



Assessment of the advancement of market-upstream innovations and of the performance of research and innovation projects

Svetlana Klessova, Sebastian Engell, Catherine Thomas

► To cite this version:

Svetlana Klessova, Sebastian Engell, Catherine Thomas. Assessment of the advancement of market-upstream innovations and of the performance of research and innovation projects. *Technovation*, 2022, 116, pp.102495. 10.1016/j.technovation.2022.102495 . hal-03636260

HAL Id: hal-03636260

<https://hal.science/hal-03636260>

Submitted on 9 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Assessment of the advancement of market-upstream innovations and of the performance of research and innovation projects

Svetlana Klessova^{a,*}, Sebastian Engell^b, Catherine Thomas^a

^a Université Côte d'Azur, CNRS, GREDEG, 250 avenue Albert Einstein, 06560, Valbonne, Sophia Antipolis, France

^b Process Dynamics and Operations Group, BCI, TU Dortmund University, Emil-Figge Str, 70, 44221, Dortmund, Germany

ARTICLE INFO

Keywords:

Innovation
Innovative technologies
Market-upstream
Maturation
R&D
Technology readiness levels
Project
Innovation performance

ABSTRACT

The assessment of the advancement of technological innovations at their development stage is a difficult task, but important to judge on the performance of innovation projects. Assessments have so far been made by assessing technical characteristics, subjectively, or by counting patents. This paper proposes an approach to assess the advancement of market-upstream innovations *directly and objectively*, through the advancement of their technological maturity. On this basis, also the innovation performance of larger projects that were put in place to progress one or several innovations, can be assessed. The paper presents an exploratory qualitative multi-case study of 54 innovative technologies at different maturity levels, that were developed in 5 market-upstream large technological research and innovation projects with mostly engineering and IT dimensions, funded by the European Union's Research and Innovation Programmes under its sub-programme "Leadership in Enabling and Industrial Technologies". From extensive documentation and data from interviews, a refined technology readiness scale and a scoring method that reflects the increase in the required efforts to advance the maturity of the innovations is developed. The findings provide groundwork for future research on market-upstream innovation and how the innovation performance of projects can be measured at the early stages of the innovation process.

1. Introduction

Scholars, practitioners and policy makers repeatedly question how to measure the progress of technological innovations during their development and validation phase. However, the assessment of the progress at this phase is complex and challenging (Henttonen et al., 2016; Birchall et al., 2011), and there is not much literature on this topic (Dziallas and Blind, 2019). For simplicity, this paper uses the term 'market-upstream innovations' to describe innovative technologies, methods, processes, software, algorithms, products ... during their development and validation stage, on the way to implementation or productisation.

A related concern is that the performance of research, development and innovation (RDI) projects is also not well understood on the innovation side, while public agencies around the world invest significant amounts of taxpayers' money into such RDI projects, with the goal to increase the maturity of technological innovations on their way to implementation, but still market-upstream. For instance, the part of the

Horizon Europe Research and Innovation Framework Programme, launched in spring 2021, that supports industrial and closer-to-market research and development was allocated 56% of €95.5 billion, the entire budget of the Programme (Science / Business, 2021). It is obvious from the sheer size of the public programmes that there is a strong interest and an urgent need to systematically understand, assess and compare the advancement of the innovations that are targeted by the projects and their innovation performance, if possible directly and objectively already at the end of the projects. Such assessments are challenging (Kim et al., 2017; Aguiar and Gagnepain, 2017), and suitable indicators are not generally agreed upon (Schwartz et al., 2011; Enzing et al., 2015).

The most often proposed way to analyse the progress of innovations is by analysing the advancement in the technical characteristics of the innovations (Bican and Brem, 2020; Lazzarotti et al., 2011; Chiesa and Masella, 1996). However, this makes it specific to each innovation, and so it is difficult to derive general principles from this analysis. Also, the development efforts of innovations depend on contextual factors

* Corresponding author.

E-mail addresses: Svetlana.Klessova@univ-cotedazur.fr (S. Klessova), Sebastian.Engell@tu-dortmund.de (S. Engell), Catherine.Thomas@univ-cotedazur.fr (C. Thomas).

<https://doi.org/10.1016/j.technovation.2022.102495>

Received 7 August 2021; Received in revised form 14 February 2022; Accepted 25 February 2022

Available online 21 March 2022

0166-4972/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

(Klessova et al., 2020) and can differ largely (Héder, 2017). Other ways of assessment are by counting patents, although scholars agree that this lens does not provide the full picture (e.g. Janger et al., 2017; Hung, 2017; Messeni Petruzzelli et al., 2018; Martin, 2016a), or by looking at the satisfaction of the users' needs (Bendoly, 2014; Hagedoorn et al., 2000), or based on the opinions of the project actors (e.g. Arroyabe et al., 2015; Arranz et al., 2018), thus subjective.

The level of advancement of innovations is often measured using the Technology Readiness Level (TRL) scale (Mankins, 2009a), which helps to situate the progress of the maturity level of an innovation during the development process. Maturity means how technologically ready (mature) the innovation is on its way towards the final use. This scale is coarse, and beyond the defence and space sectors, from which it is originated, its application is debated (Héder, 2017; Olechowski et al., 2020). Webster and Gardner (2019, p. 1231) nonetheless state "given its use in a diverse range of systems within which technology is being developed, procured, and readied for market, it is surprising that TRLs have had relatively little consideration in the field of innovation studies". In this paper, we argue that the technical progress at the development and validation stage should be measured in terms of the advancement of its maturity between the end point that is reached within the project and the starting point (the baseline).

Scholars called for research on the assessment of innovations and R&D performance at their early phases (Dziallas and Blind, 2019) and for better understanding of the innovation performance of RDI projects (Pandza et al., 2011; Spanos and Vonortas, 2012), but not much progress has been made. Innovation studies have become largely econometric (Martin, 2016b), but to capture specific characteristics of the innovation process and its dynamics, a case study-based approach can provide additional valuable insights (Messeni Petruzzelli et al., 2018) and be beneficial for innovation research (Collinson and Liu, 2019).

This paper responds to these different calls, adapts a process-oriented view on innovation (Garud et al., 2017) and investigates how to assess the progress of market-upstream innovations through the lens of the progress of their technological maturity, and, by extension, how to assess the innovation performance of market-upstream RDI projects. The research design is a comparative exploratory multiple case study (Eisenhardt, 1989), with additional quantitative supporting evidence. Empirically, our research focuses on 54 innovations at their development and validation stage pursued by 5 large RDI projects with a focus on engineering and IT with 67 participating organisations, funded by the recent European Union's Research and Innovation Framework Programmes. We analysed over 1100 pages of project documentations and data from 49 semi-structured interviews with project actors, resulting in over 46 h of recording and over 800 pages of transcripts. The existing technology readiness level (TRL) scale was applied to each of the identified innovations. As ambiguities of the levels and their definitions were observed, we refined the scale on the basis of the collected evidence. This provides the basis for an accurate assessment of the initial and the final stage of the development. Based on additional quantitative data, we found that for the set of innovations considered, the *relative* increase of the required efforts between the steps of the refined TRL scale was quite similar for each innovation *independent of their technical specificities*, while the absolute efforts for the development differed significantly. We therefore abstracted from the data values of scores for the increments between the steps of the refined TRL scale, to reflect the progression in the maturation of the technologies according to their difficulty and the required effort. By adding the scores for the maturation of the innovations between the starting point and the end point of their developments within the projects, we assessed the innovation performance of the projects can then be assessed objectively. As the proposed refinement of the TRL scale as well as the scores reflect the domain where the projects were performed (engineering and IT), the precise definition and score values may have to be adapted to other domains, but we consider the general approach to be generic and easily transferable to other types of projects.

The remainder of the paper is organised as follows: we firstly provide the literature review that guided our study. Then we explain the research setting and the research design, present and discuss our findings. We conclude by summarising our theoretical contribution as well as the contributions of the findings to policy and practice, explain the limitations of our research and suggest future research directions.

2. Literature review

2.1. Assessment of the progress of innovations throughout the development stage

2.1.1. Assessment of innovations and of R&D performance

Innovation has been defined as "a new or improved product or process (or combination thereof) that differs significantly from the unit's previous products or processes and that has been made available to potential users (product) or brought into use by the unit (process)" (OECD Oslo Manual, 2018, p. 20). In the development process, which is part of the innovation process and is also of iterative nature (Cooper, 1990, 2007; Ellwood et al., 2020), the innovations are matured, starting from first ideas and lab experiments or software prototypes to demonstrations; this implies changes in complexity, time and effort needed for the steps as well as technological risks (Mankins 2009b; Olechowski et al., 2020), time needed for maturation (Kenley and El-Khoury, 2012), implementation environment, costs (Héder, 2017; Webster and Gardner, 2019; Cooper and Kleinschmidt, 2007), knowledge base, socio-technical infrastructure, user relations, market, policies and regulations (Nakamura et al., 2012). Scholars also found that the involvement of early customers is beneficial for the innovation process (e.g. von Hippel, 1986, 1994; Hurmelinna-Laukkanen et al., 2021; Ritala and Hurmelinna-Laukkanen, 2009; Enkel et al., 2005).

