

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

*“No product is an island. A product is more than the product.
It is a cohesive, integrated set of experiences.”*

Donald Norman

Professor, consultant, and advocate for user-centred design

Within innovation research and, more specifically, living lab projects, a crucial component is to test an innovation in a real-life context with potential end users. Such a field test can validate assumptions by combining insights on behaviour and attitudes towards the innovation. This allows for iterative tailoring of the innovation to the needs and wants of the potential end users. Moreover, relevant insights can be gathered to stop or rescope the innovation project before big investments are made. Although studies indicate that testing innovations (or prototypes) in real-life contexts improves the innovation process, there is no specific framework on how to conduct a field test for an innovation. This is important because, in living lab field tests, users are actively involved in co-creating the solutions, which impacts the operational side of setting up living lab projects. Therefore, within this article, we propose a framework for field testing based on the degree to which it reflects reality and the stage within the living lab process. We distinguish four types of field tests: concept, mock-up, pilot, and go2market field test. Based on this framework, we propose some practical guidelines for setting up living lab field tests.

Introduction

Since the beginning of the digital revolution and the shift towards user involvement (Ortt & van der Duin, 2008; Rothwell, 1992), the usefulness and usability of digital systems became the object of study. In the 1950s, for example, Dreyfuss (2003) highlighted the importance of “designing for people” and emphasized the importance of creating good experiences for the end user. While the focus was on user experience, the evaluations of those experiences happened in a controlled lab (Benedek & Miner, 2002). Nowadays there is an increased tendency to extend the research process beyond the limitations of the lab towards the highly dynamic environment known as “real life”. If products are only tested in a lab setting, they often fail once introduced into the users’ natural environment. The main reason is that people are known to tailor their behaviour to the setting they are in: for example, users may exhibit different behaviour with similar technology in their home or the office (Intille et al., 2003). Additionally, there is a gap between what people say and what

they actually would do (Sanders & Stappers, 2012). Furthermore, users need to have passed the “honeymoon” period (i.e., the amount of time a user needs to get to know and form an attitude towards a new technology) before they can evaluate the technology (Spohrer & Freund, 2012). In other words, studying user interactions “in situ” over a longer period of time is indispensable. The living lab community has been aware of this from day one and recommends setting up a living lab to research the appropriation of technology in the user’s daily life (Dell’Era & Landoni, 2014). By setting up a real-life intervention (i.e., a field test) and by using multi-method approach, the likelihood of generating actionable user contributions for the innovation under development increases (Georges et al., 2016). The difference between living labs and regular social studies is the participatory aspect, where co-creation is more important than merely observing users interacting with technology. As such, a field test in a living lab, compared to a traditional field test, goes beyond gathering user feedback; it encourages users to propose improvements for the technology being tested (Spohrer & Freund, 2012).

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

However, the living labs literature is surprisingly silent in terms of the set-up of real-life experiments or testing. Living labs yield the greatest value when moving from concept to prototype in a living lab (Schuurman et al., 2016). Therefore, some living lab researchers and practitioners recommended defining hypotheses that can be tested throughout the entire living lab process in a real-life setting (Rits et al., 2015). These hypotheses can then contribute to the selection of research methods such as observation, experimentation, contextual interviews, etc. (Schuurman et al., 2018). But the principal challenge remains unanswered: how can these more “traditional” research methods be applied in real-life contexts and capture its dynamics? It is, for example, hard to define key settings in which tasks will be performed but also to collect qualitative data in the field (Brewster & Tucker, 2016; Coorevits & Jacobs, 2017). Thus, some academics have studied how different elements of context influence the user experience (Jumisko-Pyykkö & Vainio, 2012). Moreover, it has been demonstrated that evaluation methods such as the “think aloud” protocol need to be adapted and new methods that suit the challenges to evaluate technology in the field should be developed (Fields et al., 2007). Living lab researchers mention field tests as an approach to discover and understand how technology is appropriated in a real-life setting (Ballon et al., 2005; Følstad, 2008; Kjeldskov & Skov, 2014; Veckman et al., 2013). Although living lab researchers refer to real-life experimentation and testing as one of the key elements in living labs, Habibipour and co-authors (2018) did not find a common definition and therefore distilled theirs from *Merriam-Webster Dictionary*, which says that the aim of conducting a field test is “to test (a procedure, a product, etc.) in actual situations reflecting intended use”.

Most living lab researchers set up a field test towards the end of the innovation process, because it is at this point in time that the technology is mature enough to let users interact with it while taking into consideration the dynamic nature of context in which it all happens. However, Lew and colleagues (2011) argue that this should not strictly be necessary and there are possible variations in terms of the “realism” of the setting. Additionally, some studies also recommend simulations of the technology (e.g., a “Wizard of Oz” approach) or the context (e.g., a lab that looks like a living room) if the technology is not yet mature enough to make field tests possible (Coorevits & Jacobs, 2017), but they did not identify a common approach towards testing.

There is a need in the living lab community to reduce the complexity of their operations and have a more har-

monious and standardized approach (Leminen & Westerland, 2017; Mulder et al., 2008). Therefore, in this article, we seek to overcome some of the challenges related to real-life experiments and construct a framework that will encourage standardized field tests. Our approach is to use case studies to categorize field tests based on the stage of the innovation process and degree of contextual realism. The resulting framework is intended to help the living lab community maximize value from living lab processes. Accordingly, we also offer some practical guidelines for innovation practitioners.

