Review

Moving toward Generalizability? A Scoping Review on Measuring the Impact of Living Labs

Kelly Bronson [1], Rachana Devkota [1,*] and Vivian Nguyen [2]

[1] School of Sociological and Anthropological Studies, Faculty of Social Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada; kbronson@uottawa.ca

[2] Interdisciplinary Science and Practice Program, Institute of Environmental and Interdisciplinary Science, Carleton University, Ottawa, ON K1S 5B6, Canada; VivianNguyen@cunet.carleton.ca

* Correspondence: rdevkota@uottawa.ca

Abstract: The living labs (LLs) approach has been applied around the globe to generate innovation within and suited to real-life problems and contexts. 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. This article seeks to verify and address this knowledge gap by conducting an adopted scoping review method which uses a combination of tools for text mining alongside human text analysis. In total, 138 academics literature were screened, out of which 88 articles were read in full and 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. The dominant method for evaluation used in the literature is comparative qualitative using case studies. This study 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.

Keywords: living labs; evaluation; impact; environment; agriculture; sustainability; scoping review

1. Introduction

The living labs (LLs) approach has been applied around the globe to generate innovation within and suited to real-life problems and contexts. While the living lab model was started in the late 1990s, its significant application has increased only from 2006, when the European Commission launched a European Network of Living Labs (ENoLL) as part of its policy to improve competitivity [1,2]. LL research and practice has grown alongside the acceptance of collaborative and transdisciplinary approaches as effective for addressing complex problems, specifically when dealing with transitions to sustainable, resilient, and adaptive societies [3–5].

Living Labs (LLs) are a mechanism or approach that brings a diversity of stakeholders together to arrive at user-centric solutions and innovations and thus they could present a viable method for solving complex issues. Proponents of the LL approach suggest that it can increase the likelihood that innovations will meet users' needs and thus lead to technologies or practices which are adopted more quickly and widely. LLs have been used to innovate practices and tools across sectors including health care, urban planning, application design, service delivery and information management and technology [2,5]. In terms of environmental and agricultural sustainability, specific living lab studies have applied the approach to climate change adaptation and sustainable natural resource management [6]. In Canada, there are notable examples of agricultural sustainability LLs at the reginal AcadieLab (https://www.rang3.org/le-labo) and national scale. Beginning in 2018,

Agriculture and Agri-Food Canada (AAFC) launched its Living Laboratories Initiative (http://www.agr.gc.ca/livinglab) which is a large-scale application of the LL approach within an agroecosystem context.

Despite the promise of the LL approach, a few studies have suggested that there is a gap in practitioner and academic community knowledge surrounding how to measure and evaluate both the performance of LL processes and their broader impacts; it appears that this gap is even more pronounced when it comes to living labs aimed at agricultural or environmental sustainability [2,6,7]. Ballon et al. [2] (p. 1203) emphasize the need to "start evaluating thoroughly the effectiveness and impact of specific living lab experiences. However, while most scholars and practitioners appear to agree on this, no systematic impact studies of living labs exist up until this day". Additionally, Hossain et al. [6] wonder to what extent LLs focused on sustainability have received adequate academic attention? This paper has two goals. First, the paper aims to verify that a gap exists pertaining to (a) metrics, mechanisms and frameworks for evaluating and measuring the effective functioning of a LL (e.g., effectiveness of collaborations and transdisciplinary governance structure), as well as (b) the LL's longer-term impacts on society and environment. Our paper places specific attention to the possible gap vis-à-vis measurement among LLs designed for environmental or agricultural sustainability. Second, our paper aims to synthesize any existing common practices for evaluation. The major research question that guided this study was "What, if any, general evaluation methods, metrics or frameworks exist for measuring the effectiveness of LLs in general, and then among those specific to environmental and agricultural sustainability?" We present here the results of our scoping review of academic literature on living labs wherein we ultimately find limited studies of agricultural or even sustainability focused LLs which discuss measurement. We also find no universally applied and widely accepted method or framework for evaluation across our dataset. Indeed, the most common method for evaluation among the articles in our dataset is the case study, and effectively no replicability in frameworks used. Our final argument is that a tension exists in the LL literature and practice between the local and site-specificity of LLs and the seeming need for a more universal framework that could be used to evaluate LL projects against one another, thus moving this domain beyond the particular. Our paper ends by synthesizing those frameworks for evaluation which exist in order to direct future research that might develop a universal framework for evaluating LLs.

2. Methodology

To answer our above research question, we used an adopted scoping review method of available peer-reviewed literature [8]. The PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist was used for the initial screening (Figure 1). It is important to note that the study did not perform meta-analysis of included articles as these steps are not mandatory for a scoping review [9]. Additionally, we conducted text mining within using an automated computer tool called Voyant (https://voyant-tools.org/). Voyant Tools is an open-source, web-based application designed for text mining [10], which was developed by Stéfan Sinclair (McGill) and Geoffrey Rockwell (University of Alberta) [11]. 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 [10,12].