Innovations and R&D performance can be assessed at different stages of the innovation process, e.g. in the phase of the selection of the 'right' ideas and projects (e.g. Markham, 2013; Dziallas, 2020; de Oliveira et al., 2015; Kock et al., 2015) and in the phase of commercialisation, or market launch (e.g. Hittmar et al., 2015; Dewangan and Godse, 2014). These two directions are not relevant to the assessment of the progress of market-upstream innovations that is discussed in this paper: they already have been selected for further development, but on the other hand are not yet in the commercialisation phase.

The advancement of innovations at their development and validation stage has been assessed according to the number of generated patents (Cooper, 1990, 2017; Maietta, 2015; Soh and Subramanian, 2014), but the extant literature highlighted repeatedly (e.g. Janger et al., 2017; Martin, 2016a) that many innovations are not patentable or deliberately not patented. Assessments have been done by analysing the input to the process (Birchall and Tovstiga, 2005; Cooper, 1999), but the input should be compared with the output, the progress of the innovations, but they are still in development, so this refers back to the initial question: how to assess the progress of the innovations.

In the literature on New Product Development (NDP), the advancement of market-upstream innovations is often assessed through the prism of the advancement of the technical features comparing them to development cost (Markham and Lee, 2013; Henttonen et al., 2016; Ojanen and Vuola, 2006), where progress is measured by technical characteristics (Bican and Brem, 2020; Lazzarotti et al., 2011; Chiesa et al., 2009; Chiesa and Masella, 1996). However these characteristics are specific to each innovation. Similarly, the indirect assessment of the advancement of innovations by analysing how adequate the technical solution is in comparison to the needs of the users (e.g. Bendoly, 2014; Hagedoorn et al., 2000) is highly specific.

Scholars concur that assessments in the market-upstream phase are complex and challenging (Henttonen et al., 2016; Birchall et al., 2011). Methods for the assessment of the progress of the innovations in this stage of the innovation process are largely missing in the literature (Aiello et al., 2020). Dziallas and Blind (2019), in a thorough literature

analysis of over 250 papers published in 1980–2015, did not identify any indicator used to assess technologies at the development and validation stage, except patents and other indicators of IP. Literature in this stream also repeatedly highlights the need for a better understanding of the innovation process (e.g., Birchall et al., 2011; Edison et al., 2013), and calls for future research to assess innovations at earlier stages, with the results being useful for researchers, managers, and policymakers (Dziallas and Blind, 2019).

The approach proposed in this paper is based on the idea to assess the progress in the maturity of the innovations during the development and validation stage, over certain periods of time, by comparing the maturity at the beginning and at the end of this period. For measuring the level of maturity, the commonly used indicators are the technology readiness levels (TRL). It is quite surprising, that this indicator has not been paid much attention in innovation studies (Webster and Gardner, 2019) and has not been used yet in quantitative assessments. The main characteristics of the technology readiness levels and some shortcomings which are relevant for this research are described in the next section.

2.1.2. Maturity assessment: Technology readiness level

In order to structure funding schemes, to define handovers between different organisations or consortia at certain levels of maturity of innovations, to manage and reduce technological risks, and to communicate between different actors, NASA formally defined the so-called technology readiness level (TRL) scale in the late 1980s. It consists of 9 levels and describes the *maturity of a technology*, i.e. how ready (mature) it is on its way towards the final use in flight or space (Mankins, 2009a; Peters et al., 2017). These levels were tailored to flight and space programs with very high requirements on the reliability of the innovative technologies and a very high cost of the development of flight-proven systems. Héder (2017) reports that in this area, the cost of moving from TRL 6 to 9 is 90% of the total development cost; however this large increment is domain-specific (Héder, 2017; Olechowski et al., 2020).

The TRL scale is standard now in space and defense-related funding and generally accepted in this domain, but it is also used frequently by the public agencies (e.g. EC, 2014) to define the maturity levels that should be reached by projects. It must not be confused with the commercial readiness level, as TRL does not take into account other factors that are necessary for rollout and introduction of innovation to the market. It should also be noticed that IP protection, such as patenting, is not directly ‘attached’ to the TRL: sleeping or blocking patents (Torrisi et al., 2016) do not facilitate the maturation of related innovations; on the contrary, incremental innovations often progress without patenting (Martin, 2016a).

The TRL definitions leave quite some room for interpretations (Cornford and Sarsfield, 2004; Tan et al., 2011; Peters et al., 2017), often leading to subjective TRL assessments. A discussion of the difficulties of the transfer of the TRL scale to other sectors can be found in Héder (2017); other scholars also state that the TRL scale is very domain-specific, does not reflect nonlinear innovation process, and it does not necessarily work with diverse innovation processes and product cycles (e.g. Tan et al., 2011; Peters et al., 2017). Olechowski et al. (2020) present a study of the adequacy and the deficiencies of the TRL scale mostly from an innovations management point of view. From the feedback from different industries obtained by interviews and questionnaires, they extract 15 challenges of TRL implementations, structured into the domains of system complexity, planning, and review and assessment validity. The domain of system complexity in particular relates to large systems which are made up of many components. Here the issues of integration and connectivity, maturity of interfaces, lateral effects and overall system readiness level are prominent: the complexity

increases with the increase of maturity, but this is context-dependent. Olechowski et al. (2020) report that the ambiguity of the TRL definition increases at the higher TRL levels.

Mankins (2009b) discusses the relationship of TRL and technological risks. The general observation is that the technological risk decreases towards higher TRL. He proposes a risk matrix as a basis for risk management that represents the probability of failure and the required additional steps on the TRL scale, the larger the number of TRL steps that are required until the innovation is technologically ready, the higher the risk.

To investigate the questions of the assessment of market-upstream innovations in the project setting, we also draw on the literature on assessing innovation performance in market-upstream publicly-funded RDI projects.

2.2. Assessing innovation performance in market-upstream publicly-funded RDI projects

Public agencies around the world support RDI projects, often with considerable numbers of heterogeneous actors, that are generating new knowledge and innovations at the same time (vom Brocke and Lippe, 2015), with the goal to provide evidence of successful validation, test, demonstration or application of innovations that are still market upstream. For this purpose, the projects often include early customers (which can also be internal customers in large enterprises, see Section 2.1.1), i.e. end users supporting the maturation process of innovations, e.g. providing data, or introducing them into operational environments, thus speeding up the maturation of innovations. As an example, the previous European Research and Innovation Framework Programme (FP), Horizon 2020, has allocated 39,64 B€ in 2014–2021 to fund over 7000 such projects (EC, 2021a). Accountability of the projects, funded with taxpayers’ money, is important; and assessments also shall provide actionable intelligence, suggesting directions for improvement (Perkmann et al., 2011).

The performance of RDI projects has been assessed so far mostly from the ‘research side’ (e.g. Grimaldi and Von Tunzelmann, 2002; Fernandes et al., 2017). Assessment of the innovation performance requires the definition of the baseline and the end point of the developments within the project which often are not clear (Perkmann et al., 2011; Kim et al., 2017; Aguiar and Gagnepain, 2017). The number of innovations that are pushed forward within the same project often increases with the increase in the number of project organisational actors having own objectives (Mannak et al., 2019): this makes the assessments of the overall innovation performance of the projects even more challenging. The advancement of the innovations in the projects is reflected in the key performance indicators (KPIs) of the programmes, ranging from the number of testing activities and demonstrated prototypes to the number of deployments of innovations in operational environments and new products close to the introduction to the market (e.g. EC, 2015; Spanos and Vonortas, 2012; Kostopoulos et al., 2019). These KPIs indeed reflect the maturity level of the innovations at the end of the projects. As Héder (2017, p.10) states, in the interpretation of the EU, “Technological Readiness Level is a metric that shows how far a technology is from being ready for use in its intended operational environment” - this is what makes it attractive for the topic of this paper: how can the progress of innovations and thus the innovation performance of the projects be measured. However, extant studies of the RDI projects have also shown difficulties to use the coarse TRL scale in the RDI projects due to the differences in the interpretations of the TRL terminology (Olechowski et al., 2020), and also in case of the presence of several innovative technologies of different maturity in the project (Enzing et al., 2015). An

important aspect is that in order to evaluate the performance of a project, not the finally reached maturity level but the attained increase in maturity should be considered for a fair evaluation.