Field Testing within Living Labs

A living lab employs a multi-method approach, engages users, enables participation from multiple stakeholders, and operates in a real-life setting so that the different parties involved can co-create a solution (Robles et al., 2015). A study from Schuurman, De Marez, and Ballon (2016) showed that a living lab yields maximum value when evolving from concept to prototype, but if some methodological elements are missing, user contributions will be limited. This is often the case for the real-life technology intervention. The authors assign this to the lack of maturity of the innovation, making it difficult to make the evaluation realistic. Living lab researchers often only implement field tests towards the end of the development process, because they assume the complex interactions between the system, user, and environment can only be observed when the innovation has reached a certain level of maturity. The real-life aspect means the product and setting are often designed to be as close to actual usage as possible. It is very common for researchers to let users operate the technology freely and evaluate the usage via objective and subjective measurements. They do this because it enables triangulation and because real-life experience lowers the barrier for user contribution (Schuurman et al., 2016). But, when taking this into consideration towards the end of the innovation process, the need for scope change can be detected too late, leading to high development costs. Although the uncontrollable dynamics and interactions between user and system create complexity in a living lab, they also steer learning and the further development of the innovation (Leminen & Westerland, 2017). As a solution, researchers and practitioners tried to deal with the challenge of studying complex contextual requirements in the different stages of a living lab project (Coorevits & Jacobs, 2017). Attempts were made to replicate the “wild” or real-life aspect during field tests in the early phases of the Living Lab project (Mulder & Stappers, 2009). This was done by either simulating the environment in which the interaction takes place (e.g.,

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

creating a usability lab that looks like a living room) or the technology itself (e.g., a “Wizard of Oz” approach or experience-prototyping techniques) (Dell’Era & Landoni, 2014; Mulder & Stappers, 2009; Sein et al., 2011; Stewart & Williams, 2005). Replication or simulations of “real life” and “technology” in tests are accepted in the living labs literature as long as the researcher remains aware of their constraints (Coorevits, Schuurman et al., 2016). This leads to a wide array of approaches and methods being used to test innovation in the field, while the living lab community is longing for more standardization (Leminen & Westerlund, 2017). Therefore, this article will try to bring structure to the way a living lab field test can be set up.

Based on previous studies on field tests and the importance of real-life testing in early stages of the innovation process (Georges et al., 2016; Habibipour et al., 2018), we created the following definition for field tests in living labs:

“A field test is a *user study* in which the interactions of *test users* with an *innovation* in the *context of use* are tested and *evaluated*.”

Following this line of reasoning, field tests can differ in terms of the stage in the living lab process they take place in and in the degree of realism. In the following sections, we discuss both of these aspects.

Stages in the Living Lab Process

The exploration phase

New product development (NPD) starts with a problem–solution fit stage, whereas, in a living lab, this first phase is called the “exploration phase” (Figure 1). The focus is on moving from the idea towards a concept of the solution. This requires studying the “current state” of users, identifying the problem, and trying to match a new solution to the problem while taking into account the specific contexts in which these problems occur

(York & Danes, 2006). The need–solution pairing happens by iteratively reformulating problems to discover need–solution pairs. This is done by testing a point in the solution landscape (per cycle) against a point in the need landscape for viability. The trial-and-error cycle continues until an acceptable need–solution pairing is found or created (von Hippel & von Krogh, 2013). This means that the innovation, with each step of the need solution pairing, will reach a higher level of maturity. Within the exploration phase, the maturity of the technology will be rather low, mostly including basic components of the solution. To test the problem–solution fit, we can use similar technologies (i.e., a proxy technology assessment) to learn how they currently solve their problems, which needs or problems are unresolved, and which (partial) solutions work. Although in the strict sense of the definition, these type of interventions are not with the innovation at hand, we still perceive them as a field test.

The experimentation phase

The second stage within an innovation development process can be labelled as “experimentation” where we move from concept to prototype. In general, a prototype can be perceived as something being built to represent a product or experience before the actual artefact is completed (Sanders & Stappers, 2012). Prototypes of ICT products can have many variations, from paper prototypes, which are sketched representations of the graphical user interface, to functional prototypes that can be used on a device or features under development being mimicked (i.e., using a “Wizard of Oz” approach) allowing real-world tests (Coenen & Robijt, 2017). The form is influenced by the learning objectives with regards to the possible “future state”. Hence, their main goal is to facilitate hypothesis testing. In this stage, users are confronted for the first time with the solution, so user research mainly studies how users react to and interact with the new solution. In summary, the experimentation stage puts the designed solution to the test, as much as possible in a real-life context, and it

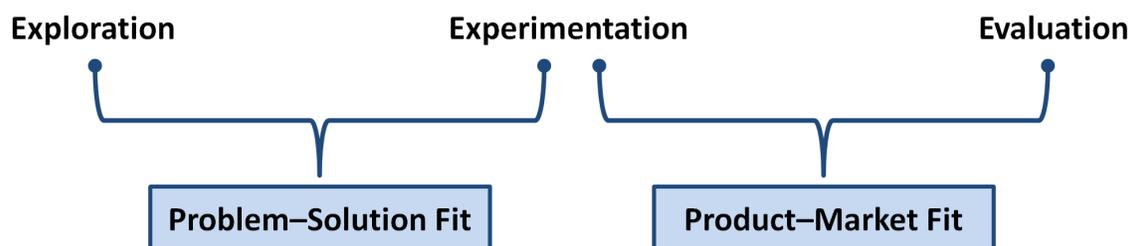


Figure 1. Overview of the NPD process and its three corresponding stages in living labs

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

allows a decision to be made on whether to head back to the exploration stage to iterate the solution or whether to proceed to the evaluation stage.