2.1. 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 multi-disciplinary academic articles. Additionally, they contain both natural and social science

articles and they are known to pull a sub-set 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.
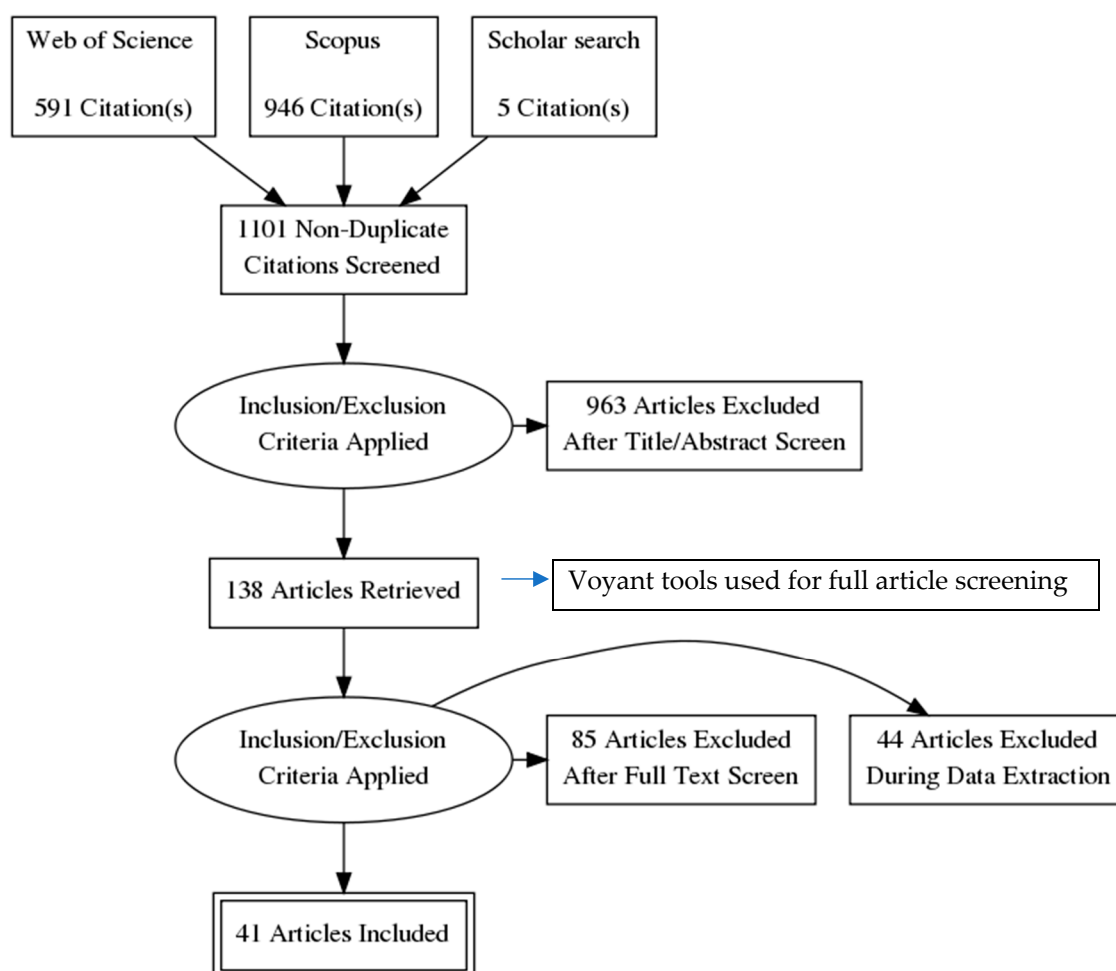

Figure 1. Flow diagram of scoping review using Prism-ScR checklist in this study.

Table 1. Finalized Search String Using Boolean Operators.

Concepts	Search Terms
(1) LLs	"living lab*"
	AND
(2) Evaluation	(evaluat* OR performance OR effective* 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 search process while two academic researchers were involved in verification of the articles.

2.2. Comprehensiveness of Searches

Deliberation within our team of researchers and benchmark papers (Supplementary Materials 1) were used to test the comprehensiveness and validity of the search strategy. Selected benchmark papers include academic literatures that were mostly published in Journals, books, and Proceedings.

2.3. Article Screening and Study Eligibility Criteria

2.3.1. 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.

2.3.2. 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 [13].

2.3.3. Eligibility Criteria

Articles were screened for inclusion or exclusion using eligibility criteria at each phase, as outlined in Supplementary Materials 2. 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). Articles screened for Include Second Opinion were further deliberated by the research team for either inclusion/exclusion.

2.3.4. 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

(Supplementary Materials 2). 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. The data coding sheet can be found in Supplementary Materials 3. 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 theme which emerged across the dataset after reading the full text. These are summarized below.

