geenhuizen (2018, p. 1285) suggests that "at least five questions need to be addressed in LL evaluation: (1) is the product/service development and design process sufficiently on schedule (working plan and budgets)?; (2) are learning results from users (user feedback) sufficiently integrated into the design process?; (3) do the designing actors remain sufficiently aligned with each other, with a common vision and common interests?; (4) what is the satisfaction of the participant actors with the results and processes so far?, and (5) is the living lab sufficiently open to attract partners in a broader network enabling support in upscaling and implementation?"

Delphi Survey and Workshop Results

Overview of participants

The way participants perceived knowledge around Living Labs were influenced by the nature of their work and backgrounds. Thus, it is important to keep in mind the composition of participants as described below and the potential biases introduced.

For the Delphi surveys, 23 individuals responded (11 in French and 12 in English). Most (77%) reported Living Labs as an area of expertise. The majority were also researchers (68%), especially among respondents of the French survey (82%). 59% of respondents also reported expertise in the field of agriculture, and 36% in the field of environment. Around a quarter (27%) reported evaluation as an area of expertise, and fewer (14%) reported having expertise as a practitioner. When asked about their involvement in Living Labs, half of participants mentioned working in organizations across Canada and Europe in the agriculture, food and environmental sectors. These include living labs at Agriculture and Agri-Food Canada, the AcadieLab, PA4ALL (Precision Agriculture Living Lab) at the BioSense Institute, VitiREV and LIT (Laboratoires d'Innovation Territoriale) at the Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement.

For the virtual Delphi workshops, 17 individuals attended (10 in the French and 17 in the English version). A little more than half of the participants (59%) worked within a Living Lab programs held by France and Canada's public organizations such as AAC Agriculture and Agri-Food Canada and INRAE Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement. The other half came essentially from Universities and colleges across Canada, France and Italy and few professionals from agricultural producer associations. The groups were composed of a majority of researchers in agri-ecosystem and sustainability LLs, evaluation and sociology. Few participants could claim to wear a practitioner's hat, since some of them evolved as LLs projects coordinators.

Expert perspectives of gaps in evaluation of Living Labs

Classification of research themes in high, medium, and low priorities

From the Delphi surveys, we generated a list of top, medium and low priority themes building on participants' ranking of themes in the survey (see Table 3). Participant views on the relative importance of research themes along with identifying literature related to those themes allow us to see where gaps exist in the literature on evaluation of Living Labs, and which gaps to prioritize. Only two themes identified in the survey were perceived to have key references in the literature: the role of stakeholders in evaluation and efficiency of open innovation approaches. References were provided for some other themes.

Table 3

Priority research themes and literature availability for evaluating LLs according to Delphi survey results.

Theme	Range of Rank
High Priority: Themes ranked as #1.	
The role of stakeholders in evaluation.	1-12.
The objectives and use of evaluation results.	1-10.
Efficiency of open innovation approaches.	1-14.
The specific objectives of evaluation.*	1-14.
Evaluation methods and tools.	1-13.
Medium Priority: Themes ranked in mid-range.	

Measuring environmental sustainability.*	1-15.
Conditions for success.	1-15.
The role of evaluators.	1-12.
Measuring social impacts.	2-11.
Scales of evaluation.	2-13.
Temporality of evaluation.*	2-14.
Lowest Priority: Themes ranked the lowest.	
The funding methods of the evaluation.*	4-13.
Evaluation repositories.*	4-15.

*No perceived references available on these research themes and a low number of references provided.

Effectiveness and the efficiency of open innovation, conditions for success, measuring social impacts and measuring environmental sustainability were generally perceived by participants as less studied topics in the context of LLs. An overarching gap identified by many participants is the need to better define effectiveness, success, social impacts and environmental sustainability in tandem with collaborators and participants, as well as how these elements could or should be measured and evaluated. This is both a practical need (e.g. clear definitions can promote effective evaluation and Living Lab practices), and a research gap (e.g. what are the different definitions of success or effectiveness of different actors). Views also converged around constant engagement, ongoing evaluation and continuous follow-up that include pre- and post-assessments with a large number of stakeholders as a key feature to promote the effectiveness of LLs and to support the evaluation of its impacts. Other key features to promote the effectiveness include, among others: collaboration and active co-creation, relationships with internal and external partners, real-life settings, adaptability, flexibility, openness, resources, preparation and accountability. Key elements for effective evaluation of impacts identified by participants include transparency, diversity of evaluation participants and the collaborative nature of the evaluation.

While recognized as key, participants identified many knowledge gaps tied to collaboration (e.g. co-governance models, co-design, co-creation of values, building trust, managing expectations). Specific processes to LLs, ability to manage change, the role of creativity and effectiveness compared to other participatory approaches also need to be better known. Participants converged on the need to better understand social and environmental impacts of LLs but they provided different potential indicators for how to measure these impacts. The lack of understanding of the overall relationships between LLs and environmental sustainability was also identified as a gap. Participants converged around the need to do more research about the

evaluation of Living Labs (including its purpose, objectives, actors, references and how to apply evaluation methods, tools and indicators), the role of data (ownership, access, information sharing, date-use agreements), funding (funding the evaluation, mismatch between risk-averse funders and the inherent unpredictability of Living Labs), temporality (of Living Labs, of evaluation, of funding, of governance) and politics, fairness and inclusivity.

Delphi workshop and research agenda findings

Guided by the broad research themes from the Delphi surveys, participants identified a wide range of research questions. Many of the key features and gaps identified by participants in the survey were revisited in workshop discussions (ie. collaboration, evaluation, temporality, impacts, effectiveness/efficiency, stakeholders, and funding). However, some of these did not make it as a standalone theme in the final research agenda (ie. creativity, politics). We reviewed each research question and identified related subtopics, which are nested under the framing of who, what, why, how, where, when. (See Appendix E for full research agenda) for the finalized research agenda outlining main themes, sub-themes, synthesis questions, and question descriptions. Table 4 illustrates an overview of this research agenda.

Table 4

Overview of research agenda findings produced from the Delphi workshops.

Research Questions	Sub-themes
Who is involved and who holds the different roles in the evaluation of LLs?	Diversity of actors, roles and involvement, role of equity and power relations in evaluations.
How can success be promoted in LLs?	Definitions, measures and conditions for success, role of mistakes and failure.
Why are LLs being evaluated?/ How is evaluation funded?	Evaluation using different interests/multiple objectives, use of evaluation results, funding influences on LL potential.
How are LLs evaluated?	Evaluation methods and frameworks, available references, perspectives of different evaluation strategies, trust and willingness to share data, comparison of evaluation approaches.
Where is the evaluation taking place?	Scales and timelines for evaluation tools and processes.
When is the evaluation taking place?	Temporality, types of measurements, timeline alignment between LL actors, evolution of LL components.
What is being evaluated?	Evaluation of general impacts, LL processes, social

	impacts, environmental impacts, social-environmental impacts.