This literature stream reflects the same difficulties to assess progress of market-upstream innovations as explained earlier in Section 2.1.1. The available objective data to measure performance in the middle of innovation processes is limited (Schwartz et al., 2012; Enzing et al., 2015). Assessment of the advancement of innovations in the projects are done indirectly, by counting patents, which is not satisfactory (Hung, 2017), or comparing between achieved and planned promised outputs (vom Brocke and Lippe, 2015; Perkmann et al., 2011), which makes sense from a project management point of view but does not measure the progress of the innovations, or subjectively, from the point of view of the project actors (e.g. Arroyabe et al., 2015; Arranz et al., 2018; Spanos and Vonortas, 2012; Kostopoulos et al., 2019), but this should be complemented by objective measures (Perkmann et al., 2011). The innovation potential of the innovations developed within the projects is also assessed by the “Innovation Radar” (EC, 2021b) which is used to categorise innovations in terms of market maturity and disruptive potential, but it does not assess the progress of innovations during the projects.

To sum up, our literature review points out to the underlying issue salient in different literature streams: the difficulty to assess the progress of innovations at their development stage and by extension the difficulty to assess the innovation performance of the RDI projects directly and objectively. Recent calls go in the same direction, calling for the assessment of innovations at their early stages (Dziallas and Blind, 2019), pointing out the needs to take into account the progress of non-patented innovations (Martin, 2016a), calling for research to provide a more fine-grained, nuanced view on innovations which will benefit the innovation research (Collinson and Liu, 2019), highlighting a need for “research on innovation outputs ... measurement which will eventually contribute to reducing uncertainty in policymaking” (Janger et al., 2017, p. 40), and to better policy learning.

In this exploratory study we investigate the progress of market-upstream innovations and pose the following research question: How can the advancement of innovations at their development and validation stage be assessed through the lens of technological maturity, objectively and directly, as well as the performance of the RDI projects on their ‘innovation side’?

3. Research setting

The research setting is defined as innovations developed within collaborative RDI projects funded by the European Commission in the context of its two previous Framework Programmes of Research and Innovation, and specifically its LEIT (Leadership in Enabling and Industrial Technologies) Programme (EC, 2020) in Horizon 2020 and its predecessor, FP7. The projects, in addition to generating new scientific knowledge, target the development of technological innovations of generic nature, market-upstream, that are often tested or deployed in pilot applications within the consortium where partners act as early customers. The projects often build on previous work that has led to different levels of maturity of the innovations, usually starting at least at TRL3 (proof of concept in lab or in computer simulations). Funded projects usually last three to four years and are carried out by inter-organisational consortia with multiple actors. This setting is particularly rich and suitable for this exploratory study: the projects in most cases generate multiple innovations and access to the project data provides access to the data on all these innovations.

4. Research design

This research is part of a larger study that takes a deeper look into

multi-actor RDI projects, focusing on their architectures (Klessova et al., 2020), size, processes and generated innovations. The research design is an exploratory inductive multiple case study (Eisenhardt, 1989), with additional quantitative evidence. Exploratory case studies (Yin, 2017) are needed when there is a lack of detailed preliminary research and of hypotheses to be tested (Mills et al., 2010). Multiple case studies provide a stronger base to build theory (Yin, 2017), enable comparison and demonstration of replication (Eisenhardt, 1991), and create more robust theory (Eisenhardt and Graebner, 2007). The replication logic here is meant as the analysis of each case on its own first, testing the emerging theory in each case separately, and focusing on “hard to measure constructs” (Gehman et al., 2018, p. 287). While quantitative studies are looking at a representative sample to make generalisations, the logic of multi-case qualitative studies is to enable analytical generalisation (Yin, 2017): the goal is to identify patterns which could confirm the presence of similar phenomena across cases, to generalise them further to theory (Eisenhardt, 1989). In the words of Eisenhardt, the study was based on “deep immersion in multiple kinds of data ... that help reveal the focal phenomena” (Gehman et al., 2018, p. 288).

The selection of the projects was done at the first stage of the larger study (Klessova et al., 2020). At that stage, we used 4 selection criteria: homogeneity, variety, comparability and availability of data. For the present research, we focused on 5 projects; some were more research-oriented, others more innovation-oriented, thus the innovations in the projects started from and were advanced to different levels of maturity. This constituted theoretical sampling: the selected cases are “particularly suitable for illuminating and extending relationships and logic among constructs” (Eisenhardt and Graebner, 2007, p. 27). Each of the projects and the innovations were unique, but the whole sample contributed to theory building, by enabling the recognition of patterns of the advancement of innovations, and a comparison between them. The five projects were from the engineering domain with a strong IT dimension (4 projects) or only in the IT domain (1 project). The choice was a result of the availability of data, of a choice made for the sake of comparability between the projects, and motivated by the pervasive presence of IT aspects in technological developments. All projects had a strong application dimension and a duration of 3–4 years. Summaries of the selected projects and the maturity of their innovations are presented in Appendix 1, where names of celestial constellations are given to the projects for their anonymisation: HERCULES, PEGASUS, GEMINI, PERSEUS, and ORION.

The 54 innovations pursued in the five studied projects were identified using the project documentation; they included innovative technologies, methods, processes, tools or models, physical devices, and software. These innovations can be characterised as non-radical, in all 5 projects they had to be embedded into larger systems with a high degree of technical complexity and many human actors involved. The existing environment consisted of specific hardware (e.g. machines, processing equipment, transport units) and/or IT systems. The interfaces to physical systems, IT systems and human users were important elements of the development. The reliability of the solutions needed to be high, for routine and in most cases daily industrial use, but not ultra-high as e.g. if rolled out in cars or aircrafts. The development processes started from existing foundations created by previous work and from different levels of maturity, the goal was to advance the maturity of the innovations. For almost all of the innovations (except one), no attempts were made to obtain a patent; the need for patenting was mostly discussed as a means to obtain freedom to operate, but it was preferred to achieve this by publications. One reason for this was explained by a project actor:

“We realised it’s such a large effort to implement it, that being first on the market is far more important than having a patent”. ID6_HERCULES_large ind

Table 1

TRL scale (EC, 2014) used for the initial identification of the maturity of innovations.

TRL 9 – Actual system proven in operational environment
TRL8 - System completed and qualified
TRL7 - System prototype demonstrated in operational environment
TRL6 - Technology demonstrated in relevant environment
TRL5 - Technology validated in relevant environment
TRL4 - Technology validated in lab
TRL3 - Experimental proof of concept
TRL2 - Technology concept formulated
TRL 1 – Basic principles observed

In the *data collection stage*, we collected empirical data about the innovations (the unit of analysis in this study) and their advancements, using over 330 pages of project proposals and over 780 pages of confidential project reports and other internal documents, e.g. exploitation plans, Innovation Radar documents (EC, 2021b) when available, registers of project IPs, meeting agendas and minutes, periodic reports ... In addition, for triangulation purpose, we also collected data on efforts, as periodically reported by partners (person months spent, according to the timesheets,¹ which corresponds to costs). These data were reported per workpackage; we thus had to reconstruct the effort per innovation with the help of the project partners in some cases. We thus were able to identify the maturity level of the innovations *at the end of the projects*, following the TRL scale from the literature (Table 1), and the relevant efforts. We then checked the *starting level* of each of these innovations in the projects, i.e. the baseline. We controlled ‘optimistic biases’ (Prater et al., 2016) that may alter the identification of the level of each innovation by checking the documents retrospectively, comparing with the initial level indicated in the final documents and by verifying initial and final levels with the project coordinator and/or one of the key partners.

Additionally, we collected data during 49 semi-structured interviews with 31 project actors. In each project, representatives of between 40% and 65% of the participating organisations were interviewed. We used theoretical sampling (Eisenhardt, 1989) to select the informants, using two criteria: different profiles of the organisations (research organisation, large industrial firms, SMEs), and different roles of the informants in the developments of the technologies (technology developer, or early user, contributing to the advancement of maturation of the innovations). Based on these criteria, 33 project actors were selected and invited to extensive interviews, with provision of detailed information about the study; only 2 of them (academic partners) were not available. In order to obtain comparable data, over 80% of the informants were CEOs of SMEs, team leaders within large corporations, or tenured university professors. A detailed modular interview guide was developed, capturing, among other, the qualitative judgements about the project results, progress in degree of maturity of technologies, paths of progression (steps passed or not, reasons, activities and difficulties experienced during the progress of technologies ...), importance of progression, relative efforts comparing between the stages. The interviews were organised in several rounds, some were complementary. All interviews were recorded, with prior consent of the informants, and transcribed verbatim. This resulted into over 46 h of recording and over 850 pages of transcripts (Table 2).