3. Results and Discussion

3.1. Overview of Results

This scoping review resulted in 41 articles that are relevant to measuring the impact of living labs. 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 living labs 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).
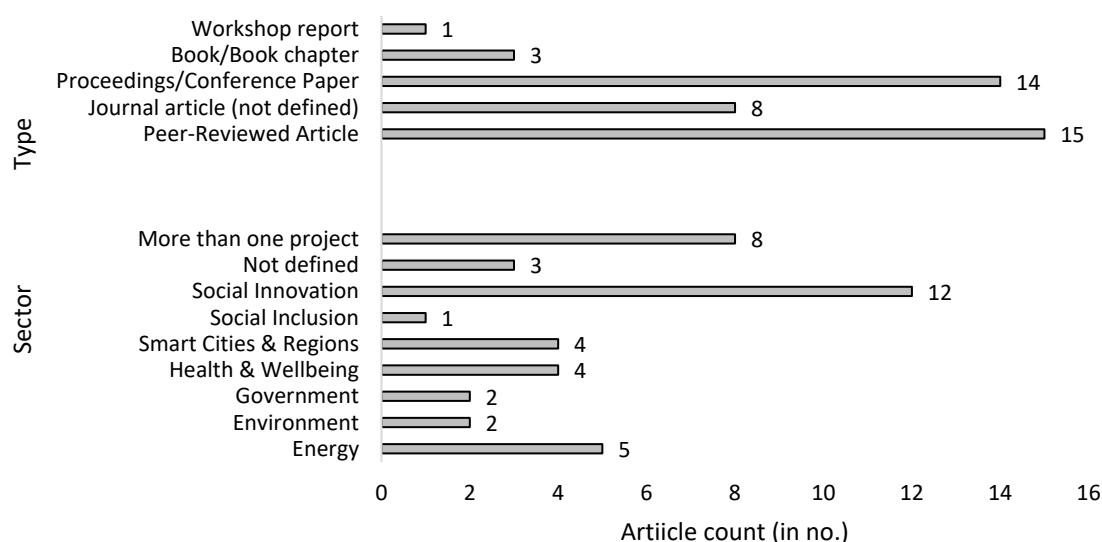

Figure 2. Type and sector relevant details on articles on Living Labs (LLs) literature (in no.).
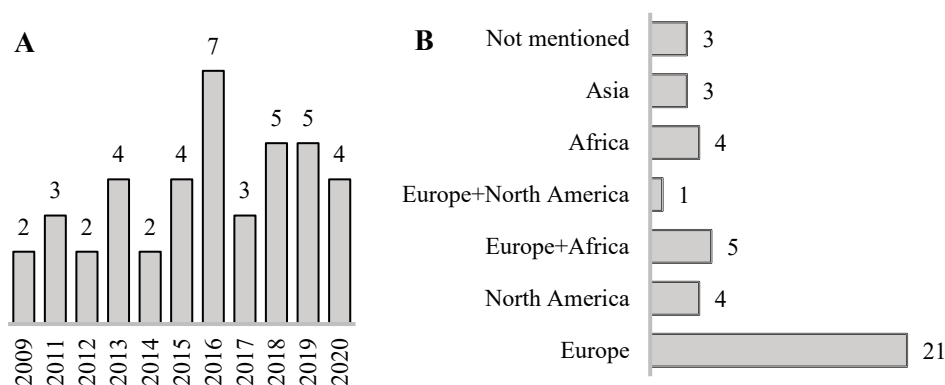

Figure 3. The total articles included in the scoping review process. Figure (**A**) shows the total articles published (in no.) by year and (**B**) shows the countries involved in publishing articles on living labs (in no.).

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/) which has expanded across Europe, rising from 20 to over 440 living labs in between 2007 to 2020, we feel this may explain the dominant presence of European publications in this sector [5,6,14].

3.2. 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 [5].

Few studies on LLs and sustainability are available [5,15–20], and even fewer of these focused on measurement/impact [16,21]. 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 (Figure 4). It is obvious from the literature that agri-ecosystem 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 [22]. 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 5 and 6).
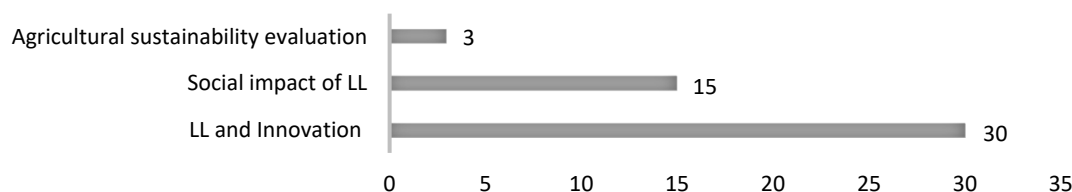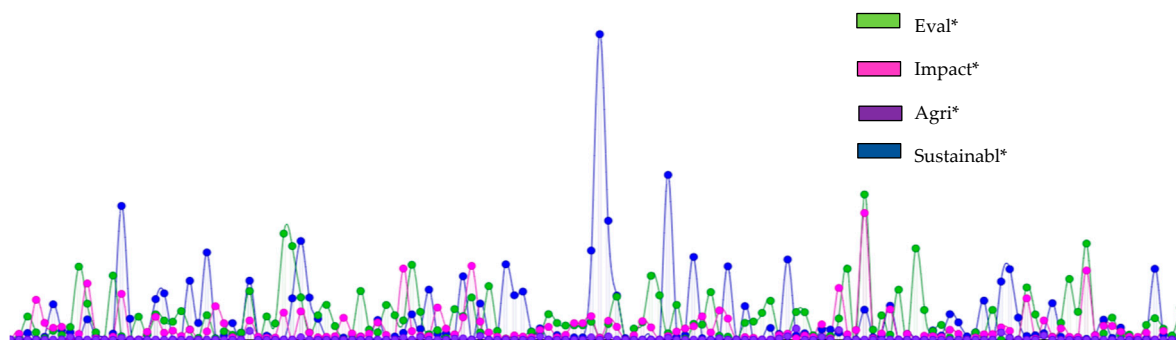