In terms of who is involved in the evaluation of LLs, knowledge gaps were identified relating to the different groups and the diversity of actors involved in evaluation (e.g. stakeholders, rightsholders, participants, partners, non-humans), the different types of involvement and roles that actors have as part of the evaluation process, and questions tied to equity and power dynamics in evaluation. Gaps in knowledge tied to the conditions for successful LLs included the different definitions of success held by different actors, the different ways of measuring and ensuring this success as well as the role of trial and error as well as failure, which are key to the LL approach but may not be perceived uniformly by actors (e.g. fundays may be risk-averse). Gaps tied to the purpose of evaluation and use of its results tied to the varying objectives and interests of actors, different possible applications of evaluation results across contexts and the influence of evaluation and its results on the LL. It could also be better known how to leverage funding in ways that promote successful LLs. Knowledge gaps tied to methods of evaluation were numerous and echo findings from the scoping review, including how to establish a common methodology for evaluation of LLs processes and impacts, strengths and limitations or different methods including qualitative and participatory methods, the role that references can play in evaluation, the tensions between subjectivity and objectivity in the evaluation of LLs and how to best enable comparison between LLs and other methods, as well as among LLs from different sectors or who use distinct processes like creation versus validation. Knowledge gaps were also identified in relation to the scales of evaluation, more specifically in relation to which method should be used at which scale and how methods could be integrated across scales to evaluate LLs processes as well as their social, environmental and socio-environmental impacts and outcomes. The alignments and mismatches between the timelines of different actors and processes, as well as the evaluation of different dimensions of LLs and the evolution of actors over time, could also be better understood. Many gaps were identified in terms of what is being evaluated in the evaluation of LLs, primarily the process of LLs and their impacts. Going beyond the initial scope of this research, participants emphasized the need to better understand how to define and measure different types of impacts (social, economic, environmental) as well as better understanding how these impacts and different components of LLs are interconnected (e.g. social-environmental impacts, connection between diversity of actors and social impacts). Participants identified the possibility for impacts to be direct or indirect and to occur at different scales (e.g. community, regional or global levels), but it wasn’t clear what these impacts are or how to evaluate them in a consistent way.

Implications

WIth the scoping review, we initially set out to verify that a gap exists within LL scholarship around tools for and approaches to evaluating both the internal dimensions of LLs (e.g., how effectively do participants communicate and build networks amongst themselves?) and their external impacts (e.g., do they lead to wider social change?). We were specifically interested in LLs focused on environmental or agricultural sustainability. Our paper also aimed to synthesize any existing best practices for evaluation of LLs. The major research question that guided this study was “What general evaluation methods or metrics exist for measuring the effectiveness of LLs in general, and then among those specific to environmental and agricultural sustainability?”.

It appears that there are no widely agreed upon and applied methods or frameworks for evaluating LLs across contexts. Indeed, the most common approach to gathering data which came out of our analysis was comparative case studies and we found that, in general, the purpose of evaluation among those articles which put forward evaluation tools was improvement of the particular LL functioning (67%) not its wider impacts. Moreover, a common entry point for evaluation among the studies in our final dataset was action research where participants of the LL help develop the metrics and indicators that come to be used to evaluate the LL.

The reason for the plurality in methods of evaluation likely has to do with the fact that LLs are by definition user-centric, thus evaluation approaches are guided by different organizations, agencies and stakeholder groups depending on the location and specific mandate of the LL (Thamhain 2013, Leminen 2016, Greve 2020). However, this may pose a problem as case study research itself is not often widely generalizable even if comparisons are made across a number of cases. More structured LL evaluation methods that have been applied across jurisdictions and individual studies do exist, but these were under-represented in our dataset and appear to be applied specifically to LLs that aim to design or prototype technologies (specifically ICTs). Some noteworthy studies which give a high level of detail regarding their qualitative methods used in LLs assessment are Callari et al. (2019), Cech and Wagner (2019), and Georges et al. (2016).

This gap in the academic literature is consequential if LLs want to move beyond particularity to make broader claims about the value of the LL approach. One paper which we found during this review (Ekins 2008) also highlights the need for a unified approach to evaluating LLs - one which might guide in comparing multiple cases by using common indicators. This need was confirmed by the Delphi process we conducted with experts. Such an approach could address the potential managerial, organizational, and design aspects of LLs and lead to overall improvement or the iteration of knowledge on LL practice over time and across jurisdictions.

The Delphi survey and workshops gathered expert knowledge and information on LL evaluations, where participant discussions further specified current knowledge gaps in academic literature. The Delphi workshop was a collaborative platform for practitioners across various LL sectors to discuss LL evaluations based on their unique experiences and knowledge-bases. It allowed participants to build a network with others involved in the LL research field. The results of our scoping review were supported with the Delphi workshop’s resulting research

agenda, where there was a high priority to identifying evaluation tools to measure social, environmental, and social-environmental impacts.

Conclusion

Ultimately, these findings from the Delphi surveys and workshops were synthesized into a research agenda that outlines what research questions need to be addressed to evaluate the effectiveness and impacts of LLs. Our knowledge synthesis project provides insights on how to evaluate the social processes underlying the functioning of LLs and the long-term social impacts of both the processes and outcomes of LL innovations. The fact that Canada is leading a pair of complementary, nation-wide agroecosystem LL initiatives serves as a platform and opportunity to understand its effectiveness on sustainable agricultural practices, and how it may serve as a model for other environmental and natural resource sectors.

A common finding across the scoping review and the Delphi research agenda is the use of empowerment and the need for diversity in LL processes. Application of the Delphi workshop findings include mobilizing the networks of experts that developed during the workshops to act on the resulting research agenda on evaluation of LLs. The diversity of actors and their definitions, expectations, and perspectives surrounding the effectiveness and impacts of LLs should be considered in the evaluation process. This inclusion can support the development of relevant metrics for evaluating social and sustainability impacts from LL in an equitable way, which leads to improved and effective evaluation processes.

Also, the Delphi research agenda offers nuances for research and operations that capture the diverse nature of social and environmental impacts of LLs over time and at different scales, while also offering potential for greater qualitative and participatory methods in the evaluation and reporting of LL processes and outcomes. Actioning and consideration of the themes from the agenda can open spaces in policy to co-design questions and methods, co-create processes, and co-evaluate impacts and outcomes.

Several large networks of LL initiatives have recently been formed in North America and across Europe, some of which focus on social innovation, rural innovations, and sustainability (Leminen et al. 2016, Greve et al. 2020). Our project has potential to increase visibility and leverage LL approaches for environmental and agricultural sustainability, in Canada and globally. Overall, this project identified a gap in evaluation of LLs in agricultural and environmental sustainability. At the same time, we have identified opportunities for researchers and practitioners to use the resulting research agenda as a platform to collaborate, leverage the LL approach for sustainability, and build upon lessons learned for both the implementation and evaluation of LLs.