In the *data analysis stage*, several steps were performed. The first step aimed at understanding the progress of each of the 54 innovations, and whether their start and end maturity level can be *homogeneously and unambiguously* identified using the existing TRL scale. When the project

actors struggled to identify the levels, or when the levels comprised heterogeneous degrees of maturity of the innovations (visible from documentary analysis), we searched for the reasons for the confusion and noticed the need for a finer-grained classification.

The data analysis was based on the abstractive process of open coding (Strauss and Corbin, 1998, adapted by Gioia et al., 2013). The Eisenhardt Method (Eisenhardt, 2021) is often used in multiple case studies, and the Gioia Methodology (Gioia et al., 2013), with first order data and second order themes, is considered to be well adapted to single cases. In this study, we built on the similarities between the methods: deep diving into the phenomena, theoretical sampling, and grounding of the concepts in the data in a plausible and realistic way, for theory building (Gehman et al., 2018). We thus combined two methods to analyse and present the structured evidence for the conclusions and to systematise the thinking: the approach of Eisenhardt (1989) provided the general framework of this research, while the Gioia methodology was mobilised to pass from the data to the concepts, and to highlight the hidden part of the elaboration of the concepts and the interpretation of the investigator. We grouped 22 concepts of 1st order into 10 more abstract themes of the 2nd order, which were combined into 4 aggregate dimensions (see the data structure in Appendix 2). The explanations provided by the project actors about the levels of developments and about the advancement between the levels were analysed; this resulted in a finer-grained scale for the advancement of the innovations which was augmented by an exemplary *scoring system* (weighing of the steps between levels), according to the average efforts (human resources) needed to progress in the maturity of the innovations from one level to the next. The result was compared across the innovations and between the projects. The scores were developed using both theoretical constructs, knowledge of the field, and empirical data, i.e. qualitative feedback triangulated with quantitative data for selected innovations in three projects (38 innovations): these projects were from a comparable domain (engineering with a strong IT component) with comparable type of the funding scheme (‘research and innovation action’). We took care to make sure that the scores reflect the “objective” effort that is needed to progress, i.e. not taking into account contextual conditions related e.g. to collaboration issues or external factors: we therefore particularly considered those innovations where the collaboration was smooth.

After it had been developed, the scoring system was applied to all innovations in the largest project, ORION, where a broad range of maturity levels of technologies was achieved. We sent the working paper, presenting the fine-grained scale and the scores (the framework), to the ORION project partners, collected comments, and presented and discussed the revised results at a meeting with 36 project actors. The framework was agreed upon by the participants, with some suggestions of adjustments and clarifications which were taken into account. Then the materials (scale and scores based on the previously collected data) were sent to key project partners and discussed with them. Once the refined scale and the scoring were available, compared and analysed, the assessment framework was applied firstly to measure the advancement of the individual innovations, and then to measure the overall innovation performance of the projects.

5. Results

5.1. Refinement of the maturity scale

5.1.1. Identified ambiguities of the TRL scale

The straightforward application of the TRL categories to classify the identified innovations was difficult for the reasons already suggested in the extant literature: the existing categories were not homogeneous, did not fully fit to the studied innovations, and the definitions were partly

¹ We considered person months as minimum and indicative parameter: for instance, industrial partners did not necessarily book efforts of technicians on the project, although these staff was involved in the projects; efforts of organisational support were not reported in person month, as it was accounted as overheads; sometimes, the project budget was reached, and the effort of the project staff was not recorded completely.

Table 2
Overview of interviews per case study.

Anonymous Name	Partners (organisations, number)	Interviewed organisations per case		Interviews (incl. compar.), number	Number of interviews per profile of informants				Recording, min.	Transcripts, pages
		Number	%		Research	Industry		Consultant		
						Large	SME			
HERCULES	10	4	40	9	3	4	2	0	548	168.8
PEGASUS	12	6	50	7	3	3	1	0	384	109,1
GEMINI	13	6	46	7	1	1	4	1	447	134.1
PERSEUS	15	7	47	10	6	4	1	0	638	189.0
ORION	17	11	65	16	5	5	5	0	787	258.6
				49	18	17	13	1	2804 (46.7 h)	859.6

confusing for the project actors and not specific enough:

“For me, it was an operational environment in the pilot, but it wasn’t a system. For him, it was a system but not in an operational environment.”
ID52_GEMINI_SME

Because of the nonlinear nature of the innovation process, some TR levels were skipped: e.g. some innovations did not go through the near-operational environment ...

“TRL level 6, does it exist, I don’t know for us. We went directly from five to seven”. ID51_ORION_SME

... while in another project, a component was demonstrated and then integrated into a system only later on, at a higher level of maturity.

Several problems of the TRL scale emerged, as pointed out by the project actors (Appendix A, on-line supplementary materials). Categories at the high maturity level describe different achievements that required different efforts: (1) Scale: components versus systems; (2) Type of the environment where the innovations were simulated, validated, demonstrated or deployed: synthetic, near-operational, operational; (3) Number of deployed applications: single or multiple; (4) Type of product: improved own product versus completely new product; (5) Type of deployment: temporary or permanent. Thus, more differentiation between categories was needed, comparing to the existing TRL scale, especially on the higher TRLs, in line with the observations reported by Olechowski et al. (2020).

5.1.2. Fine-grained categories of maturity levels

Based on interviews and analysis, we refined the categories of the scale to make it fit better to the innovations from the projects. In what follows, we proceed from the highest to the lowest considered level (also see illustrative quotes in Appendix B, on-line supplementary materials).

Refinement of the category “System completed and qualified” (TRL8). The wording here comes from the world of aircraft, spacecraft, also automotive. In the context of the RDI projects analysed in this study, the highest achieved level of maturity was that of a beta-version (prototype) of a product, or of a significant improvement of an existing product, given for testing to potential users or customers. Another case of a high level of maturity was that the innovative technology was rolled out within the productive environment of one or more project partners, e.g. it is actually used in day-to-day operations in an industrial plant over an indefinite period of time. Both new and improved products and a rollout in an industrial environment require a high level of robustness and reliability but there is a difference between a product which is given to several external users and an internal solution, where some updates and adaptations are tolerable. Still the requirements in all cases are significantly higher than for a pilot application (see below), in particular there must not be a need for frequent maintenance and the necessary

features for long-term productive use must all be present. As mentioned above, in all cases of interest, the innovative technologies had to be integrated into a complex environment with which they interacted intensely. Consequently, we refined the category TRL8 into 4 new sub-categories: *Rollout of one application (TRL8.1)*, *Rollout of multiple applications (TRL8.2)*, *Prototype of improved own product given for testing to trial customers (TRL8.3)*, *Prototype of new product given for testing to trial customers (TRL8.4)*.

Refinement of the category TRL 7, “System prototype demonstrated in operational environment”. The notion of prototype is vague, prototypes can be anything from a first realisation to a serious implementation. We therefore use the term “pilot deployment” which is common in developments for and in industry, in line with the feedback of the project actors. The understanding of a “pilot” is a realisation of the innovative technology that has all key features that are needed to prove its value in the real application environment. The purpose of a pilot is to demonstrate that the expected benefits materialise and that it integrates well into the workflow and the hard- and software environment. Pilots requires an integration effort into the environment where the pilot is deployed. Deployment means that the pilot is in operational use for a significant period of time, but while and after gaining experience it may have to be developed further to better meet the requirements of the operational environment, before rollout. It can also happen that the pilot is not developed further into a permanent solution, e.g. because it did not fully meet the expectations, or the benefits did not warrant the necessary development efforts towards a long-term stable solution, or because there was no convincing concept for continuous maintenance and adaptation.

“Pilot deployment” was refined into two sub-categories: *Deployment of one pilot (TRL7.1)* and *Deployment of multiple pilots (TRL7.2)*.

The differences between the categories “Technology demonstrated in relevant environment” (TRL6) and “Technology validated in relevant environment” (TRL5) are not obvious from the wording. According to Oxford Reference,² validation means “The act or process of making something valid, ratifying it, or checking that it satisfies certain standards or conditions”, other sources define it as a proof of correctness. Demonstration can be understood as showing the working of the technology to the users. If one goes back to Mankins (2009a) where the TRL scale is discussed in detail, the wording is: TRL5 “Component and/or breadboard validation in relevant environment” and TRL6 “System/subsystem model or prototype demonstration in a relevant environment (ground or space)”, and in the discussion, the difference between the two categories mostly is in the definition of the «relevant environment».

To clarify the difference between the categories, we use the term

² <https://www.oxfordreference.com>.