Figure 4. No. of articles published on living lab focusing on innovation, social impact and agricultural sustainability, Articles are not mutually exclusive. Source: Scoping review.
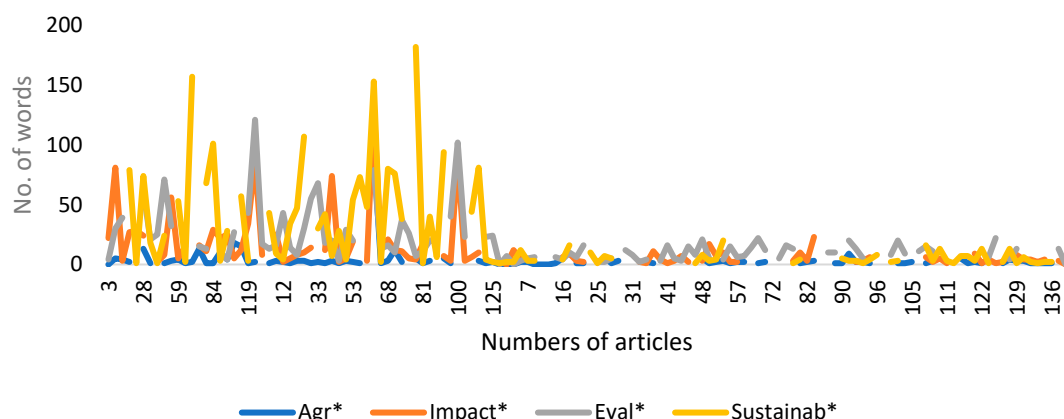

Figure 5. Frequency of terms in articles included in this review. Figure generated from excel with manual counting.

Figure 6. Frequency of words (Eval*, Agri*, Impact* and Sustainabl* in 138 articles included in this study. Figure generated using Voyant.

The reason for the plurality of evaluation methods likely has to do with the fact that LLs are by definition user-driven, thus evaluation approaches are guided by different organizations, agencies and stakeholder groups depending on the location and specific mandate of the LL [23]. 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%) while 78 percent of studies conducted an evaluation study after the LL project was completed (Figure 7) and based upon the specific goals of those particular projects.
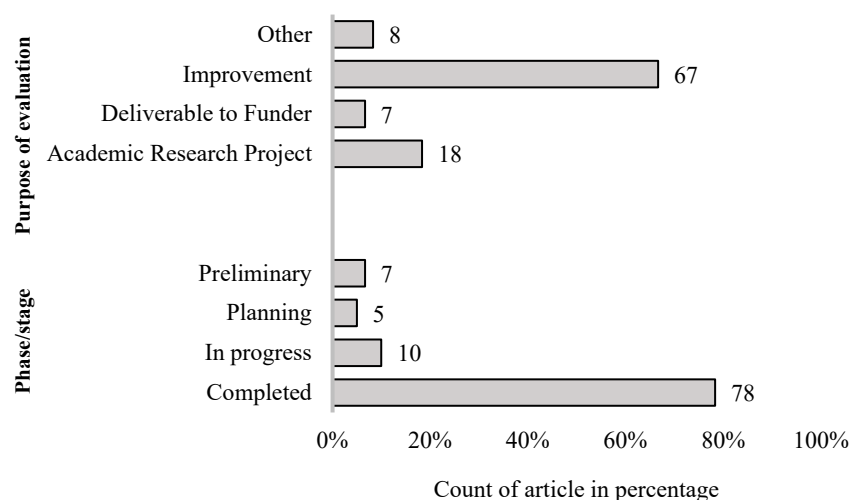
Figure 7. Shows the different phase/stage of evaluation and purpose of evaluation for research studies included in this scoping review.

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 8) shows the snapshot of methods used in evaluating LLs among our datasets. The most common methods used in LL evaluation are discussed below.
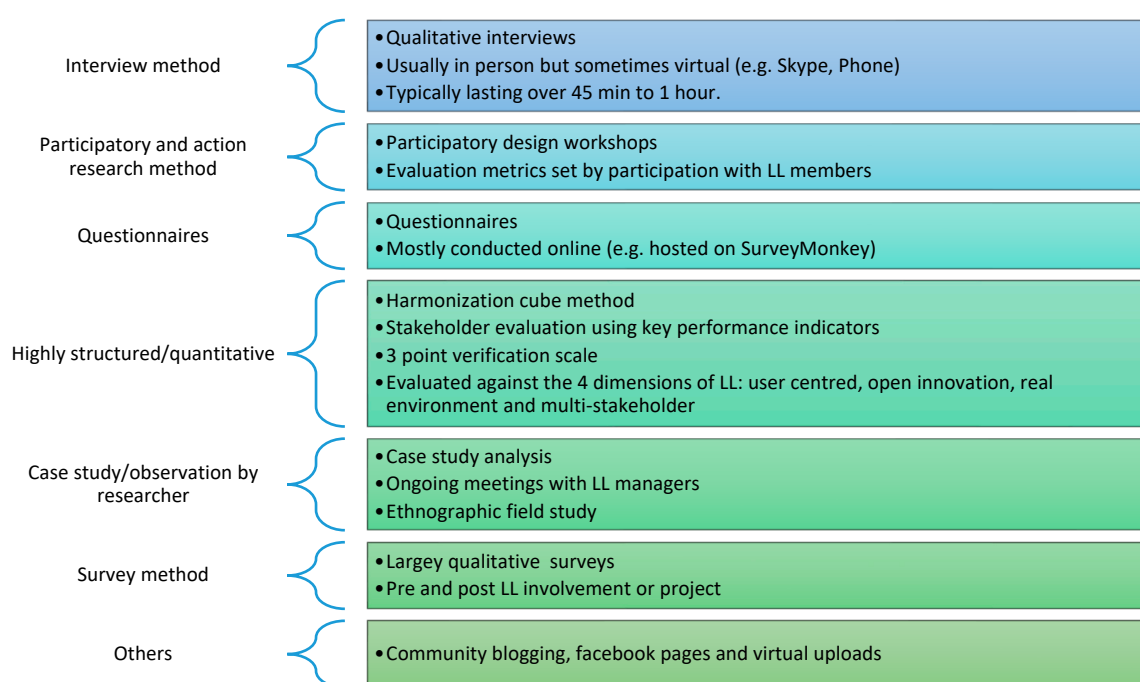
Figure 8. Summary of different evaluation approaches used in LLs studies which discuss evaluation.