The evaluation phase

The third and final stage consists of evaluating the innovation in terms of market fit. Within this phase, the innovations have a rather high level of maturity. The focus is on how to enter the market, including determining which users will adopt first, how to communicate with them, and which features should be launched to maximize uptake and continued use. York and Danes (2006) refer to this as “customer validation”, which means the identification of a scalable and repeatable sales model, where the goal is to establish product–market fit and find a viable business model. A key question at this stage is: what advantages is the innovation able to deliver? This facilitates the determination of pricing levels, given that the impact of the solution can be quantified. This stage can also consist of the post-launch activities, where actual adoption and usage of the innovation is monitored in order to re-design or add new functionalities according to the needs of existing or new market groups.

Schuurman, Ballon, and De Marez (2016) showed that it is more challenging to organize a field test in the early stages of the NPD process. Extra effort and expertise are required to make the test possible. The framework in this article will help researchers and practitioners to gain more expertise on how to organize a field test in each phase of the process.

Degree of Realism

The second parameter that will determine the type of field test that can be set up in a living lab project is related to context. For some innovations, a particular use context will be simulated to test the innovation. The most important thing is to determine the degree of realism (i.e., how close the test is to the actual use and context) required for an evaluation to be meaningful and which aspects of use are important enough to preserve in the evaluation setup (Coorevits & Jacobs, 2017). For example, the physical location cannot be similar to the one in which the final product will be used, the test users are not representative to real users, the tasks will not be the same, the motivations and other concurrent activities of participants are different in the test situation compared to real-life, etc. Kjeldskov and Skov (2014) as well as Korn and Bodker (2012) called for greater awareness of the trade-offs you make when simulating a context. They state that, the better the under-

standing of the context in which an activity takes place, the better the evaluation of a system. Coorevits and Jacobs (2017) provided a framework to understand context in living labs. The framework goes beyond the traditional understanding of a real-life setting (the physical environment) and highlights the importance of social, task, time, and other elements that can influence the interaction with a system. If one or some of these elements are not realistic in a living lab field test, they might also influence the outcome of the study. Unrealistic content, for example, can feel artificial to the user and can lead to atypical behaviour because they perceive the system itself as unrealistic. They might start to explore the boundaries of the system out of curiosity. If users are asked, as part of a usability test, to perform a series of tasks that are not relevant to them, this might create boredom or displeasure, which might be wrongly seen as an outcome of the study instead of the treatment, and as such it compromises the external validity of the usability test.

There are five components of context that can influence the interaction with a system:

1. *Temporal context*: the interaction of the user with the system in relation to time (Tamminen et al., 2004). Time can be simulated by giving users dedicated moments in time where they have to perform actions, by establishing the duration in which the field test takes place, etc.
2. *Physical context*: the apparent features of a situation or physically sensed circumstances in which the user/system interaction takes place (Dourish, 2004). A physical context can be simulated by making a lab look like a living room, for example, or by limiting the physical context to a certain area of the real physical context.
3. *Technical/information context*: the relationship to other services and systems that are relevant to users' systems. It also refers to the interoperability, informational artefacts, and access between devices, services, platforms, etc. Simulations can happen by mimicking the autonomy of a system or features but also the aesthetics and content available in the system.
4. *Social context*: the other people present, their characteristics, and roles but also the interpersonal interactions and culture surrounding the user systems interactions. When simulating the social context, for example, social interactions can be reduced by test-

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

ing with the user alone, or users can be asked to test with a friend, family member, or colleague.

5. *Task context*: all the tasks surrounding the user's interaction with the system. Simulation of the task context means, for example, that the user is asked to perform certain tasks during the field test (Bailey & Konstan, 2006).

Although simulation of these five contextual elements and the decision to simulate particular elements while not controlling others will vary depending on the living lab requirements and as such require a custom approach, there is still a common trend. If the maturity of an innovation is high, fewer simulations will be required.

Methodology

Based on the above elements, we composed a high-level framework composed of four quadrants along two axes: degree of realism (high vs. low) and phase in the living lab project (early vs. late). This leads to four "archetypes" of living lab field tests: low realism and early phase, high realism and early phase, low realism and late stage, and high realism and late stage. In order to validate and fine-tune the framework, we performed a qualitative multiple

illustrative case study. Yin (2009) defines the case study research method as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used". The goal was to determine whether these four archetypical field tests could be found in living labs practice and to better understand their potential differences and value. We used action research to analyze the cases, which is particularly relevant when producing guidelines for best practices (Sein et al., 2011). We composed a sample of 17 field tests out of more than 100 living lab innovation projects from imec.livinglabs (see also Schuurman et al., 2016 and the imec.livinglabs website: imec-int.com/en/livinglabs). Out of these cases, the author team selected four field tests that best matched the four archetypes.

Results and Discussion

In this section, we identify the four types of field tests that resulted from our coding – concept, mock-up, pilot, and go2market (Figure 2) – and describe them with illustrative case studies. We then elaborate on the operationalization of these four types of field tests.