Table 3

Refined maturity scale. The blue color shows the refined categories.

Categories of innovation outputs (innovations)	Short explanation of the category	Means of identification / verification	Corresponding TRL
PRODUCT	Products refers to the external use of innovations by (potential) customers		
TRL8.4 Prototype of new product given for testing to trial customers	New software, hardware (e.g. a device) or method that can be offered to trial customers as a product, i.e. is robust and mature enough for guaranteeing the correct function when used by customers, or as a standardized service to customers.	Innovative solution has passed the stage ‘pilot deployment in operational environment’ and has a defined short time horizon for putting the product to the market	8
TRL8.3. Prototype of improved existing product given for testing to trial customers	Same as above, but new function or element of a service integrated into an existing product. Example: New feature in a large software system. Could also be a new model library as part of a larger software system that is planned to be marketed commercially.	Same as above for improved product	8
DEPLOYMENT OF INNOVATIVE TECHNOLOGIES IN THE OPERATIONAL ENVIRONMENT	In contrast to products, deployment refers to the internal use of innovative technologies by project partners by implementing them in their daily operations (operational environment).		
Rollout (permanent deployment in operational environment)	Project partners implement the innovative technology from the project for long-term productive use within the organisation. The installation must be robust, errors and need for maintenance occur only rarely.	In operation for several months within the duration of the project.	
TRL8.2 Rollout of multiple applications	Several installations of the innovative technology in permanent use.	See above	8
TRL8.1. Rollout of one application	One installation of the innovative technology in permanent use, typically a continuation of a pilot.	See above	8
TRL7. Pilot deployment (temporary), or integration into the operational environment	Pilot: Putting the innovative technology to work in the operational environment over a significant period of time to prove its potential and to provide information on further necessary improvements. Still frequent interactions with the developers. Pilots can also be done by technology providers (e.g. project partners) at their customers. Example: Novel measurement device installed temporarily at a production plant. Pilot means an implementation for validation of the benefits by the end users, and requires integration in their operational environment under their supervision. Must be significantly more stable and user friendly than a prototype.	The technology is deployed temporarily and tested over a significant period of time. The purpose is to demonstrate the usability and the benefits to the end users and to generate feedback on necessary further improvements.	
TRL7.2. Deployment of multiple pilots	As above, multiple installations.	As above, multiple installations.	7
TRL7.1. Deployment of one pilot	As above, one installation	As above, one installation	7
PROTOTYPE VALIDATION AND DEMONSTRATION	A technology prototype is an intermediate step in the development process where the essential features are present but the pilot implementation in the real environment needs further improvements. The installation is temporal and provisional. Only the essential features must be present.		
TRL6. Prototype demonstrated in operational or near-operational environment	A prototype of an innovative technology is demonstrated in an operational or near-operational environment, or operated with real-time data from the operational environment. Near-operational environment means pilot facility or other setting that exhibits the main features of the operational environment, i.e. the environment where the innovation will operate when deployed as a pilot. Demonstration means that the prototype functions according to the specification when integrated with the real equipment or using real-time data that is provided by the end users (usually project partners). The goal is to get the green light to progress to the pilot deployment.	Demonstration of a prototype in operational or near-operational environment over a sufficiently long period of time to decide on the next steps in the development process.	
TRL6.2. System prototype demonstrated in operational or near-operational environment	See explanation above. The system prototype consists of several elements (components) that have to be integrated into a system to perform the demonstration. A system prototype requires efforts to integrate components. Example: Measurement system that consists of hard- and software and is connected to the hardware and IT systems of a pilot facility.	See above: demonstration of system prototype with at least two components	6
TRL6.1 Component prototype demonstrated in operational or near-operational environment	The prototype consists only of one new component. Example: Software that provides one novel function	See above: demonstration of component prototype	6
TRL5. Prototype validated in a synthetic environment	Validation means that the prototype meets the success criteria for moving on to an installation in the operational or near-operational environment and adding additional features. Synthetic environment can mean a test rig or a virtual environment, e.g. a computer simulation that represents (simulates) activities at a high level of realism, and enables extensive testing under realistic conditions.		
TRL5.2 System prototype validated in a synthetic environment	Requires effort to integrate validated components into a system and to validate it.	Prototype can be operated in the synthetic environment and provides the specified functionality.	5
TRL5.1 Component prototype validated in a synthetic environment	Validation of one component in the synthetic environment, e.g. extensive testing of a piece of software with realistic data.	See above	5
TRL4. Technology validated in lab or testing environment (computer)	Technology prototype validated in the laboratory or testing environment, i.e. the basic functionality is proven under ideal conditions or in simple setups.	Serious tests under laboratory or simulated conditions are successful.	4
TRL3. Proof of concept	The principles and methods that were proposed are shown to work in principle for the targeted application	Preliminary tests are successful.	3

“*Prototype demonstrated in operational or near-operational environment*”, which is understood as the prototype working successfully in a (temporary) installation at an industrial site or in a demonstration setting (pilot plant), or working with real live data from the application in case of software solutions. A prototype is a temporary, simplified realisation of the innovative technology (component or system) that does not have to have the full set of features (e.g. it could only have a rudimentary user interface or, if the focus is on the user interface, the data processing may not be fully robust yet). In contrast to the pilot, a prototype installation is done under the assumption that the prototype will be removed and replaced by an operational installation (the pilot) later in the process. Demonstrated means that the main functionality of the prototype was in line with the technical requirements in the operational or near-operational environment. We subdivided this category into two: *Component prototype demonstrated in operational or near-operational environment (TRL6.1)* and *System prototype demonstrated in operational or near-operational environment (TRL6.2)*. A system prototype integrates one or more innovative components into a larger system as well as with the environment.

The category “Technology validated in relevant environment” (TRL5) in the original literature differs from TRL6 mostly in the type of environment where the validation takes place. We precise it here by stating that the prototype is demonstrated in a *synthetic*, but realistic environment (mock-up), or in extensive computer simulations. This is a standard step in the development process, where, for the first time, a solution is shown to work as a device or a piece of software, under serious realistic testing conditions, e.g. in a test-rig, or coupled to a computer simulation (e.g. hardware in the loop simulation) or in extensive tests. This test in a synthetic environment may come with efforts for a first integration of several components or with the (synthetic) environment. Therefore, this category was split into two sub-categories: *Component prototype validated in a synthetic environment” (TRL5.1)* and *System prototype validated in a synthetic environment” (TRL5.2)*.

The categories “Technology validated in lab” (TRL4) and TRL3 (Experimental proof of concept) were kept, but the explanation was revised, to fit it to digital innovations, so the word “*experimental*” was dropped for TRL3 and “*or testing environment (computer)*” was added for TRL4. TRL4 is understood as showing that a first realisation of the innovative technology performs as expected in a laboratory setting or with realistic data sets, while TRL3 is understood as providing sufficient confidence that the envisioned technology will be applicable to the problem at hand by means of simple tests or computations. TRL 3 was the lowest starting point to generate innovations on higher TRLs within the studied projects.

Putting the categories together: a fine-grained scale. We combined the refined categories into a finer-grained multi-dimensional scale (Table 3). Instead of 6 categories of maturity, the refined scale now has 12 sub-categories. The purpose of the refined scale is to identify in a more precise way the initial and the final maturity levels, independent of iterations between the levels. The labels of the levels reflect the common ground of the projects that were included in the study, in other domains slightly different terms may be used. Then we classified the initial and the final level of maturity of the 54 innovations according to the finer-grained scale (Appendix C, on-line supplementary materials).

5.2. Efforts for the maturation of the innovations

We firstly assessed the efforts for maturation from the initial to the final stage per innovation, in terms of the average amount of the human resources, related to the progress in the level of maturity. Both the amount of efforts and nature of effort changed with the increase of

maturity:

“*The academic prototype is ten percent of the work; but getting it into application is ninety percent ... For me, the most memorable point (of the project) was to figure out how large the gap between the research and implementation really was ... There was a moment when I talk with (multinational firm, project partner), they said okay doing this if we put everything behind, it will take 30 years ... This is sort of level ... it's at the TRL four or five now*”. ID9_PEGASUS_SME

The innovations and the development processes were different, and thus the efforts for the different steps and for different innovations were different. For instance, the internal integration (from level 5.1 to 5.2 or 6.2) in some cases, e.g. for pure software development was much more demanding than the pilot deployment (integration into the operational environment of the end user, progress to level 7.1) while in other cases, the internal integration was less difficult than the pilot deployment. Progressing from level 7.1 (pilot installation in one case) to 8.1 was described as easier if the application is the same, then the progress to multiple rollout required higher additional efforts compared to going from 7.2 to 8.2. Also some pilots were already quite refined (e.g. had an almost perfect user interface in order to get the approval of the end users), while others were preliminary in some aspects for a faster evaluation in the sense of rapid prototyping, thus shifting some effort to the next level.