Knowledge Mobilization Activities

The following list outlines all knowledge mobilization activities using our research findings as of June 2021.

Papers:
Bronson, K., Devkota, R., & Nguyen, V. Moving toward Generalizability? A Scoping Review on Measuring the Impact of Living Labs. Sustainability 2021, 13, 502. https://doi.org/10.3390/su13020502

We expect two more submissions of peer-reviewed manuscripts publishing the research agenda, and another describing challenges/enablers for effective LLs using additional data from the scoping review.

Presentations to Practitioners/End-Users:
Nguyen, V., & Bronson, K. Environmental and Agricultural Sustainability: Preliminary Scoping Review on Evaluating Effectiveness of Living Labs. Living Labs LTAR Socio-economic Workshop. 15 September 2020.

Bronson, K. A Scoping Review of Research on Measurement of Living Labs Impacts. Agriculture and Agri-Food Canada Living Labs Socio-economic Working Group. 26 February 2021.

Bronson, K., & Nguyen, V. Environmental and Agricultural Sustainability: Scoping Review on Evaluating Effectiveness of Living Labs. SSHRC KSG Forum - Living Within Earth's Carrying Capacity, 1 April 2021.

Conferences:

McPhee, C. 2020. Hot Topic Discussion: How can we measure performance and impact in living labs? XXXI ISPIM Innovation Conference (Virtual), 8 June 2020.

McPhee, C., Bancerz, M., & Chrétien, F. 2021. Les caractéristiques et les impacts sociaux des laboratoires vivants dans les agroécosystèmes. 88e Congrès de l'Acfas. 6 mai 2021.

Chris McPhee, Vivian Nguyen, Christine Beaudoin, Albana Berberi, Kelly Bronson, Rachana Devkota, Jean-François Jasmin, Steve Joncoux, Dominira Saul, R. Sandra Schillo, and Mika Westerlund. 2021. Evaluation of Living Labs Focused on Agricultural or Environmental Sustainability. XXXII ISPIM Innovation Conference (Virtual), 20-23 June 2021.

Bibliography

Almirall, E., Lee, M., & Wareham, J. (2012). Mapping living labs in the landscape of innovation methodologies. *Technology innovation management review*, *2*(9).

Ballon, P., & Schuurman, D. (2015). Living labs: concepts, tools and cases. *Info*.

Ballon, P., Van Hoed, M., & Schuurman, D. (2018). The effectiveness of involving users in digital innovation: Measuring the impact of living labs. *Telematics Informatics*, *35*(5), 1201-1214.

Boon, W. P., Chappin, M. M., & Perenboom, J. (2014). Balancing divergence and convergence in transdisciplinary research teams. *Environmental science & policy*, *40*, 57-68.

Bronson, K., Devkota, R., & Nguyen, V. (2021). Moving toward Generalizability? A Scoping Review on Measuring the Impact of Living Labs. *Sustainability*, *13*(2), 502.

Bulkeley, H., Coenen, L., Frantzeskaki, N., Hartmann, C., Kronsell, A., Mai, L., ... & Palgan, Y. V. (2016). Urban living labs: governing urban sustainability transitions. *Current Opinion in Environmental Sustainability*, *22*, 13-17.

Burbridge, M. (2017). If Living Labs are the Answer–What's the Question? A Review of the Literature. *Procedia engineering*, *180*, 1725-1732.

Callari, T. C., Moody, L., Saunders, J., Ward, G., Holliday, N., & Woodley, J. (2019). Exploring participation needs and motivational requirements when engaging older adults in an emerging Living Lab. Technology Innovation Management Review, 9(3), 38-4

Chen, Y. T., & Chou, W. H. (2010, June). Constructing living labs analysis model for designing and evaluating living labs systems. In *2010 IEEE International Technology Management Conference (ICE)* (pp. 1-8). IEEE.

Cech, F., & Wagner, M. (2019, June). eRollin'On Green: A Case Study on Eco-Feedback Tools for eMobility. In *Proceedings of the 9th International Conference on Communities & Technologies-Transforming Communities* (pp. 121-125).

Dell'Era, C., Landoni, P., & Gonzalez, S. J. (2019). Investigating The Innovation Impacts of User-Centred And Participatory Strategies Adopted By European Living Labs. *International Journal of Innovation Management*, *23*(05), 1950048.

Ekins, P., Dresner, S., & Dahlström, K. (2008). The four-capital method of sustainable development evaluation. *European Environment*, *18*(2), 63-80.

Gamache, G., Anglade, J., Feche, R., Barataud, F., Mignolet, C., & Coquil, X. (2020). Can living labs offer a pathway to support local agri-food sustainability transitions?. *Environmental Innovation and Societal Transitions*, *37*, 93-107.

Georges, A., Schuurman, D., & Desmet, M. (2015, June). Uncovering the needs and wants of end-users towards green apps: A Living Lab approach. In *2015 IEEE International Conference on Engineering, Technology and Innovation/International Technology Management Conference (ICE/ITMC)* (pp. 1-7). IEEE.

Greve, K., Leminen, S., DE VITA, R. I. C. C. A. R. D. O., & Westerlund, M. (2020). Unveiling the diversity of scholarly debate on living labs: A bibliometric approach. *International Journal of Innovation Management*, 2040003.

Guzman, J. G., Schaffers, H., Bilicki, V., Merz, C., & Valenzuela, M. (2008, June). Living labs fostering open innovation and rural development: Methodology and results. In *2008 IEEE International Technology Management Conference (ICE)* (pp. 1-8). IEEE.

Hagy, S., Bard, F., Sasic, A., Sredanovic, E., & Camarasa, C. (2017). Next Generation Living Labs.

Hetenyi, G., Lengyel, A. D., & Szilasi, M. D. (2019). Quantitative analysis of qualitative data: Using voyant tools to investigate the sales-marketing interface. *Journal of Industrial Engineering and Management*, *12*(3), 393-404.

Hofte, H., Jensen, K. L., Nurmi, P., & Froehlich, J. (2009, September). Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild: http://mll09. novay. nl. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-2).

Holappa, N., & Sirkka, A. (2017). Living Lab as an Agile Approach in Developing User-Friendly Welfare Technology. In *Harnessing the Power of Technology to Improve Lives* (pp. 654-659). IOS Press.

Hossain, M., Leminen, S., & Westerlund, M. (2019). A systematic review of living lab literature. *Journal of cleaner production*, *213*, 976-988.

Hyysalo, S., & Hakkarainen, L. (2014). What difference does a living lab make? Comparing two health technology innovation projects. *CoDesign*, *10*(3-4), 191-208.