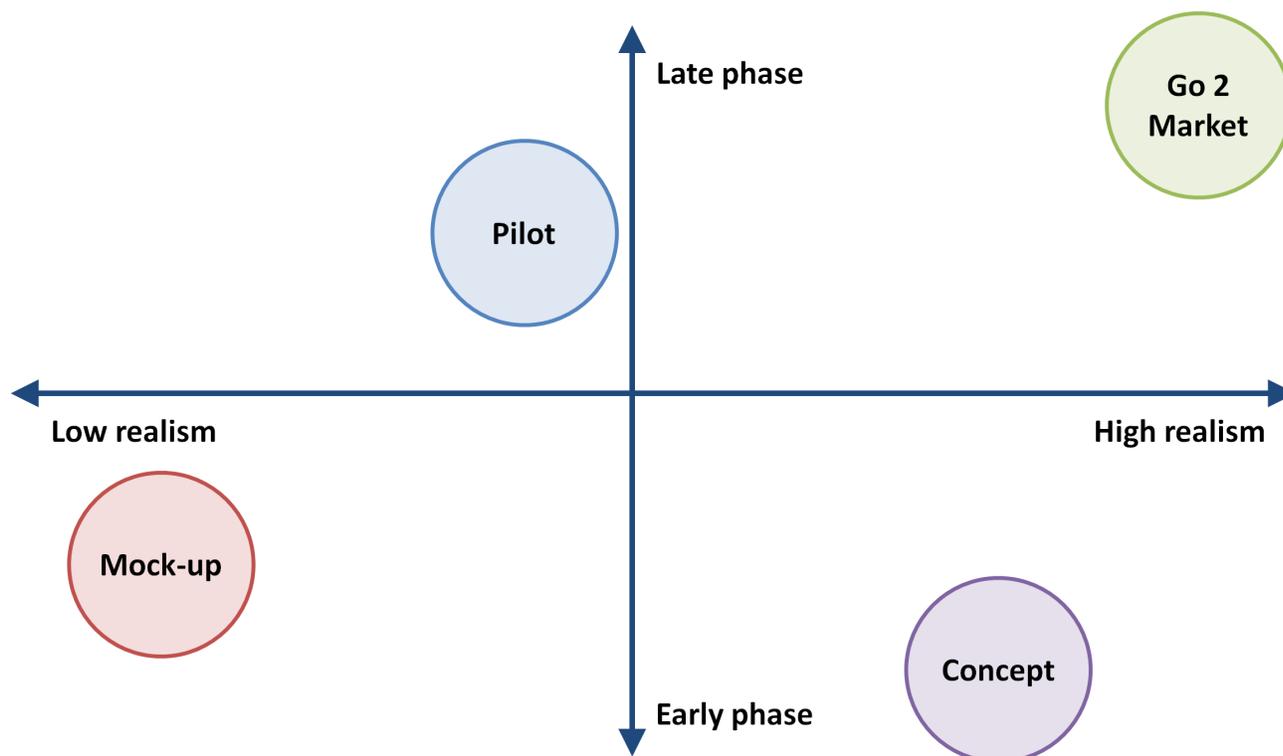


Figure 2. The four types of field tests in living labs, characterized by their phase and degree of realism

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

Concept field test

Concept tests are, in the strict sense of the field test definition above, not a field test because the intervention happens with existing technologies and not with the innovation itself, but we include them in the model because they share other elements of the definition. Concept tests will help identify the user's problem in the early stages of new product development. By focusing on a preliminary idea and applying lightweight technological interventions that attempt to investigate current practices and experiences, the output of this test will inform the development of the value proposition the innovation should focus on. It is a good way to gather feedback before wireframes or prototypes are developed. The intention of a concept test is to evolve from idea to mock-up. They are mostly done with 5–8 people (per persona) in a real environment. An example of a concept test is the proxy technology assessment (Bleumers et al., 2010; Brown et al., 2011). A proxy technology assessment lets future users experience one or more related technologies (i.e., hardware or software) that already exist today. Crucial is that these technologies share as many characteristics as possible with the technology under development. These types of technologies are described as proxy technologies. Both the way in which the proxy technology is appropriated and the users' experience-based reflections on these technologies can be used to inform and inspire the development of new technologies in an early stage. Smoke testing will help to quantitatively validate and measure the needs, value promise, and initial interest in a product (Gothelf, 2013). The goal is to justify building the product. A smoke test is typically a one-page website describing the product or service before it is actually available. The potential customer or user is at that point in time not aware that it does not yet exist but must give some form of payment to access the product or service. Ideally, smoke tests happen in an "A/B format" that compares two or three different value promises and the potential uptake with enough users to statistically validate the results (e.g., $n=30$ per format).

Within our sample of field tests from imec.livinglabs, we selected NowYu. This was a project to identify how users can gain greater control over their data on social media. The project examined how and what people are willing to share as well as the value they expected in return. A proxy technology assessment was set up where we asked several users to test different data-sharing platforms. The platforms were selected in a way that we could test user preference for different potential re-

wards or values, data sharing and control mechanisms, etc. The users were given assignments, but they were free to choose whether they wanted to perform the action on the platform and when they wanted to perform it. They received screenshots of the platform on which they could write feedback related to their experiences, reasons for taking or not taking actions, etc. In other words, the degree of realism was rather high. This allowed us to create clickable mock-ups and interesting navigation flows and make decisions in relevant features to accomplish a problem–solution fit.