3.2.1. Case Study Analysis and Action Research for LL Evaluation

Our scooping 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 [24] (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. [25] 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 [26], 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. [27] (p. 1) 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 evaluation of LLs [27].

Another key insight from our review is that employing a diversity of tools is a key attribute of LL evaluation, for example studies use self-reported methods (for e.g., diaries, experience sampling) alongside researcher measurement (e.g., observation, ethnography) [27,28].

3.2.2. Qualitative Evaluation Tools

Structured and systematic evaluation methods exist in the literature, but they are under-represented [26]. As such, our review shows that qualitative research methods are most commonly used for evaluating LLs (Figure 9). Some noteworthy studies which detail their qualitative methods used in LLs assessment are Callari et al. [29], Cech & Wagner [30], and Georges et al. [31]. Within those qualitative methods used, participatory design but also workshops and open-ended qualitative interviews were the most common methods deployed (Figure 9). Figure 9 was generated from the full article screening and indicates the high number of word counts returned from our corpus (138 articles) for the terms "participatory", "workshop", and "qualitative". Around 74 percent of articles reviewed in this study used participatory action research, workshops, email surveys, phone surveys and semi-structured questionnaires. Around 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 [29]. 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 and Sirkka, [32]). Very few studies (see Callari et al. [29] used a deductive, concept-driven coding frame to analyze interview transcripts.
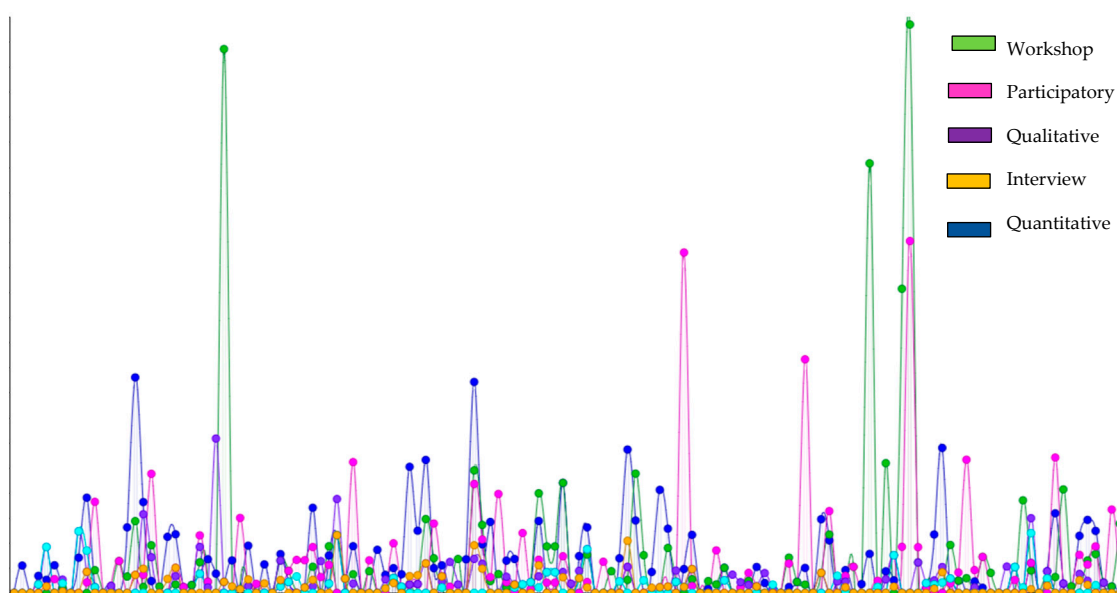
Figure 9. Word counts using Voyant ["interview (dark blue color), workshop (green), participatory (pink), qualitative (purple), quantitative (light blue), and interview (yellow)].

3.2.3. Quantitative Evaluation Methods/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 and Chou [33], Hagy et al. [34]). Moreover, these quantitative methods are mostly combined with qualtitative methods (as a mixed method of data collection).

One study by Dell'Era et al. [35] 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 [35]. 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 Living Lab 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 [33] 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. [2] recommended a logit model to measure the effectiveness of involving users in digital innovation process. Something similar to this logit model is called "Reference Model" which is recommended by Guzmán et al. [36] for user-driven innovation assessment that is highly structured. Kovacs [37] used an "Alcotra and Harmonization cube method" to evaluate the interactive value production coming from LL. Maciuliene and Skaržauskiene [38] 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 & Protogeros [39] 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.