While the different innovations required different *absolute* efforts, we found that, for the same innovation, the increase of the efforts comparing with *previous* efforts, i.e. the relative increase of the efforts, was comparable across the innovations, as the additional challenges for the progression on the maturity scale were similar. The development passed from a protected environment (own computer, or lab), to the near-operational environment, then to operational but temporary environment (pilot(s)), and then, for a limited number of innovations, to a permanent deployment in an operational environment of a partner (rollout), or to the operational environment of first users. Additional factors such as e.g. inefficient collaboration, or specificities of the technology, affected the efforts needed for specific innovations, but for a larger set of innovations, this averaged out (see Fig. 3 below).

5.3. A scoring system to assess the progress in the maturation of innovations

Based on these findings, to measure the progress in the maturation of the innovations, we employed the metric of average relative efforts per innovation, compared to the effort required to progress the same innovation from TRL3 to TRL4, and developed a non-linear scale that assigns different scores to the progression from one level of maturity to the next (Table 4). So the progression by one TRL obtains a different score depending on the initial level. When (one or several) innovative components are integrated into a larger system, this is reflected by increasing scores (see Table 4 below).

These scores reflect the (average) relative increase in the efforts and resources that is required to advance from one level to the next when innovations are brought to higher levels of maturity. They do not represent the absolute effort. The values of the scores were derived from the qualitative judgements of the project actors and were calibrated using quantitative data about reported resources of partners in the projects (see below). These scores are supposed to reflect the objective effort that is needed to progress, not taking into account contextual (non-technical) conditions. The logic of the constructions of the scores for the steps of advancement is visualised in Fig. 1; an example can be found in Fig. 2.

Table 4

Incremental scores for reaching different stages of maturity of innovations: from TRL3 to TRL8.

From	To	Score	Minimum or maximum range of relative efforts, as observed in the study	Remark
TRL 3 Proof of concept	TRL 4 (Technology validated in lab or testing environment, e.g. software validated for representative data)	1	1	Lowest initial stage for the scale in the studied projects.
TRL 4	TRL 5.1 (Component prototype validated in a synthetic environment)	2	2	
TRL 5.1	TRL 5.2 (System prototype validated in a synthetic environment)	1 per component	0.5 to 2 (for complex system, equal to the number of components integrated into the system)	5.2 is an optional level. This score is only added if the integration takes place on this level. It can also be accounted for on the next higher level.
TRL 5.1	TRL 6.1 (Component prototype demonstrated in operational or near-operational environment).	3	2 to 4	This is counted for each component of a system.
TRL 6.1	TRL 6.2 (System prototype demonstrated in operational or near-operational environment)	Additional 1 per new component	0.5 to 2 per component	TRL 6.2 is needed for progress to higher levels as the innovations on these levels are always systems. If integration is done on level 5.2, there is no additional score here.
TRL 6.2	TRL 7.1 (Deployment of one pilot)	6	4 to 7	
TRL 7.1	TRL 7.2 (Deployment of multiple pilots)	3	2 to 3	7.2 is an optional level
TRL 7.1	TRL 8.1 (Rollout of one application)	5	4 to 6	
TRL8.1	TRL8.2 (Rollout of multiple applications)	3	2 to 4	Level 8.2 can also be reached going from 7.1 to 8.1 and then to 8.2
TRL 7.1	TRL 8.2 (Rollout of multiple applications)	8	6 to 10	
TRL 7.1	TRL 8.3 (Prototype of improved own product given for testing to trial customers)	8	4 to 8	
TRL 7.1	TRL 8.4 (Prototype of new product given for testing to trial customers)	12	8 for software	

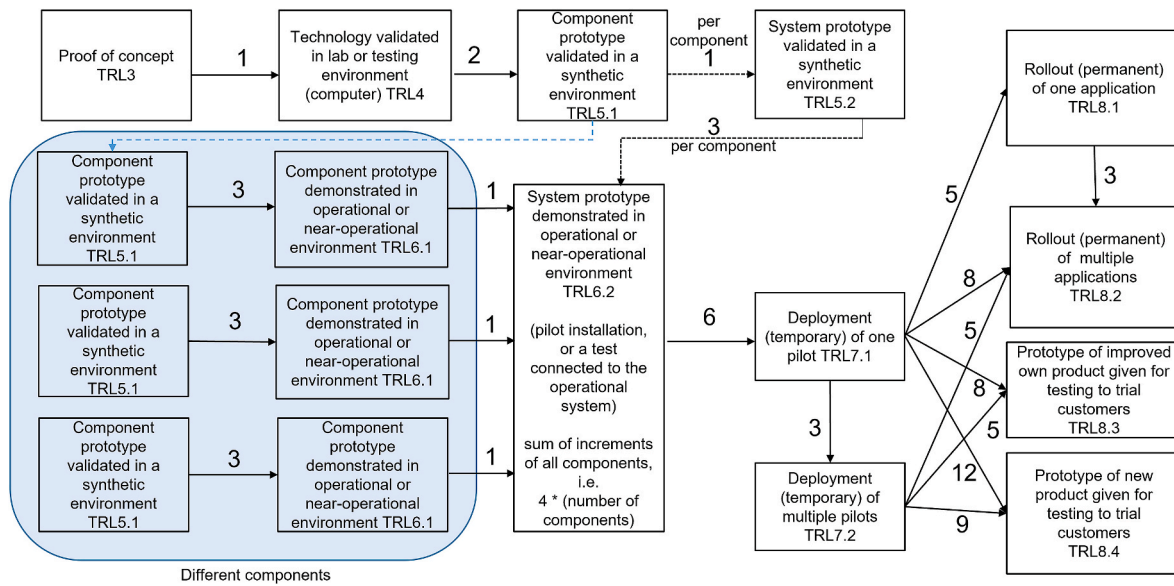


Fig. 1. Logic of construction of scores.

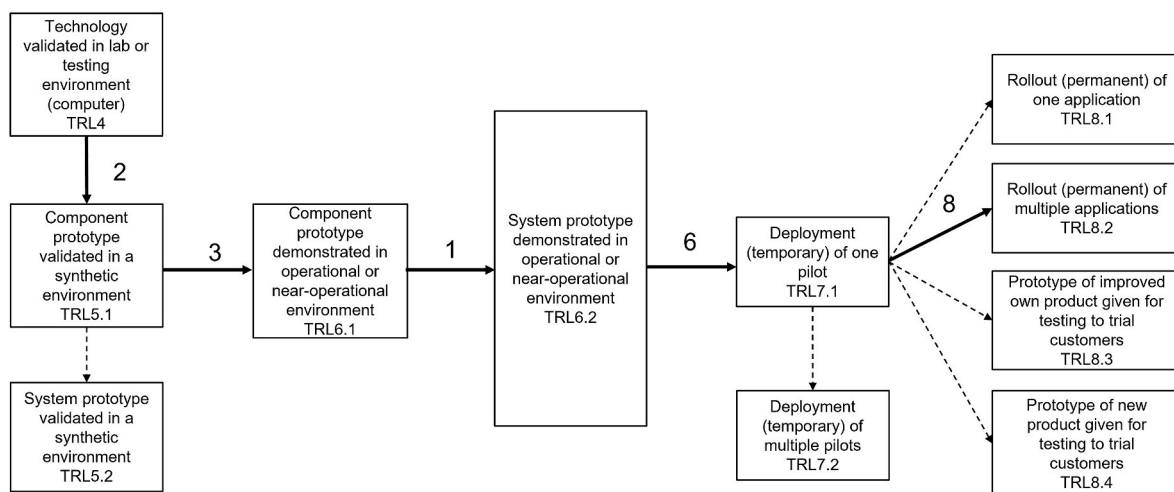


Fig. 2. An example of maturation of an innovation and related scores. From TRL4 to TRL8.2 (rollout of multiple applications), the total score is 20: 2 + 3 + 1 + 6 + 8.

During the last round of interviews with project participants, we collected feedback on the proposed scores, whether they reflect the experience of the project actors in the studied projects. The feedback was that the scores were in line with their experience, with some variations between the innovations: these were taken into account, and as a consequence, minimum and maximum relative efforts are also shown in Table 4 (see also Appendices D and E, on-line supplementary materials). As expressed by the project participants, the anticipated higher effort may lead to a re-evaluation of the ratio of the expected benefit to the effort.

When an innovation is advanced over several stages, the scores from the initial stage to the final stage of the development process can be summed up, giving the total advancement of this innovation during the project or over a certain period of time.