Mock-up field test

Mock-up tests can help to gather information about the nature of the interaction and test it before the functional model is built. Additionally, they can investigate aspects of the product form such as visual affordances. These tests are especially relevant if they happen before the actual development takes place as they can guide the development in the right direction. The IEEE's report "Why Software Fails" points out that an estimated 50% of rework time could have been avoided had testing been done in the early design stages (Charette, 2005). Mock-up tests are mostly done with 5–8 people and focus on testing the intended interactions in a semi-real environment. Two examples of mock-up tests are "Wizard of Oz" and augmented reality (AR) simulations. The Wizard of Oz is a technique that enables the evaluation of an unimplemented technology by using a human to simulate the response of a system. The AR simulation can create a mock object that simulates the behaviour of complex, real objects. This is useful when it is impractical or impossible to incorporate the object in a real test. For example, when the test requires structural changes to infrastructure in a city, which is impossible.

As an illustrative case study, we chose GARbage. This was a project in which we simulated a screen on a Big Belly (this is a type of smart garbage bin) via AR. The goal was to identify how smart garbage bins could be made more interactive. The simulated screen allowed citizens to report litter or call the emergency numbers. During the field test, we simulated the technology in AR because it was difficult to make structural changes to the environment, and tasks were simulated by asking the users to walk through a given scenario while imagining them really happening because the likelihood of occurrence is rare. In other words, time, task, technical, and social context were simulated. The physical location of a city context remained natural. The test allowed us to identify a non-fit between problem and solution, as well as suggestions from participants on how to rescope.

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

Pilot field test

Pilot tests should provide insights into anything that might be missing in the innovation, so this can be adjusted before the complete roll-out to a larger group of test users. Pilot testing focuses on testing the entire system with a subset of users in real-life conditions and can be perceived as the dry-run test of the innovation. This should improve the likelihood of an optimized user experience. As the goal at this point is to quantitatively gain insights, involving 20–30 people will be required to statistically infer conclusions. One example of a pilot test is setting up test marketing. Test marketing is a method wherein the product is launched in a selected (geographical) area that is representative of the final market to check the viability of the product and the demand among the selected group of people. Test marketing is relevant when you decided to go to the market but, of course, the test can alter the plans by giving a no-go. In other words, this test allows testing in a real (sub) context with the minimal viable product.

iCinema was a project in which we wanted to create an application that allowed interactions via a second screen (smartphone) in a movie theatre to increase audience engagement. Because of the potential contextual barriers, we invited several users to come and see a movie. Most contextual elements had a high realism such as the people they came with, but the test was not completely natural. For example, the time and information context (i.e., the movie being played) was simulated because of the “test setup”. During the test, the movie theatre screen invited them to interact via their smartphones, while we measured not only the number of people actually interacting, but also their experiences. The outcome indicated that second-screen interaction is acceptable, but only before and after the movie, so innovations should only focus on those time periods. The outcome allowed us to make some minor tweaks and launch the application during Ghent’s film festival.

Go2market field test

Go2market field tests are mostly used to validate the innovation concept when the maturity is at a higher level. The research questions are related to the product–market fit, focusing on the willingness-to-pay, retention, growth, and how to put the innovation in the market. Often, these tests will have an A/B testing scenario to estimate, for example, how new features are adopted by users and whether or not they increase retention. Go2market field tests are characterized by a high level of maturity resulting in the fact that the test can have a high degree of realism. As the goal is to make predic-

tions for the entire population, samples start at a minimum of 50 users, while experts claim that a higher rate of sampling is often even better.

SPOTT was a project in which an application was tested that allowed users to buy products being shown on television while they are watching their favourite television show. Given that the users could test the application at home during the course of a month and no instructions were given, the context was completely natural. This also implied that the content of certain television shows was made interactive, so anyone downloading the app and watching these programs could participate. The test was intended to validate learnings from previous steps and provide insights into the willingness to pay per adoption profile. The most important outcome of this test was answers to questions about how to accomplish growth and retention.

Guidelines for Operationalizing the Different Types of Field Tests

The four types of tests indicate some differences in setup. The early stages of the living lab process deal with innovations that have a low maturity. Also, the degree of realism will be simulated to a greater extent. In the early stages, the focus is on validating assumptions about customer needs, on identifying target segments for a new product or idea, and on gathering insights to define an innovation with a competitive value promise.

Early-stage field tests share the following characteristics, which take the form of practical guidelines for setting up living lab field tests:

1. *Small-scale and closed:* When setting up a field test in the early stages, a smaller number of test users is needed. First of all, the input you will receive from a larger number of users is limited. Second, as most living lab researchers and practitioners operate with a tight budget, it is better to spread that budget over different steps of your iteration process. When selecting this small number of users, it is important to focus on specific user profiles or personas (Coorevits et al., 2016) to join your test. This will allow you to identify the most promising target groups, their needs, and how the innovation should be formed to reach maximum potential.
2. *Higher degree of guidance:* Because the maturity of the innovation is low and there still are many uncertainties, you will have to select the most critical assumptions or uncertainties to test in this stage of the

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

process. It is about diving deeper into the habits of users while putting them in context. As a researcher, you will often spend time preparing, for example, a storyboard representing the situation and taking the user through the journey by asking the user to perform certain tasks. This means that users will be given more specific guidelines on how to test the innovation, to gain answers on your specific questions (e.g., Is the use flow correct and does it make sense? Is the design understandable?). This also means that, as a researcher, you will have to be aware of not biasing the outcome because the test will be more intrusive for the user.