Keyson, D. V., Al Mahmud, A., & Romero, N. (2013, December). Living lab and research on sustainability: Practical approaches on sustainable interaction design. In *International Joint Conference on Ambient Intelligence* (pp. 229-234). Springer, Cham.

Kovács, K. (2016). Evaluation and Practice of Interactive Value Production in Living Labs. *Periodica Polytechnica Social and Management Sciences*, *24*(1), 52-59.

Lam, J., Koustas, E., Sutton, P., Johnson, P. I., Atchley, D. S., Sen, S., ... & Woodruff, T. J. (2014). The Navigation Guide—evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth. *Environmental health perspectives*, *122*(10), 1040-1051.

Leminen, S., & Westerlund, M. (2019). Living labs: From scattered initiatives to a global movement. *Creativity and Innovation Management*, *28*(2), 250-264.

Turoff, M., & Linstone, H. A. (2002). The Delphi method-techniques and applications.

Mačiulienė, M., & Skaržauskienė, A. (2020). Sustainable urban innovations: digital co-creation in European living labs. *Kybernetes*.

McPhee, C., Bancerz, M., Mambrini-Doudet, M., Chrétien, F., Huyghe, C., & Gracia-Garza, J. (2021). The defining characteristics of agroecosystem living Labs. *Sustainability*, *13*(4), 1718.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22*(3), 276-282.

Miller, A. (2018). Text Mining digital humanities projects: Assessing Content analysis capabilities of voyant tools. *Journal of Web Librarianship*, *12*(3), 169-197.

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology*, *18*(1), 1-7.

Ondiek, M. A., & Moturi, C. (2019). An assessment of the sustainability of Living Labs in Kenya. *Innovation & Management Review*.

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley & Sons.

Parkinson, S., & Ramirez, R. (2006). Using a sustainable livelihoods approach to assessing the impact of ICTs in development. *The Journal of Community Informatics*, *2*(3).

Patton, M. Q. (2008). *Utilization-focused evaluation*. Sage publications.

Pohl, C. (2005). Transdisciplinary collaboration in environmental research. *Futures*, *37*(10), 1159-1178.

Puerari, E., De Koning, J. I., Von Wirth, T., Karré, P. M., Mulder, I. J., & Loorbach, D. A. (2018). Co-creation dynamics in urban living labs. *Sustainability*, *10*(6), 1893.

Sampsel, L. J. (2018). Voyant Tools. *Music Reference Services Quarterly*, *21*(3), 153-157.

Schliwa, S., Evans, J., McCormick, K., & Voytenko, Y. (2015, March). Living labs and sustainability transitions—Assessing the impact of urban experimentation. In *Proceedings of the INOGOV Workshop: Climate Change Policy and Governance: Initiation, Experimentation, Evaluation, Helsinki, Finland* (pp. 12-13).

Schuurman, D., Baccarne, B., Kawsar, F., Seys, C., Veeckman, C., De Marez, L., & Ballon, P. (2013). Living labs as quasi-experiments: results from the Flemish LeYLab. In *XXIV ISPIM Conference: Innovating in Global Markets: Challenges for Sustainable Growth*.

Schuurman, D., De Marez, L., & Ballon, P. (2016). The impact of living lab methodology on open innovation contributions and outcomes. *Technology Innovation Management Review*, *6*(1).

Ståhlbröst, A. (2012). A set of key principles to assess the impact of Living Labs. *International Journal of Product Development*, *17*(1-2), 60-75.

Thamhain, H. J. (2013). Contemporary methods for evaluating complex project proposals. *Journal of Industrial Engineering International*, *9*(1), 1-14.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, *169*(7), 467-473.

Van Geenhuizen, M. (2018). A framework for the evaluation of living labs as boundary spanners in innovation. *Environment and Planning C: Politics and Space*, *36*(7), 1280-1298.

Veeckman, C., Schuurman, D., Leminen, S., & Westerlund, M. (2013). Linking living lab characteristics and their outcomes: Towards a conceptual framework. *Technology Innovation Management Review*, *3*(12).

Vontas, A., & Protogeros, N. (2009, June). Evaluating living labs core competences and assets. In *2009 3rd IEEE International Conference on Digital Ecosystems and Technologies* (pp. 558-562). IEEE.

von Wirth, T., Fuenfschilling, L., Frantzeskaki, N., & Coenen, L. (2019). Impacts of urban living labs on sustainability transitions: Mechanisms and strategies for systemic change through experimentation. *European Planning Studies*, *27*(2), 229-257.

Voytenko, Y., McCormick, K., Evans, J., & Schliwa, G. (2016). Urban living labs for sustainability and low carbon cities in Europe: Towards a research agenda. *Journal of cleaner production*, *123*, 45-54.

Yin, R. (1984). Introduction in Yin, R Case study research: design and Methods. *Beverly Hills: Sage Publications*, 13-26.

Appendix A: Benchmark Papers

The following is a comprehensive list of benchmark articles referenced when developing a search string.

References:

Ballon, P., Van Hoed, M., & Schuurman, D. (2018). The effectiveness of involving users in digital innovation: Measuring the impact of living labs. Telematics and Informatics, 35(5), 1201-1214.

Edwards-Schachter, M. E., Matti, C. E., & Alcántara, E. (2012). Fostering quality of life through social innovation: A living lab methodology study case. *Review of Policy Research, 29*(6), 672-692.

G20 MACS. (2019). Agroecosystem Living Laboratories: Executive Report. G20 Meeting of Agricultural Chief Scientists (MACS) International Agroecosystems Living Laboratories (ALL) Working Group.

Hossain, M., Leminen, S., & Westerlund, M. (2019). A systematic review of living lab literature. *Journal of Cleaner Production*, 213, 976-988.

Ibarra, E. R. B. (2020). Revisión sistemática del concepto de laboratorios vivos. *Dimensión Empresarial, 18*(1).

Verloo, H., Lorette, A., Gomes da Rocha, C., Amoussou, J. R., Gillès de Pélichy, E., Matos Queiros, A., ... & Von Gunten, A. (2020). A comprehensive scoping review protocol of using living labs to explore needs and solutions for older adults with dementia. *Smart Homecare Technology and TeleHealth*, 7, 19-27.

Zen, I. S. (2017). Exploring the Living Learning Laboratory: An approach to strengthen campus sustainability initiatives by using sustainability science approach. *International Journal of Sustainability in Higher Education, 18*(6), 939-955.

Zivkovic, S. (2018). Systemic Innovation Labs: A lab for wicked problems. Social Enterprise Journal.

Appendix B: Eligibility Criteria

The following outlines the screening eligibility criteria to include or exclude articles. If any of the questions are answered with “NO”, then the article was excluded. If all the questions are answered “YES”, then the article is included. If there is any uncertainty, the reviewer assigned the article as *Include Second Opinion*, and the article was further assessed by the review team.