The assignment of the scores was validated, to the extent possible, by the quantitative data. We calculated the scores for each innovation using Table 4, and compared the scores with the estimated input in terms of person months per innovation, investigating 38 innovations from three projects,³ PERSEUS, HERCULES and ORION. In Fig. 3, these values are plotted over the scores. In accordance with the literature and the interviews, the efforts were higher when increasing the level of maturity at higher maturities further, so a curve of the required effort over the increase in maturity would be non-linear. The scores were defined to reproduce this behavior, so when the efforts are plotted over the scores,

³ See Section 4 for the explanation on how the input per innovation was estimated from the available data, and the reason of choosing these three projects.

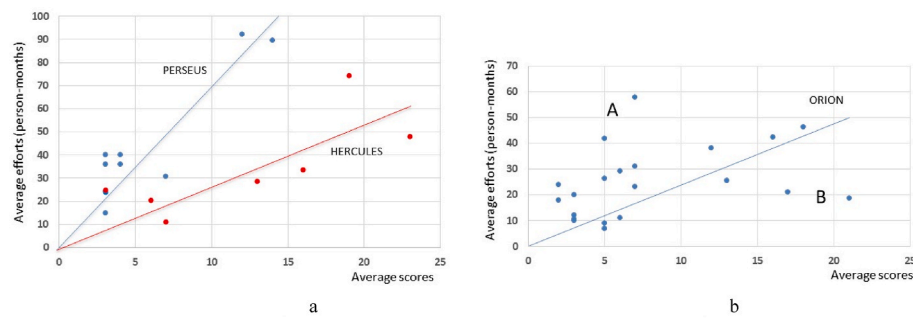


Fig. 3. Total scores for the maturation of the innovations versus absolute efforts invested in the projects for these innovations: (a) PERSEUS and HERCULES; (b) ORION. The trendline is shown per project.

a linear relationship should result. In Fig. 3, the data for PERSEUS and HERCULES on the average follows the postulated linear relationship (Fig. 3a). The slope of the linear trend gives the average effort for a score of 1, due to the differences between the projects, these values differ. For the ORION project (Fig. 3b), the overall trend is linear as well, but larger deviations were observed for some innovations. This is due to a several factors: the difficulty to attach the efforts to the innovations⁴ and inaccuracy in the reporting, variability due to differences between technologies, qualifications of the people involved and efficiency of collaboration. The data confirms the claim that the increase in the required effort to progress from one level to the next is relatively constant while the absolute effort may differ.⁵ In Fig. 3b, A refers to innovations where groups of several partners worked on a particularly difficult and challenging problem, B refers to innovations that were developed by single partners.

5.4. Towards an assessment of the innovation performance of the projects

Adding all scores for all innovations in the projects results in the ‘innovation performance score’, or overall performance of the project on the ‘innovation side’, reflecting both the number and the progress in maturity of the innovations. The summary of the results, both in terms of numbers of innovations and scores for their advancement, are shown in Table 5.

The table shows that the overall innovation performance of the five projects differs significantly when observed with the lens of “numbers” and with the lens of “scores”. For example, both PEGASUS and HERCULES generated 7 innovations, but their advancement (innovation performance score) in HERCULES (87) is almost three times higher than in PEGASUS (30), as several innovations in HERCULES reached high maturity levels. The difference in the performance of the projects is in line with the judgement of the participants and the formal reviews of the projects by the funding agency: HERCULES and ORION were seen as particularly successful with respect to their innovation performance. Of course, when comparing projects, measuring the innovation performance by the scores is only justified if the projects are of a similar nature so that the difficulties of the development and the required maturity of the innovations are comparable.

Table 5

Innovation performance score and number of innovations generated by the five projects under study.

Project	HERCULES	PEGASUS	GEMINI	PERSEUS	ORION
Innovations in progress, number	7	7	9	10	21
Innovation performance score	87	30	54	56	166

6. Discussion

We have collected rare and rich field data and used it in an exploratory study to contribute to the advancement of innovation, R&D assessment and project studies. The study improves the understanding of the progress of diversified innovations, contributes to a better understanding of the innovation process (Birchall et al., 2011), is of benefit for innovation research (Messeni Petruzzelli et al., 2018; Collinson and Liu, 2019), and contributes to reducing uncertainty in policymaking (Janger et al., 2017). The study provides several theoretical contributions.

The first and main theoretical contribution of this paper is to the literature on the R&D assessment: the paper enriches it, by providing the new way on how to assess the progress of innovations at the development and validation stage, which until now was not paid sufficient attention in the literature (Aiello et al., 2020). The established indicators are insufficient (Dziallas and Blind, 2019; Schwartz et al., 2011; Enzing et al., 2015) and the objective assessment of innovations at this stage is challenging (e.g. Henttonen et al., 2016; Birchall et al., 2011). This holds not only for collaborative multi-actor projects but also for single actor projects. The innovations are often assessed via patent indicators, which is not enough, as pointed out by e.g. (Hung, 2017; Janger et al., 2017), and sometimes of little relevance, or in terms of their technical characteristics (e.g. Bican and Brem, 2020; Lazzarotti et al., 2011; Chiesa et al., 2009; Chiesa and Masella, 1996), but this is technology-specific. We offer a novel approach of how to assess the advancement of innovations directly and objectively, in generalisable way, and propose to measure the progress of the maturity of technological innovations, i.e.

⁴ Reporting was done not per innovation but per activity.

⁵ The available data mostly is on advancing the innovations by several levels, corresponding to summing up the respective scores. Nonetheless, as this is a linear transformation, the trend of efforts over scores should be linear.

the progress between the initial maturity level and the end level reached in the project. We thus introduced a baseline, the need of which has been highlighted by [Perkmann et al. \(2011\)](#). With this approach, the issues of the specificity of innovations ([Bican and Brem, 2020](#)) or of indirect assessments of the advancement of innovations (e.g. [Bendoly, 2014](#); [Hagedoorn et al., 2000](#)) are avoided, as the construct ‘advancement of maturity’ is not domain specific and enables to assess the progress of a variety of innovations. The approach is applicable to both single actor privately funded projects and inter-organisational multi-actor projects that develop multiple innovations.

The second theoretical contribution is to the literature on technology maturity. Scholars repeatedly noted that the conventional TRL scale ([Mankins, 2009a](#)) leaves room for interpretations, often leading to subjective TRL assessments ([Tan et al., 2011](#); [Peters et al., 2017](#)), provides rather coarse levels of technological maturity ([Olechowski et al., 2020](#); [Enzing et al., 2015](#)), and is debated controversially beyond the defense and space sectors ([Héder, 2017](#)). Our contribution shows that the main principles of the construction of the TRL scale are pertinent in the type of engineering projects investigated here, but the scale should be adapted. To appreciate the progress of the innovations in a nuanced, more accurate and unambiguous way, we propose refined TRL categories which are adapted to incremental innovation activities ([Martin, 2016a](#)), notably to innovations in engineering and IT which are embedded into larger pre-existing systems; the need for a refinement especially on higher TRL levels was also found by [Olechowski et al. \(2020\)](#) in a cross-sectorial study. In addition, scholars repeatedly noted that the advancement of the maturation of innovations is nonlinear (e.g. [Tan et al., 2011](#); [Peters et al., 2017](#)). To assess and compare the nonlinear progress in terms of advancement of maturity of the innovations, we introduced a scoring system which reflects the relative increase of the efforts needed to progress from one level to the next without contextual, non-technical influences ([Webster and Gardner, 2019](#); [Klessova et al., 2020](#)). This also enriches the literature on technology maturity. From the statements of the project actors and from the data on efforts spent, we found that while the absolute amount of resources spent to advance from one maturity level to the next significantly differed for different innovations, the *relative increase* of the efforts when advancing to higher (refined) TRL levels was similar. This happens because the increase of the efforts is related to similar changes of the implementation environment and similar additional requirements, as pointed out by [Olechowski et al. \(2020\)](#), [Héder \(2017\)](#), [Webster and Gardner \(2019\)](#), [Cooper and Kleinschmidt \(2007\)](#), [Nakamura et al. \(2012\)](#) and [Klessova et al. \(2020\)](#). The proposed values of scores hold for the domain of innovations that are embedded into larger systems and do not require an extremely high level of reliability. The increments do not increase as sharply as discussed in the TRL literature for aircraft and space applications, because of the lower requirements for reliability at high TRLs in the domains considered here. The construction logic of the scores is general and domain-independent, but their values, i.e. the increments between levels, may have to be adapted based upon empirical studies to other domains where the efforts and thus the scores may increase faster or slower towards higher levels of maturity ([Héder, 2017](#)).