3. *Qualitative*: During these early stages, we often try to answer questions that are related to the “why” and “how”, so we can find a better problem–solution fit. For example, what problems are users currently facing and how are they trying to solve them? Therefore, profound qualitative research methodologies are more appropriate. The more the innovation takes form, the higher the level of maturity and the higher the degree of realism that can be accomplished in the field tests. In this phase, research steps are focusing more on validating the assumptions and creating minor tweaks so uptake of the innovation can reach its maximum potential.

Field tests during the later stages will show the following characteristics:

1. *Large-scale and open*: As the main focus will be to validate the value promise on a larger scale, these types of field tests will include a larger group of test users in which the field test also has a more open character and everyone who qualifies can participate in the test. You will often choose a specific group of users or all users as they use the product over time. You can gain insights into bugs, issues they face while using them, or needs for further improvement. This larger group of test users is needed to get a statistical validation of the proposed innovation, potential future roadmap based on adoption potential per target group, and to operationalize the willingness to pay (De Marez & Verleye, 2004).
2. *Limited to no guidance*: As the research questions are mainly related to finding a product–market fit, the test subjects should be asked to act freely to avoid “surprises” during market launch. The main focus is to make sure your product can stand the highly dynamic contextual requirements that can function as

a driver or barrier to interactions and, therefore, the test should be as natural as possible, meaning limited involvement of the researchers and limited-to-no guidelines should be given to the users in how to test. This also implies the test is less intrusive for the user.

3. *Quantitative*: As the focus is on validation and larger user groups are involved, the methods used will be more quantitative in nature. Questions about “what” and “how many” will be answered during these field tests. Log data from the system and measurements (in the form of a survey) will take place at several time intervals or when certain events take place to learn about how users behave, their attitudes towards the technology, and their wishes about how to improve the technology towards an optimized product.

Conclusion

Within this article, we proposed a framework for field testing based on two axes, the phase in the living lab process and the degree of realism. Based on these two axes, and by means of four illustrative case studies, we identified four types of field tests: concept, mock-up, pilot, and the go2market. The goal of this framework is to guide practitioners to set-up field tests at every stage in the living lab process. At this moment, we see that field tests are mostly used to evaluate innovations, however, we believe that conducting field tests in an earlier phase of the innovation process can help fit the solution better to the problem.

Although increasing realism is important, not all modifications can justify the needed time and resources. Therefore, we recommend using the framework of Coorevits and Jacobs (2017) to become aware of all contextual elements that might potentially influence the interaction and make trade-offs accordingly. This will allow the researcher to become more aware of bias in their study and reduce the impact on outcomes. Additionally, it will allow field tests to be set up in the early stages of development in the living lab, because it enables decisions about what to simulate, while remaining aware of the influence that non-finished or semi-real elements can have on the outcome of a test. The earlier in the development stage, the more trade-offs will have to be made, but it will allow the researcher to take into consideration the appropriation of technology sooner, and it will ultimately reduce the likelihood of product failure.

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

Even though this framework can guide practitioners in setting up field tests, we are aware that other factors can influence the set-up of the field test, such as the duration of the test. This is something that needs careful consideration. It depends on the complexity of the product, but it should last until the user feels confident that they know how to use the product. Also, other elements such as learning of completely new behaviours, the impact of the innovation on the daily life, the social character of the innovation, the installation or use of specific hardware, etc. can influence that setup, and therefore further research is needed to enrich the framework.

There is also the substantial challenge of measuring the behaviour of people in a context when testing innovations. Therefore, new methods and tools such as experience sampling and wearables can contribute to study the behaviour of test users. More research is needed to determine which methods could be used best in each type of field test.

Acknowledgements

This article was developed from work presented at the ISPIM Innovation Conference in Stockholm, Sweden, June 17–20, 2018. ISPIM (ispim-innovation.com) – the International Society for Professional Innovation Management – is a network of researchers, industrialists, consultants, and public bodies who share an interest in innovation management.

About the Authors

Lynn Coorevits is an Innovation Manager at imec.livinglabs, where she is responsible for tools and methodologies that improve testing in living labs. She is also affiliated to the imec-mict-ugent department Ghent University in Belgium as a Senior User Researcher. She holds master degrees in Psychology and Marketing Analysis from Ghent University and has over 10 years of experience in innovation research and consultancy.

Annabel Georges is a User Specialist at imec.livinglabs. Within this “living-lab-as-a-service” she conducts user research to structurally support innovation development for SMEs. Next to this, she is also involved in several smart city projects. Her main research topics are drop-out and user engagement within field tests and working on the processes of current living lab practices. Annabel holds a Master in Communication Sciences degree with a specialization in New Media and Society at Ghent University.

Dimitri Schuurman is the Team Lead of the Business Model and User Research Team at imec.livinglabs. He holds a PhD and a Master’s degree in Communication Sciences from Ghent University in Belgium. Together with his imec colleagues, Dimitri developed a specific living lab offering targeted at entrepreneurs in which he has managed over 100 innovation projects. He is also active in the International Society for Professional Innovation Management (ISPIM) and in the European Network of Living Labs (ENoLL) as a living labs specialist. His main interests and research topics are situated in the domains of open innovation, user innovation, and innovation management.