Question 1: Is this study about *“true” Living Labs*?	
Responses	● **Yes (Include)** ● **No (Exclude)** ● **Unclear (Include Second Opinion)**
Notes	Living Lab definition: a mechanism or approach that brings ordinary citizens and technology users together with stakeholders from the private and public sectors, and they show great potential for bringing forward user-centric solutions and innovations for solving complex environmental issues, for example generating climate change adaptation, and more sustainable natural resource management (Hossain et al. 2019). ● Include if mentioned terms such as ○ Co-creation / participatory / workshops ○ Urban Living Lab / smart cities / mobility as a service ○ Rural Living Lab or agricultural/environmental Living Lab ○ User’s real-life environment / user-centred ○ Public private partnership ○ Quadruple helix ○ ENoLL / European Commission

Question 2: Is this study about *the evaluation of Living Labs*?	
Responses	● **Yes (Include)** ● **No (Exclude)** ● **Unclear (Include Second Opinion)**

Notes	● Include if **directly** evaluating Living Labs ○ E.g., interviews, user feedback, indicators or measures ● Include if **indirectly** evaluating Living Labs ○ E.g., once innovation is produced, the study evaluates how well it is received by the wider world / non-participants (implying the Living Lab was successful) ○ E.g., studying the wider benefits of Living Labs ● Exclude if the paper's focus is ONLY on improving a product/innovation through the Living Labs' activities

Question 3: Is this study about *the effectiveness of Living Labs*?	
Responses	● **Yes (Include)** ● **No (Exclude)** ● **Unclear (Include Second Opinion)**
Notes	● Include if **directly** discusses effectiveness of Living Labs ○ E.g., whether Living Labs facilitated collaborations, adoption, learning, etc. ● Include if **indirectly** discussing effectiveness of Living Labs ○ E.g., once a product or service of Living Labs goes into the real world and if it was effective, impactful, adopted, etc. ● Exclude if paper focus on effectiveness of the product or service or innovation ○ E.g., the innovation was effective at reducing chemical A, the product worked for lowering energy consumption etc.

Appendix C: Delphi Survey

The following section is the Delpi survey in English (the same surveys were also translated and distributed in French).

KSG Delphi - Effectiveness and social impacts of "Living Labs" in agriculture and the Environment

Introduction

You are here because you accepted to participate in a series of discussions on the effectiveness and social impacts of Living Labs within the agricultural and environmental fields. This proposal is part of a SSHRC-funded Knowledge Synthesis project, led by Carleton University, Agriculture and Agri-Food Canada, and the Rivière-du-Loup Living lab en innovation ouverte (LLio).

In parallel with a systematic literature review, these Delphi-method inspired discussions aim to co-construct a research agenda to fill the knowledge gaps on the following question: How to measure the effectiveness and social impacts of “Living Labs” in the agricultural and environmental fields?

Who are you?

- *Name*
- *Profession*
- *Connecting organization*
- *Area of expertise (click the following the apply)*
    - *Researchers*
    - *Practitioner*
    - *Living Lab*
    - *Agriculture*
    - *Environment*
    - *Social Impacts*
    - *Other? (Please specify)*
- *Describe in a few lines your involvement in Living Labs domain (Open answer)*

Part One: Key Elements for Living Labs (Open answer)

- *What are the key elements to promote the effectiveness of Living Labs? (5 maximum)*
- *What are the key elements to evaluate the social impacts of Living Labs? (5 maximum)*
- *Regarding these key elements, which need to be better known? (5 maximum)*

Part Two: Ranking Evaluation Themes (ranking 1-15)

- *Rank the following themes in order of importance for advancing knowledge about Living Labs (1 being the highest priority). You may add missing themes before proceeding with ranking.*
    - *The role of evaluators*
    - *The role of stakeholders in evaluation*

    - *The temporality of the evaluation*
    - *The scales of evaluation*
    - *The objectives and use of the evaluation results*
    - *The funding methods of the evaluation*
    - *Evaluation repositories*
    - *Evaluation methods and tools*
    - *The specific objectives of the evaluation in collaborative approaches*
    - *Conditions for success*
    - *Measuring social impacts*
    - *Measuring environmental sustainability*
    - *Other 1*
    - *Other 2*
    - *Other 3*

If you added other options to the ranking, please specify: (Open answer)

- *Other 1*
- *Other 2*
- *Other 3*

Part Three: Further Assessment of Themes Within Living Labs Approaches

The role of evaluators

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

The role of stakeholders in evaluation

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

The temporality of the evaluation

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

The scales of evaluation

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

The objectives and use of the evaluation results

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

The funding methods of the evaluation

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

Evaluation repositories

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

Evaluation methods and tools

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

The specific objectives of the evaluation in collaborative approaches

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*

    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

Conditions for success

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

Measuring social impacts

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

Measuring environmental sustainability

- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

Other Themes

- *Theme(s) (Open answer)*
- *Are there any major references in the literature or practices on this theme?*
    - *Yes*
    - *No*
    - *I don't know*
- *If so, what do you think are the top 3? (Open answer)*
- *What aspects of this theme remain to be developed? (Open answer)*

Snowball Sampling

- *Who in your network do you think would have relevant experience to participate in this Delphi process?*
    - *Provide contact details if possible (Open Answer)*

Other Comments? (Open answer)

Appendix D: Scoping Review Articles

The following section is a reference list for all 41 included articles used in the Living Labs scoping review.

References:

1. Ballon, P., Van Hoed, M., & Schuurman, D. (2018). The effectiveness of involving users in digital innovation: Measuring the impact of living labs. Telematics Informatics, 35(5), 1201-1214.
2. Benavent, J., Colobrans, J., Mari, S., Castro, J., & Colome, J. (2011). Lessons learned from users: The development of the LivingLab4carers platform case. 1–8.
3. Callari, T.C., Moody, L., Saunders, J., Holiday, N. and Woodley, J. (2019). Exploring Participation Needs and Motivational Requirements When Engaging Older Adults in an Emerging Living lab. Technology Innovation Management Review, 9 (3): 38-49
4. Cech, F., & Wagner, M. (2019, June). eRollin'On Green: A Case Study on Eco-Feedback Tools for eMobility. In Proceedings of the 9th International Conference on Communities & Technologies-Transforming Communities (pp. 121-125).
5. Chen, Y.T. & Chou, W.H. (2016). Constructing living labs analysis model for designing and evaluating living labs system. ICE: 2010 IEEE International Technology Management Conference, 21-23, 1-8.
6. Davies, A. (2018). HomeLabs: domestic living laboratories under conditions of austerity. In Urban Living Labs (pp. 126-146). Routledge.
7. Delina, L. (2020). A rural energy collaboratory: co-production in Thailand's community energy experiments. Journal of Environmental Studies and Sciences, 10(1), 83–90. https://doi.org/10.1007/s13412-019-00572-x
8. Dell'Era, C., Landoni, P., & Gonzalez, S. J. (2019). Investigating The Innovation Impacts Of User-Centred And Participatory Strategies Adopted By European Living Labs. International Journal of Innovation Management, 23(05), 1950048.
9. García-Guzmán, J., del Carpio, A. F., De Amescua, A., & Velasco, M. (2013). A process reference model for managing living labs for ICT innovation: A proposal based on ISO/IEC 15504. Computer Standards & Interfaces, 36(1), 33-41.
10. Georges, A., Schuurman, D., & Desmet, M. (2016). Uncovering the needs and wants of end-users towards green apps: A living lab approach. 2015 IEEE International Conference on Engineering, Technology and Innovation/International Technology Management Conference, 22-24, 1-7.
11. Giboreau, A. (2018). Situational factors and the design of in situ evaluations. In Methods in Consumer Research, Volume 2 (pp. 109-116). Woodhead Publishing.
12. Guzmán, J.G., del Carpio, A. F., Colomo-Palacios, R., & de Diego, M. V. (2013). Living labs for user-driven innovation: a process reference model. Research-Technology Management, 56(3), 29-39. http://dx.doi.org/10.5437/08956308X5603087
13. Guzmán, J.G., Schaffers, H., Bilicki, V., Merz, C. & Valenzuela, M. (2016). Living labs fostering open innovation and rural development: Methodology and results. 2008 IEEE International Technology Management Conference (ICE)

14. Hagy, S., Morrison, G. M., & Elfstrand, P. (2017). Co-creation in living labs. In Living Labs (pp. 169-178). Springer, Cham.
15. Hofte, H., Jensen, K. L., Nurmi, P., & Froehlich, J. (2009, September). Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild: http://mll09. novay. nl. In Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 1-2).
16. Holappa, N., & Sirkka, A. (2017). Living Lab as an Agile Approach in Developing User-Friendly Welfare Technology. In Harnessing the Power of Technology to Improve Lives (pp. 654-659). IOS Press.
17. Hyysalo, S., & Hakkarainen, L. (2014). What difference does a living lab make? Comparing two health technology innovation projects. CoDesign, 10(3-4), 191-208.
18. Leonardi, C., Doppio, N., Lepri, B., Zancanaro, M., Caraviello, M., & Pianesi, F. (2014, October). Exploring long-term participation within a living lab: satisfaction, motivations and expectations. In Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (pp. 927-930).
19. Ley, B., Ogonowski, C., Mu, M., Hess, J., Race, N., Randall, D., Rouncefield, M. and Wulf, V. (2015). At home with users: A comparative view of living lab. Interacting with Computers, 27(1), 21 – 35
20. Logghe, S., & Schuurman, D. (2017). Action research as a framework to evaluate the operations of a living lab. Technology Innovation Management Review, 7(2), 35-41.
21. Luo, K., Lin, S., Shao, K., & Lin, H. (2012). Developing an Engineering Data Bank Service for the Precision Machinery Industry Cluster Using the Living Lab Concept. International Journal of Automation and Smart Technology, 2(3), 265–275. https://doi.org/10.5875/ausmt.v2i3.147
22. Jiang, C., Xiao, Y., & Cao, H. (2020). Co-Creating for Locality and Sustainability: Design-Driven Community Regeneration Strategy in Shanghai's Old Residential Context. Sustainability, 12(7), 2997.
23. Kovács, K. (2016). Evaluation and Practice of Interactive Value Production in Living Labs. Periodica Polytechnica Social and Management Sciences, 24(1), 52-59.
24. Mačiulienė, M., & Skaržauskienė, A. (2020). Sustainable urban innovations: digital co-creation in European living labs. Kybernetes.
25. Mastelic, J., Sahakian, M., Bonazzi, R. (2015). How to keep a living lab alive?. Info,17(4), 12 - 25.
26. Moore, T.; Horne, R. and Doyon, A. (2020). Housing Industry Transitions: An Urban Living Lab in Melbourne, Australia. Urban Policy and Research, 38:2,118-131, DOI: 10.1080/08111146.2020.1730786"
27. Moro, A., & Puerari, E. (2015). Ecosystem Innovation as Triggers of New Paths and Practices for Urban Space. 10th International Forum on Knowledge Asset Dynamics, pp. 1886-1897.
28. Nesti, G. (2018). Co-production for innovation: the urban living lab experience. Policy and Society, 37(3), 310-325.
29. Ogonowski, C., Ley, B., Hess, J., Wan, L., & Wulf, V. (2013, April). Designing for the living room: long-term user involvement in a living lab. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1539-1548).

30. Osorio, F., Dupont, L., Camargo, M., Palominos, P., Peña, J. I., & Alfaro, M. (2019). Design and management of innovation laboratories: Toward a performance assessment tool. Creativity and Innovation Management, 28(1), 82-100.
31. Pigot, H. and Giroux, S. (2015). Living labs for designing assistive technologies. HealthCom, 2015(7454493), 170-176
32. Salminen, J., Konsti-Laakso, S., Pallot, M., Trousse, B., & Senach, B. (2011). Evaluating user involvement within living labs through the use of a domain landscape. 1–10.
33. Schaffers, H., Merz, C. & Guzman, J.G. (2016). Living labs as instruments for business and social innovation in rural areas. ICE : 2009 IEEE International Technology Management Conference, 22-24, 1-8.
34. Schuurman, D., Baccarne, B., Kawsar, F., Seys, C., Veeckman, C., De Marez, L., & Ballon, P. (2013). Living labs as quasi-experiments: results from the Flemish LeYLab. In XXIV ISPIM Conference: Innovating in Global Markets: Challenges for Sustainable Growth.
35. Schuurman, D., De Marez, L., & Ballon, P. (2016). The impact of living lab methodology on open innovation contributions and outcomes. Technology Innovation Management Review, 6(1).
36. Smit, D., Herselman, M., Eloff, J., Ngassam, E., Venter, E., Ntawanga, F., Cheng-Hui Chuang, & Van Greunen, D. (2011). Formalising living labs to achieve organisational objectives in emerging economies. 1–7.
37. Ståhlbröst, A. (2012). A set of key principles to assess the impact of Living Labs. International Journal of Product Development, 17(1-2), 60-75. https://doi.org/10.1504/IJPD.2012.051154
38. Van Geenhuizen, M. (2018). A framework for the evaluation of living labs as boundary spanners in innovation. Environment and Planning C: Politics and Space, 36(7), 1280-1298.
39. Veeckman, C., Schuurman, D., Leminen, S., & Westerlund, M. (2013). Linking living lab characteristics and their outcomes: Towards a conceptual framework. Technology Innovation Management Review, 3(12).
40. Vontas, A., & Protogeros, N. (2009). Evaluating living labs core competences and assets. 3rd IEEE International Conference on Digital Ecosystems and Technologies, 558–562. https://doi.org/10.1109/DEST.2009.5276772
41. Von Wirth, T., Fuenfschilling, L., Frantzeskaki, N., & Coenen, L. (2019). Impacts of urban living labs on sustainability transitions: Mechanisms and strategies for systemic change through experimentation. European Planning Studies, 27(2), 229-257.