3.2.4. Evaluation Methods for LLs Specifically Related to Agri-Ecosystems

Out of 41 final articles which we found relevant to LL and evaluation, only two study were found relevant to agri-ecosystems and sustainability (Ondiek and Moturi [21], and Hagy et al. [34]). For instant, Hagy et al. [34] (p. 18) did a study on an innovation agro-ecosystem 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 [34], 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 agro-ecosystem 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.

3.2.5. Evaluation Frameworks for LLs

There are several approaches to LL evaluation which have been studied for decades [23,40]. For example, the World bank and UN have their own project and program evaluation guidelines to follow while evaluating the outcome of the technological processes or programs. Similarly, the Rapid Impact Evaluation method, developed by Dr. Andy Rowe in 2004, is used by Government of Canada for evaluating its project and program (www.canada.ca). 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. [25], Schuurman et al. [28]). However, our review of 138 articles on LLs indicated

that there is 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. [2], Guzmán et al. [36], Kovacs [37], Mačiulienė and Skaržauskienė [38], Osorio et al. [41], Schuurman et al. [28], Schuurman et al. [25]). 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 the uniformity in research process/study. 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 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 [21] 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 need of the users, and LLs' efficiency, effectiveness, utility and sustainability aspect of the LLs are some of the important factors to be accounted while evaluating any LLs [21,42].

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 [43,44]. For example, von Wirth et al. [44] assessed the initial strengths and weaknesses of the living labs in their study and proposed a set of practices which were believed to support the living labs 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 living labs 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 [21] employed the four-capital method of sustainable development evaluation framework recommended by Ekins et al. [42] 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. [42] argued that this model is helpful in showing the relationships between key elements of projects in describing how sustainable development can be realized.

Stahlbrost [7] 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 [26] (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?".

Table 2. Summary of relevant evaluation frameworks and model used in LLs evaluation literature.

Evaluation Framework/Principles/Model	Key Focus	Key Elements	Authors
Digital Co-Creation Index (DCCI) framework for evaluation in 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 [38]
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 [21]
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. [35]
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. [2]
A maturity grid-based assessment tool	Framework developed by reviewing eight frameworks that focus specifically on innovation laboratories	Guidance tool to evaluate the maturity degree of an innovation laboratory or to adapt an existing LL project	Osorio et al. [41]
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.	Kovacs [37]
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. [45]
Living Lab triangle- Conceptual framework	Linking living lab environments, approaches and innovation outcomes to each other.	For data coding, parameters were divided on 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. [46]
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. [47]

Table 2. *Cont.*

Evaluation Framework/Principles/Model	Key Focus	Key Elements	Authors
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 [7]
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. [36]
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 [33]
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 [39]
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 [48]

4. Conclusions and Future Perspectives

We initially set out in this review paper to verify that a gap exists within living laboratory 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-driven, thus evaluation approaches are guided by different organizations, agencies and stakeholder groups depending on the location and specific mandate of the LL [23,49,50]. 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. [29], Cech and Wagner [30], and Georges et al. [31].

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 [42] also highlights the need for a unified approach to evaluating LLs−one which might guide in comparing multiple cases by using common indicators. 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.

Additionally, our review uncovered very few articles on agricultural and environmental sustainability and within those even fewer that measure impacts. LLs focusing on social innovation, environment and/or sustainability used qualitative methods of evaluation such as participatory design workshops, semi structured interviews, focus group discussion, email surveys and ethnographic studies. These evaluation methods are unstructured and inductive in nature. This may be due to the fact that social innovation, rural innovations, and the environment are complex subjects that need more academic attention to arrive at a structured evaluation framework.

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 [49,50]. Future work could develop a unifying framework for evaluating sustainability LLs by focusing on three key elements synthesized from best practices to date: (1) level of participant involvement and empowerment, (2) time-series analysis and (3) long-term viability of the LL project.

Supplementary Materials: The following are available online at https://www.mdpi.com/2071-1050/13/2/502/s1, List of articles S1: Benchmark Articles on LL for search process, Checklist S2: Data coding sheet, Screening outline S3: Eligibility criteria for exclusion or inclusion of articles.

Author Contributions: Conceptualization, V.N. and K.B.; methodology, K.B. and R.D.; software, R.D.; validation, R.D. and K.B.; formal analysis, R.D. and K.B.; investigation, V.N.; resources, V.N. and K.B.; data curation, V.N. and K.B.; writing—original draft preparation, R.D. and K.B.; writing—review and editing, K.B. and V.N.; visualization, R.D., K.B. and V.N.; supervision, V.N. and K.B.; project administration, V.N.; funding acquisition, V.N. and K.B. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schuurman, D.; de Marez, L.; Ballon, P. Living Labs: A Systematic Literature Review. In Proceedings of the Open Living Lab Days 2015, Istanbul, Turkey, 25–28 August 2015.
2. Ballon, P.; van Hoed, M.; Schuurman, D. The Effectiveness of Involving Users in Digital Innovation: Measuring the Impact of Living Labs. *Telemat. Inform.* **2018**, *35*, 1201–1214. [CrossRef]
3. Pohl, C. Transdisciplinary Collaboration in Environmental Research. *Futures* **2005**, *37*, 1159–1178. [CrossRef]
4. Boon, W.P.C.; Chappin, M.M.H.; Perenboom, J. Balancing Divergence and Convergence in Transdisciplinary Research Teams. *Environ. Sci. Policy* **2014**, *40*, 57–68. [CrossRef]
5. Gamache, G.; Anglade, J.; Feche, R.; Barataud, F.; Mignolet, C.; Coquil, X. Can Living Labs Offer a Pathway to Support Local Agri-Food Sustainability Transitions? *Environ. Innov. Soc. Transit.* **2020**, *37*, 93–107. [CrossRef]
6. Hossain, M.; Leminen, S.; Westerlund, M. A Systematic Review of Living Lab Literature. *J. Clean. Prod.* **2019**, *213*, 976–988. [CrossRef]
7. Ståhlbröst, A. A Set of Key Principles to Assess the Impact of Living Labs. *Int. J. Prod. Dev.* **2012**, *17*, 60–75. [CrossRef]

8. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [CrossRef]

9. Munn, Z.; Peters, M.D.J.; Stern, C.; Tufanaru, C.; McArthur, A.; Aromataris, E. Systematic Review or Scoping Review? Guidance for Authors When Choosing between a Systematic or Scoping Review Approach. *BMC Med. Res. Methodol.* **2018**, *18*, 143. [CrossRef]

10. Miller, A. Text Mining Digital Humanities Projects: Assessing Content Analysis Capabilities of Voyant Tools. *J. Web Librariansh.* **2018**, *12*, 169–197. [CrossRef]

11. Sampsel, L.J. Voyant Tools. *Music Ref. Serv. Q.* **2018**, *21*, 153–157. [CrossRef]

12. Hetenyi, G.; Lengyel, A.; Szilasi, M. Quantitative Analysis of Qualitative Data: Using Voyant Tools to Investigate the Sales-Marketing Interface. *J. Ind. Eng. Manag.* **2019**, *12*, 393–404. [CrossRef]

13. McHugh, M.L. Interrater Reliability: The Kappa Statistic. *Biochem. Med.* **2012**, *22*, 276–282. [CrossRef]

14. Leminen, S.; Westerlund, M. Living Labs: From Scattered Initiatives to a Global Movement. *Creat. Innov. Manag.* **2019**, *28*, 250–264. [CrossRef]

15. Bulkeley, H.; Coenen, L.; Frantzeskaki, N.; Hartmann, C.; Kronsell, A.; Mai, L.; Marvin, S.; McCormick, K.; van Steenbergen, F.; Voytenko Palgan, Y. Urban Living Labs: Governing Urban Sustainability Transitions. *Curr. Opin. Environ. Sustain.* **2016**, *22*, 13–17. [CrossRef]

16. Voytenko, Y.; McCormick, K.; Evans, J.; Schliwa, G. Urban Living Labs for Sustainability and Low Carbon Cities in Europe: Towards a Research Agenda. *J. Clean. Prod.* **2016**, *123*, 45–54. [CrossRef]

17. Keyson, D.V.; Al Mahmud, A.; Romero, N. *Living Lab and Research on Sustainability: Practical Approaches on Sustainable Interaction Design*; Springer: Cham, Switzerland, 2013; Volume 8309, pp. 229–234. [CrossRef]

18. Burbridge, M. If Living Labs Are the Answer—What's the Question? A Review of the Literature. *Procedia Eng.* **2017**, *180*, 1725–1732. [CrossRef]

19. Puerari, E.; de Koning, J.; Mulder, I.; Loorbach, D. *Shaping Spaces of Interaction for Sustainable Transitions*; Universidade de Lisboa: Lisboa, Portugal, 2017.

20. Schliwa, G.; Evans, J.; McCormick, K.; Palgan, Y.V. 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, 12–13 March 2015; pp. 1–21.

21. Ondiek, M.A.; Moturi, C. An Assessment of the Sustainability of Living Labs in Kenya. *Innov. Manag. Rev.* **2019**, *16*, 391–403. [CrossRef]

22. AgriLink. AgriLink Home—AgriLink. Available online: https://www.agrilink2020.eu/ (accessed on 26 November 2020).

23. Thamhain, H.J. Contemporary Methods for Evaluating Complex Project Proposals. *J. Ind. Eng. Int.* **2013**, *9*, 1–14. [CrossRef]

24. Yin, R. *Case Study Research: Design and Methods*; Sage Publications: Beverly Hills, CA, USA, 1984; Volume 5.

25. Schuurman, D.; de Marez, L.; Ballon, P. The Impact of Living Lab Methodology on Open Innovation Contributions and Outcomes. *Technol. Innov. Manag. Rev.* **2016**, *6*, 7–16. [CrossRef]

26. Van Geenhuizen, M. Article Politics and Space A Framework for the Evaluation of Living Labs as Boundary Spanners in Innovation. *Politics Space* **2018**, *36*, 1280–1298. [CrossRef]

27. Hofte, H.; Jensen, K.L.; Nurmi, P.; Froehlich, J. Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild. In *Proceedings of the MobileHCI09—The 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, Bonn, Germany, 15–18 September 2009*; ACM Press: New York, NY, USA, 2009; p. 2. [CrossRef]

28. Schuurman, D.; Baccarne, B.; Kawsar, F.; Seys, C.; Veeckman, C.; de Marez, L.; Ballon, P. Living Labs as Quasi-Experiments: Results from the Flemish LeYLab. In Proceedings of the XXIV ISPIM Conference: Innovating in Global Markets: Challenges for Sustainable Growth, Helsinki, Finland, 16–19 June 2013.