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

References

- Bailey, B. P., & Konstan, J. A. 2006. On the Need for Attention-Aware Systems: Measuring Effects of Interruption on Task Performance, Error Rate, and Affective State. *Computers in Human Behavior*, 22(4): 685–708.
<https://doi.org/10.1016/j.chb.2005.12.009>
- Ballon, P., Pierson, J., & Delaere, S. 2005. *Test and Experimentation Platforms for Broadband Innovation: Examining European Practice*. Paper presented at the 16th International Telecommunications Society Europe Conference, September 4, 2005, Porto, Portugal.
<https://doi.org/10.2139/ssrn.1331557>
- Benedek, J., & Miner, T. 2002. *Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting*. Paper presented at the Usability Professionals Association 2002 Conference, July 8–12, 2002, Orlando, FL.
<https://doi.org/10.1080/02772248.2012.710412>
- Bleumers, L., Naessens, K., & Jacobs, A. 2010. How to Approach a Many Splendored Thing: Proxy Technology Assessment as a Methodological Praxis to Study Virtual Experience. *Journal of Virtual Worlds Research*, 3(1): 3–24.
- Brewster, M., & Tucker, J. M. 2016. Understanding Bystander Behavior: The Influence of and Interaction Between Bystander Characteristics and Situational Factors. *Victims and Offenders*, 11(3): 455–481.
<https://doi.org/10.1080/15564886.2015.1009593>
- Brown, B., Reeves, S., & Sherwood, S. 2011. Into the Wild: Challenges and Opportunities for Field Trial Methods. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, May 7–12, 2011, Vancouver, BC.
<https://doi.org/10.1145/1978942.1979185>
- Charette, R. N. 2005. Why Software Fails. *IEEE Spectrum*, 42(9): 42–49.
<https://doi.org/10.1109/MSPEC.2005.1502528>
- Coenen, T., & Robijt, S. 2017. Heading for a FALL: A Framework for Agile Living Lab Projects. *Technology Innovation Management Review*, 7(1): 37–43.
<http://doi.org/10.22215/timreview/1048>
- Coorevits, L., & Jacobs, A. 2017. Taking Real-Life Seriously: An Approach to Decomposing Context Beyond “Environment” in Living Labs. *Technology Innovation Management Review*, 7(1): 26–36.
<http://doi.org/10.22215/timreview/1047>
- Coorevits, L., Schuurman, D., Oelbrandt, K., & Logghe, S. 2016. Bringing Personas to Life: User Experience Design through Interactive Coupled Open Innovation. *Persona Studies*, 2(1): 97–114.
<https://doi.org/10.21153/ps2016vol2no1art534>
- De Marez, L., & Verleye, G. 2004. Innovation Diffusion: The Need for More Accurate Consumer Insight. Illustration of the PSAP Scale as a Segmentation Instrument. *Journal of Targeting, Measurement and Analysis for Marketing*, 13(1): 32–49.
<https://doi.org/10.1057/palgrave.jt.5740130>
- Dell’Era, C., & Landoni, P. 2014. Living Lab: A Methodology between User-Centred Design and Participatory Design. *Creativity and Innovation Management*, 23(2): 137–154.
<https://doi.org/10.1111/caim.12061>
- Dourish, P. 2004. What We Talk About When We Talk About Context. *Personal and Ubiquitous Computing*, 8(1): 19–30.
<https://doi.org/10.1007/s00779-003-0253-8>
- Fields, B., Amaldi, P., Wong, W., & Gill, S. 2007. In Use, In Situ: Extending Field Research Methods. *International Journal of Human-Computer Interaction*, 22(1–2): 1–6.
<https://doi.org/10.1080/10447310709336952>
- Følstad, A. F. 2008. Living Labs for Innovation and Development of Information and Communication Technology: A Literature Review. *Electronic Journal for Virtual Organizations and Networks*, 10: 99–131.
- Georges, A., Schuurman, D., & Vervoort, K. 2016. Factors Affecting the Attrition of Test Users During Living Lab Field Trials. *Technology Innovation Management Review*, 6(1): 35–44.
<http://doi.org/10.22215/timreview/959>
- Gothelf, J. 2013. *Lean UX: Applying Principles to Improve User Experience*. Sebastopol, CA: O’Reilly Media.
- Habibipour, A., Georges, A., Ståhlbröst, A., Schuurman, D., & Bergvall-Kåreborn, B. 2018. A Taxonomy of Factors Influencing Drop-Out Behaviour in Living Lab Field Tests. *Technology Innovation Management Review*, 8(5): 5–22.
<https://doi.org/10.22215/timreview/1155>
- Intille, S. S., Tapia, E. M., Rondoni, J., Beaudin, J., Kukla, C., Agarwal, S., Bao, L., & Larson, K. 2003. Tools for Studying Behavior and Technology in Natural Settings. In A. K. Dey, A. Schmidt, A., & J. F. McCarthy (Eds.), *UbiComp 2003: Ubiquitous Computing Lecture Notes in Computer Science*, 2864: 157–174.
https://doi.org/10.1007/978-3-540-39653-6_13
- Jumisko-Pyykkö, S., & Vainio, T. 2012. Framing the Context of Use for Mobile HCI. *International Journal of Mobile Human Computer Interaction*, 2(4): 1–28.
<https://doi.org/10.4018/jmhci.2010100101>
- Kjeldskov, J., & Skov, M. B. 2014. Was It Worth the Hassle? Ten Years of Mobile HCI Research Discussions on Lab and Field Evaluations. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*: 43–52, September 23–26, 2014, Toronto, ON.
<https://doi.org/10.1145/2628363.2628398>
- Korn, M., & Bodker, S. 2012. Looking Ahead: How Field Trials Can Work in Iterative and Exploratory Design of Ubicomp Systems. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*: 21–30, September 5–8, 2012, Pittsburgh, PA.
<https://doi.org/10.1145/2370216.2370221>
- Leminen, S., & Westerlund, M. 2017. Categorization of Innovation Tools in Living Labs Categorization of Innovation Tools in Living Labs. *Technology Innovation Management Review*, 7(1): 15–25.
<https://doi.org/10.22215/timreview/1046>
- Lew, L., Nguyen, T., Messing, S., & Westwood, S. 2011. Of Course I Wouldn’t Do That in Real Life: Advancing the Arguments for Increasing Realism in HCI Experiments. In *CHI ’11 Extended Abstracts on Human Factors in Computing Systems*: 419–428, May 7–12, 2011, Vancouver, BC.
<https://doi.org/10.1145/1979742.1979621>
- Mulder, I., & Stappers, P. J. 2009. Co-Creating in Practice: Results and Challenges. In *Proceedings of the 2009 IEEE International Technology Management Conference (ITMC)*: 1–8, June 22–24, 2009, Leiden, Netherlands.
<https://doi.org/10.1109/ITMC.2009.7461369>