Appendix E: Research Agenda

The following appendix is the finalized research agenda from the Delphi survey and workshops on evaluations of LL effectiveness and impacts. Each research theme is further categorized into subthemes, synthesis questions developed from raw questions in the Delphi workshop, and a description of these developed synthesis questions.

Theme	Sub-theme	Synthesis question	Description
WHO: The role and diversity of actors in the evaluation / QUI: Le rôle et la diversité des acteurs dans l'évaluation	Role of the different actors / Rôle des différents acteurs	What are the right conditions for each category of actors to participate in the evaluation of living labs?	Consider the different categories of actors (including but not limited to stakeholders and rights holders, participants and partners from public and private sectors), as well as the role and barriers for each category or group in the processes of living labs and their evaluation.
	Differentiated actors involvement / Implication différenciée des acteurs	What forms and which moments of evaluation are most conducive to including actors in the process?	Given that there are multiple phases of evaluation, the involvement and influence of different actors may differ at different stages.
		How do and can evaluations take into account differing needs and priorities of actors who work within different timelines and timescales?	Define the concept of timeline, timescale and temporalities, and how these relate to the various actors involved in the evaluation (e.g., evaluators and participants).
	Role of the evaluators / Rôle des évaluateurs	What issues are tied to the different perspectives and positions of evaluators?	Evaluators may qualify their theoretical perspectives and their position in the living labs and evaluation processes. Positions of evaluators can be considered as internal, external,

			both or other.
	Diversity of actors / Diversité des acteurs	What types of diversity are important and can be measured in the evaluation of living labs?	Diversity can be tied to attributes of the individual, but also to the context within which they were trained and worked. Examples of types of diversity can include gender, culture, experience, community representation, discipline, sectors, among others.
		How can the contributions of non-humans, as participating actors of living labs, be evaluated?	Non-humans (e.g. soil, agricultural infrastructure, plants, animals, water) may be considered as participating actors and not only passive objects.
	Equity and power relations / Équité et relations de pouvoir	How can it be ensured that representation and power are balanced between the different actors in the evaluation process?	Representation of different actors can be balanced in the evaluation process. There may be attempts to distribute power and decision-making to make the process equitable and representative.
		How does the process of evaluation and its use influence the balance of relationships and power among actors? How can it be taken into account?	The evaluation process itself, such as the choice of indicators and reporting metrics, may influence the behavior of actors, power dynamics and structural (in)equity within the evaluation and living labs.
HOW MUCH: Conditions for success / COMBIEN: Les conditions de réussite	Definition and measure of success / Définitions et mesure de succès	What dimensions influence the success of a living lab?	Many different dimensions and perspectives can be defined and measured when thinking of success in a living lab, including but not limited to social, economic, temporal,

			procedural, institutional and organizational dimensions.
		How can the diversity of definitions of success among actors be considered in the evaluation of living labs? How do these definitions influence evaluation?	Actors may have different definitions and ways to measure success, which can influence the evaluation of living lab processes, outcomes and impacts.
	Conditions for success / Conditions pour le succès	What are enabling conditions for effective and successful living labs?	Various short and long term conditions may enable success of living lab initiatives, including but not limited to willingness to change, certain investments or collaborative approaches.
		What participant characteristics enable effective and successful living labs?	Specific social and psychological characteristics of participants may play a key role in enabling success and effectiveness of a living lab.
	Role of mistakes and failure / Le rôle des erreurs et échecs	What are the roles of mistakes and failure in developing, implementing and evaluating living labs?	Trial and error, as well as failure, are key to the living lab approach and evaluation of living labs, though some actors may be averse to failure (e.g., funders).
WHY: The use and funding of evaluation / POURQUOI: L 'utilisation et le financement de l'évaluation	Why evaluate / Pourquoi évaluer	How can the different objectives and interests of actors be considered and integrated in the evaluation of living labs?	Reflect on why the evaluation is being conducted, as multiple objectives and different interests in living labs may be conflicting at times (e.g., evaluate adaptive learning, validate desired outcomes).
	Use of evaluation results / Utilisation des résultats de l'évaluation	What are different uses and applications of living labs evaluations in diverse contexts?	Results of evaluation can be used in different ways (e.g., to benefit different actors) and in different contexts (e.g., social,

			economic).
		How can evaluation itself influence the process and results of the living lab?	Evaluation and its results may influence living labs, for example by influencing its evolution, generating best practices and/or enabling knowledge sharing across sectors.
	Funding / Financement	What are the current and potential influences of funding contexts on living lab impacts and evaluations?	Funding (its timelines and agenda) can influence the success of living labs as well as the evaluation process, may require tradeoffs and could be leveraged to improve the evaluation and the impacts of living labs.
HOW: Methods and tools for evaluation / COMMENT: Les méthodes et outils d'évaluation	Methods / Méthodes	How can a common methodology and vision for the evaluation of living lab processes and impacts be established?	Common evaluation methods and frameworks for all living labs, as well common visions for each living lab, may be established at the start of the process.
		What are the strengths and limitations of different methods to evaluate processes and impacts of living labs?	Different methods may be used to evaluate Living Labs, but it is not always clear what their roles and effects are. Examples include co-construction, qualitative methods and methods tied to social-ecological systems.
		How might existing frameworks from other fields be used to evaluate the "building blocks" of living labs across sectors and contexts?	Multiple methods and frameworks from different fields may be used to evaluate the “building blocks” of living labs (e.g., infrastructure, level of openness, real-world

			context, etc.) across sectors and contexts (e.g. health versus agriculture).
	References / Référentiels	How can a collection of references and practical tools support the evaluation of living labs and the inclusion of new actors?	Various references and guides available for the evaluation of living labs or similar models may play different roles in evaluation. A collection of references and tools like academic papers and handbooks may take various forms (e.g., repositories, archives or inventories, toolkits).
		How can evaluation support improved understanding of the different points of reference of actors in living labs?	Actors in living labs may have different disciplinary, political , social and cultural points of references as they come from a wide range of disciplines, sectors and contexts.
	Perspectives / Perspectives	What are the roles of subjectivity and objectivity in the different evaluation processes of living labs?	Subjectivity and objectivity are in tension in living labs and may influence evaluation processes (e.g. self-evaluation, comparison and setting objectives).
	Trust / Confiance	What role do trust and willingness to share data play in the evaluation of living labs?	The relationship between trust and willingness to share data (e.g., does sharing data strengthen trust? does trust imply willingness to share?) may play a role in the evaluation process.
	Comparison / Comparaison	How does the evaluation of living labs compare with evaluation of other approaches?	Living labs are distinct but share commonalities with other approaches; some lessons from other fields or models may