29. Callari, T.C.; Moody, L.; Saunders, J.; Ward, G.; Holliday, N.; Woodley, J. Exploring Participation Needs and Motivational Requirements When Engaging Older Adults in an Emerging Living Lab. *Technol. Innov. Manag. Rev.* **2019**, *9*, 38–49. [CrossRef]

30. Cech, F.; Wagner, M. Erollin' on Green: A Case Study on Eco-Feedback Tools for Emobility. In *ACM International Conference Proceeding Series*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 121–125. [CrossRef]

31. Georges, A.; Schuurman, D.; Desmet, M. Uncovering the Needs and Wants of End-Users towards Green Apps: A Living Lab Approach. In *Proceedings of the 2015 IEEE International Conference on Engineering, Technology and Innovation/ International Technology Management Conference, ICE/ITMC 2015, Belfast, UK, 22–24 June 2015*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2016. [CrossRef]

32. Holappa, N.; Sirkka, A. Living Lab as an Agile Approach in Developing User-Friendly Welfare Technology. In *Harnessing the Power of Technology to Improve Lives*; Cudd, P., de Witte, L., Eds.; IOS Press: Amsterdam, The Netherlands, 2017; Volume 242, pp. 654–659. [CrossRef]

33. Chen, Y.T.; Chou, W.H. Constructing Living Labs Analysis Model for Designing and Evaluating Living Labs Systems. In *Proceedings of the 2010 IEEE International Technology Management Conference, ICE 2010, San Diego, CA, USA, 21–23 June 2010*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2016. [CrossRef]

34. Hagy, S.; Bard, F.; Sasic, A. *Next Generation Living Labs Comprehensive Report*; Chalmers University of Technology and Climate-KIC: Gothenburg, Sweden, 2017. [CrossRef]

35. Dell'Era, C.; Landoni, P.; Gonzalez, S.J. Investigating the innovation impacts of user-centered and participatory strategies adopted by European Living labs. *Int. J. Innov. Manag.* **2019**, *23*, 235. [CrossRef]

36. Guzmán, J.G.; Schaffers, H.; Bilicki, V.; Merz, C.; Valenzuela, M. Living Labs Fostering Open Innovation and Rural Development: Methodology and Results. In Proceedings of the 2008 IEEE International Technology Management Conference (ICE), Lisboa, Portugal, 23–24 June 2008; pp. 1–8.

37. Kovács, K. Evaluation and Practice of Interactive Value Production in Living Labs. *Period. Polytech. Soc. Manag. Sci.* **2016**, *24*, 52–59. [CrossRef]

38. Mačiulienė, M.; Skaržauskienė, A. Sustainable Urban Innovations: Digital Co-Creation in European Living Labs. *Kybernetes* **2020**, *49*, 1969–1986. [CrossRef]

39. Vontas, A.; Protogeros, N. Evaluating Living Labs Core Competences and Assets. In Proceedings of the 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies, Istanbul, Turkey, 1–3 June 2009; pp. 558–562. [CrossRef]

40. Patton, M.Q. *Qualitative Research & Evaluation Methods*, 15th ed.; Sage Publications: Thousand Oaks, CA, USA, 2002.

41. Osorio, F.; Dupont, L.; Camargo, M.; Palominos, P.; Peña, J.I.; Alfaro, M. Design and Management of Innovation Laboratories: Toward a Performance Assessment Tool. *Creat. Innov. Manag.* **2019**, *28*, 82–100. [CrossRef]

42. Ekins, P.; Dresner, S.; Dahlström, K. The Four-Capital Method of Sustainable Development Evaluation. *Eur. Environ.* **2008**, *18*, 63–80. [CrossRef]

43. Hyysalo, S.; Hakkarainen, L. What Difference Does a Living Lab Make? Comparing Two Health Technology Innovation Projects. *CoDesign* **2014**, *10*, 191–208. [CrossRef]

44. Von Wirth, T.; Fuenfschilling, L.; Frantzeskaki, N.; Coenen, L. Impacts of Urban Living Labs on Sustainability Transitions: Mechanisms and Strategies for Systemic Change through Experimentation. *Eur. Plan. Stud.* **2019**, *27*, 229–257. [CrossRef]

45. Mastelic, J.; Sahakian, M.; Bonazzi, R. How to keep a living lab alive? *Emerald Insight.* **2015**, *17*, 12–25. [CrossRef]

46. Veeckman, C.; Schuurman, D.; Leminen, S.; Westerlund, M. Linking Living Lab Characteristics and Their Outcomes: Towards a Conceptual Framework. *Technol. Innov. Manag. Rev.* **2013**, *3*, 6–15. [CrossRef]

47. Guzmán, J.G.; Fernández del Carpio, A.; Colomo-Palacios, R.; Velasco de Diego, M. Living Labs for User-Driven Innovation: A Process Reference Model. *Res. Manag.* **2013**, *56*, 29–39. [CrossRef]

48. Parkinson, S.; Ramirez, R. Using a sustainable livelihoods approach to assessing the impact of ICTs in Development. *J. Community Inform.* **2007**, *2*, 12. [CrossRef]

49. Leminen, S.; Westerlund, M.; Nyström, A.-G.; Kortelainen, M. The Effect of Network Structure on Radical Innovation in Living Labs. *J. Bus. Ind. Mark.* **2016**, *31*, 743–757. [CrossRef]

50. Greve, K.; Leminen, S.; De Vita, R.; Westerlund, M. Unveiling the diversity of scholarly debate on living labs: A bibliometric approach. *Int. J. Innov. Manag.* **2020**. [CrossRef]