A Framework for Field Testing in Living Lab Innovation Projects

Lynn Coorevits, Annabel Georges, and Dimitri Schuurman

- Mulder, I., Velthausz, D., & Kriens, M. 2008. The Living Labs Harmonization Cube: Communicating Living Lab's Essentials. *The Electronic Journal for Virtual Organization & Networks*, 10(November): 1–14.
- Ortt, J. R., & van der Duin, P. A. 2008. The Evolution of Innovation Management Towards Contextual Innovation. *European Journal of Innovation Management*, 11(4): 522–538. <https://doi.org/10.1108/14601060810911147>
- Rits, O., Schuurman, D., & Ballon, P. 2015. Exploring the Benefits of Integrating Business Model Research within Living Lab Projects. *Technology Innovation Management Review*, 5(12): 19–27. <http://doi.org/10.22215/timreview/949>
- Robles, A. G., Hirvikoski, T., Schuurman, D., & Stokes, L. 2015. *Introducing ENoLL and Its Living Lab Community*. Brussels: European Commission.
- Rothwell, R. 1992. Successful Industrial Innovation: Critical Factors for the 1990s. *R&D Management*, 22(3): 221–240. <https://doi.org/10.1111/j.1467-9310.1992.tb00812.x>
- Sanders, E. B.-N., & Stappers, P. J. 2012. *Convivial Toolbox: Generative Research for the Front End of Design*. Amsterdam: BIS.
- 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): 7–16. <http://doi.org/10.22215/timreview/956>
- Schuurman, D., Herregodts, A. L., Georges, A., & Rits, O. 2018. *Innovation Management in Living Lab Projects*. Paper presented at Open Living Lab Days, August 22–24, 2018, Geneva, Switzerland.
- Sein, M. K., Henfridsson, O., Rossi, M., & Lindgren, R. 2011. Action Design Research. *MIS Quarterly*, 35(1): 37–56.
- Spohrer, J. C., & Freund, L. E. (Eds.) 2012. *Advances in the Human Side of Service Engineering*. Boca Raton, FL: CRC Press. <https://doi.org/10.1201/b12315>
- Stewart, J., & Williams, R. 2005. The Wrong Trousers? Beyond the Design Fallacy: Social Learning and the User. In H. Rohracher (Ed.), *User Involvement in Innovation Processes. Strategies and Limitations from a Socio-Technical Perspective*: 39–71. Munich: Profil Verlag.
- Tamminen, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A. 2004. Understanding Mobile Contexts. *Personal and Ubiquitous Computing*, 8(2): 135–143. <https://doi.org/10.1007/s00779-004-0263-1>
- Veekman, C., Schuurman, D., Leminen, S., Lievens, B., & Westerlund, M. 2013. Characteristics and Their Outcomes in Living Labs: A Flemish-Finnish Case Study. In *Proceedings of the XXIV ISPIM Conference – Innovating in Global Markets: Challenges for Sustainable Growth*, June 16–19, 2013, Helsinki, Finland. <http://doi.org/10.13140/2.1.3147.1047>
- von Hippel, E. A., & von Krogh, G. 2013. *Identifying Viable 'Need-Solution Pairs': Problem Solving Without Problem Formulation*. MIT Sloan School Working Paper 5071-13. Cambridge, MA: Massachusetts Institute of Technology (MIT). <https://doi.org/10.2139/ssrn.2355735>
- Yin, R. K. 2009. *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage.
- York, J. L., & Danes, J. E. 2006. Customer Development, Innovation, and Decision-Making Biases in the Lean Startup. *Journal of Small Business Strategy*, 17(2): 89–103.

Citation: Coorevits, L., Georges, A., & Schuurman, D. 2018. A Framework for Field Testing in Living Lab Innovation Projects. *Technology Innovation Management Review*, 8(12): 40–50. <http://doi.org/10.22215/timreview/1204>



Keywords: living labs, field test, user innovation, context research, testing

Academic Affiliations and Funding Acknowledgements



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

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



Carleton
UNIVERSITY



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

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