			apply to the evaluation of living labs and vice versa.
		What methods, metrics, and criteria of evaluation for living labs are needed to compare between projects, sectors, contexts, specific processes and overall approaches?	Evaluation methods and tools (e.g., metrics, indicators, measurements) may enable comparison and transfer of knowledge between living labs and other approaches, and among living labs from different sectors, different contexts, or which use distinct processes (e.g. centralized versus decentralized, creation versus validation).
WHERE: The scale of evaluation / OÙ: L'échelle de l'évaluation	Scale / Échelle	Which evaluation methods and tools should be used at which scales?	Evaluation of living labs may occur at different scales, understood as a complex and multidimensional concept which includes temporal scales, social levels (micro, meso, macro) and location (geography, institutions).
	Integration / Intégration	How can evaluation methods and tools be integrated within and across scales to generate a holistic understanding of living lab processes, outcomes and impacts?	Social, environmental and social-environmental impacts and outcomes may occur at various scales and timelines, for example in regional and national networks and among the different layers of living labs.
WHEN: The temporality of evaluation / QUAND: La temporalité de l'évaluation	Temporality / Temporalité	What are the evaluation methods and tools specific to the different stages of evaluation?	Different methods, tools, and criteria can be used at different stages of the evaluation process. These may contribute to the growth of living labs, and there may be a range of views among participants around the expected length of each stage.

	Measurement / Mesure	How can the different dimensions of Living Labs be measured at each stage?	Different dimensions of living labs (activities, process, products, outcomes, impact) may be evaluated over time. Specific examples include sustainability, governance, innovation and the projects themselves.
	Alignment / Alignement	What is the alignment between the timelines of living labs, funders, users and communities? What are the impacts of mismatches?	Alignment or mismatch between the timelines of different actors and processes (e.g., evaluation, funding, management activities and cycles of living systems like crops) may influence living labs processes and evaluation.
	Evolution / Évolution	How do behaviours, perspectives, and knowledge of living lab actors change over time?	The views, values, knowledge and behaviors of actors may evolve over the course of living lab processes and evaluation.
WHAT: The objects of evaluation / QUOI: Les objets d'évaluation	Impacts in general / Impacts en général	What methods are needed to evaluate the impact of living labs?	Different impacts (e.g. innovations, practices or living labs processes) can be evaluated with different methods that may contribute to the development of living labs.
		How could mixed methods design provide tools for the evaluation of living lab outcomes?	Mixed methods designs may support the evaluation of practices and impacts of living labs. For example, systems thinking or network analysis could be combined with qualitative or participatory methods.
		How do specific mechanisms of living labs relate to the various types of innovation adoption?	Different participatory approaches and mechanisms may promote different types of innovation adoption and may

			reveal different strategies and system dynamics.
	Process / Processus	How can impact goals be developed? How can cost-benefit analysis be used to outline realistic project goals?	Cost-benefit analyses and impact goals may improve alignment and sustainability of living labs and their evaluation. For example, cost-benefit analyses could reveal if and how the cost of a living lab influences its environmental goals.
		How can concerns for efficiency in co-creation be balanced in living labs and their evaluation? What are the trade-offs?	Assumptions about the need to compromise between the burden of co-creation (resource intensive, e.g. time and money) and the efficiency of participatory approaches may influence the evaluation of living labs.
		How can participation in the evaluation of living labs be encouraged? How can tensions related to the lack of willingness to participate be overcome?	Actors may have different levels of willingness to participate in living labs processes and evaluation. For example, tensions may arise around lack of willingness to participate.
	Social impacts / Impacts sociaux	How does the value placed on social impact differ according to the specific characteristics of the actors in a living lab?	Actors with different characteristics (e.g. type, diversity, benefits sought, etc.) may hold differing views and values about social impacts. These values may change over time and be influenced by living labs processes.

		What key dimensions of evaluation can capture the social impacts of living labs?	Multiple dimensions may be considered when evaluating the social impacts and outcomes of living labs. These include organizational transformation, the evolution of relationships between actors and groups, dynamics within social systems and innovation ecosystems, among others.
		What are measures and indicators of social impacts for actors with varying levels of involvement and for different scales of living labs?	Appropriate indicators to investigate the social impact of living labs may include well-known indicators of social impacts in other fields. Social impacts may differ according to scales (e.g., time, geography and individual, community, or global level) and for actors with different levels of involvement (e.g., participants, partners, facilitators).
		What are the best methods to evaluate specific social processes and outcomes of living labs?	Specific social processes and outcomes of living labs may include influence on mental health and well-being, transfer and acquisition of skills and knowledge, relationships between actors, collaboration, trust, willingness to participate, data-sharing and behavior change, among others.
		What are the connections and mutual influences between key dimensions of living labs? How can these connections be established and influenced?	Impacts of living labs are complex and multi-dimensional, and there may be different connections and mutual influences between key elements of living labs (e.g., social impacts and

			sustainability, diversity of actors and social impacts).
		What key criteria of best management practices can be measured and compared? How can we operationalize these criteria?	Key criteria to measure and compare best management practices (e.g.. environmental sustainability, cost) may be identified and operationalized in the evaluation of living labs.
		What indirect impacts and outcomes do living labs have on individuals, groups and society?	Living labs can have indirect impacts (e.g., social learning, social impact, knowledge sharing). Evaluation of these indirect impacts may be inclusive of different actors.
	Environmental impacts / Impacts environnementaux	How is sustainability defined in the evaluation of living labs?	Definitions of sustainability may be considered in the evaluation of living labs, along with the measurement of environmental outcomes and impacts at different scales (e.g., carbon soil sensors, ecosystem service indicators, sustainable development goals).
	Social-environmental impacts / Impacts socio-environnementaux	How are social, environmental and socio-environmental impacts defined in the evaluation of living labs?	Social, environmental and social-environmental impacts of living labs may be defined and measured with indicators to support evaluation.
		How can social and ecological impacts of living labs be assessed simultaneously?	Some methods may simultaneously assess the interrelated impacts of living labs across communities and ecosystems (e.g., ties between social, economic and environmental impacts).

		What are the qualitative approaches used to measure social, environmental and socio-environmental impacts?	Qualitative and technology-oriented approaches may have distinct methods, with different strengths and limitations, for the evaluation of social impacts of living labs.
		How do living labs and their evaluation relate to solving wicked problems?	Living labs and their evaluation may play a role in helping understand and solve complex issues (e.g., water quality, lack of social-ecological resilience, complex networks and communications, innovations in gene technology).