According to [Mankins \(2009b\)](#) and [Olechowski et al. \(2020\)](#), the TRL scale alone does not provide a suitable tool to measure the technological risk for further advancements. While, indicated by the scores, the efforts increase when advancing to higher TRL, [Mankins \(2009b\)](#) postulates a decreasing risk towards higher TRL, due to the decreased technical difficulties. However, the anticipated higher effort may lead to a re-evaluation of the ratio of the expected benefit to the effort, and the risk that an innovation is abandoned was observed to increase at the

transition from TRL 6 to 7 because of the required efforts on higher TRL levels.

These two contributions, taken together, lead to the additional theoretical contribution to the literature on assessment of the innovation performance in complex multi-actor RDI projects, complementing existing views and perspectives (e.g. [EC, 2021b](#); [Spanos and Vonortas, 2012](#); [Pandza et al., 2011](#)). Scholars repeatedly pointed out (e.g. [Janger et al., 2017](#); [Hung, 2017](#)) that the innovation performance in such projects is multi-faceted and is not adequately reflected by the number of publications and patents, and that the project assessments shall be completed by objective measures ([Perkmann et al., 2011](#)). This became especially important in the recent years, as public agencies are massively supporting market-upstream RDI multi-actor projects with research and industrial organisations that are supposed to generate both new knowledge and innovations, and the innovations under development are diverse and specific ([Kostopoulos et al., 2019](#)). The paper enriches this literature by contributing to a better understanding of the innovation performance in such complex projects: we have shown that the number of innovations alone is not a reliable indicator to assess the project results on the innovation side: the project value creation ([Laursen and Svejvig, 2016](#); [Martinsuo et al., 2013, 2019](#)) is captured more reliably by the overall progress in the maturity of the innovations. When the number of organisational partners in the projects increases ([Mannak et al., 2019](#)), it can be expected that the number, variety and progress of different project innovations increase too. As the assessment is done at the level of each innovation, the proposed approach helps to overcome previously expressed concerns about the use of the TRL scale in multi-actor projects ([Enzing et al., 2015](#); [Héder, 2017](#)).

Our proposed scoring system adds scores for the advancement of each innovation within a project. When scores are compared across projects and even within a project, the innovations must be of comparable ‘weight’ in terms of difficulty and required effort. This is related to the definition of a component in the definition of the TRL scale ([Olechowski et al., 2020](#)) and of modularity in innovation (e.g. [Baldwin and Clark, 1997](#)).

With this kept in mind, the proposed approach helps to analyse and compare the progress between different innovations and projects, and it can be applied to projects with few or with many innovations: projects aiming at developing a small number of innovations with a high progress in their maturity achieve high scores when their innovation performance is assessed. This is in line with the goals of many funding agencies, in particular the European Commission, to push for projects that achieve innovations, in order to bridge the ‘valley of death’ between R&D and innovation in Europe. The identification and understanding of the nature of inefficiencies (low innovation performance score comparing with what can be expected) may help to point out directions for improvements ([Perkmann et al., 2011](#)); applying the findings of the study to publicly-funded projects can support policymaking practices and policy learning, and is important for designing policy interventions ([Dziallas and Blind, 2019](#); [Gault, 2018](#)). Last but not least, a better understanding of the innovation performance of the projects is also an enabler for answering different calls for research where innovation performance of market-upstream projects is part of the analysis.

7. Implications

7.1. Policy implications

The study has several policy implications. We organise them in three groups: implications on the assessment of multi-actor RDI projects, implications on the monitoring of running projects, and implications on

future funding policies.

Regarding the assessment of the projects, policymakers shall clarify what is taken as evidence of their success. The study suggests that at medium and high levels of maturity of innovations, the assessment should be based upon the *progress* of the innovations that was achieved during the project. We showed that the assessment of the success of the projects is different when considering the number of innovations or the advancement of their maturity. Therefore the funding agencies should define not only the target level of maturity, as it is often the case nowadays, but also the expected progress of maturation, i.e. the distance between the initial level and the final level. The progress of the innovations is an adequate indicator of the performance of the projects, which is not simply proportional to the number of innovations; increases of the maturity at higher levels require larger efforts. That is why the target level of maturity in the funding projects should be precisely defined, and the funding level must be consistent with the expected increase of efforts, i.e. the scores.

As a consequence, in our view policymakers should consider shifting the focus of attention of the assessments during and at the end of the projects to what is or will be achieved at the end of the project, in terms of the advancement of innovations compared to the relevant baseline, instead of comparing outputs to the initial plans, i.e. whether what was proposed initially was realised several years later as proposed. An assessment of the advancement of the degree of maturity reached by the innovations would help not only to better appreciate the innovation performance of the projects, but also stimulate targeting innovations with a high promise for progress up to rollout and prototype products rather than “sticking to the plan”. This means that the funding agency should accept that different innovations than originally planned may result from the project and some of the planned ones will have to be frozen at certain degrees of maturity. In the same vein, more accurate capturing of the innovation performance of the project with multiple innovations, especially at the medium and high level of maturity, may be facilitated with result-oriented reporting, i.e. reporting by innovations. This would lead to more transparency and better accountability, and a better analysis. The authors of the study spent significant efforts on reconstructing effort data per innovation: reporting was always done per workpackage (a workpackage “packages” work by topics), but the maturation of the innovations was often related to several workpackages.

For project monitoring and eventual interventions by the funding agencies and their evaluators, the refined TRL scale provides a more transparent classification of the achieved or targeted maturity levels than the standard TRL scale, and the scores can be used to compare the efforts spent with what can be expected in a certain domain. This alone does not define the necessary interventions, as there can be specific technological and non-technological factors that lead to deviations from the average efforts, but it provides a starting point for an in-depth discussion and further analysis, focusing on those innovations where unusual evolutions were observed.

Regarding implications for future funding policies, the scoring system, together with the refined TRL scale, enables a retrospective and fact-based assessment of the progress of maturation of innovations. As such, it provides information about which innovations, consortia, projects ... were successful with respect to this metrics. Considered *in combination* with additional data and analysis, the score provides additional ground for further analysis, policy learning and related policy decisions. For instance, at the level of the projects, it provides a basis for an analysis of why and how some innovations were pushed forward and some not. At the level of the domains, if adequate scores are available

across projects and sectors about the individual innovations, they invite to the analysis whether advancements in some domains are more difficult than in others, why, and what the policy decision shall be for a desired outcome. This can be used to decide on the relative need for funding of projects. Of course, future funding decisions shall be taken in combination with other factors: not only the expectation of effective advancements of innovations per se, but also the economic, societal, environmental and other impacts of the innovations. Moreover, the policy of the funding agency, whether to support a limited number of innovations to make a large step in maturity or to support a broader range of innovations with incremental advancement and a long term implementation horizon, whether to accept higher or lower levels of uncertainty and risks of abandoning, etc., will also influence the definition of the funding strategies. These are outside of the scope of this paper, but the proposed refined assessment scheme helps to combine these aspects with a facts-based analysis of the success of past projects, and the scores provide an indication of the relative increase in the required efforts for maturation up to higher maturity levels.

The same overall innovation performance score can be reached by projects by aiming at a large progress in maturation of a small number of innovations, and by aiming at smaller incremental advancements of maturation of a larger number of innovations. So there is a need to balance the expectations about the number of innovations within the projects and their desired maturation. Having a larger set of projects with advancement at lower TRLs (and thus lower budgets) and then, a few years later, selecting a smaller number of promising and well progressed innovations to continue with additional budget at higher TRLs might be worth considering. The proposed scores are well-suited to compare projects within a certain domain. The main advantage of the proposed scoring system is that the total (achieved) score vs. the total effort (funding) can be used to compare projects with maturation steps at *different TRL levels* within the same domain. However it is beyond the scope of the scores to take the (societal) relevance of the achieved progress or other impact-related criteria into account.

If the funding institutions target higher maturity levels of innovations at the end of the projects, abandoning of developments must be considered as a possible outcome that occurs for the reasons discussed in the previous sections (Klessova et al., 2020), and thus be acknowledged as part of the process, and mechanisms to deal with this risk must be implemented. Putting it more generally: what is required in publicly funded projects to advance the innovations as far as possible, and to minimise the risk of abandoning of developments, within and after the project?

7.2. Managerial contributions

The study also has several implications for managerial practice. These implications can be divided into three categories: Assessment of the projects, definition of targets and efforts, and interventions.

The proposed refinement of the TRL scale and the scores enables project managers to measure systematically and precisely the progress of the innovations, and thus to assess the value for money for the company or for the taxpayer. The innovation performance scores are proportional to average efforts. With regard to the related incurred project costs, the method enables assessing project efficiency (total score over budget): the scores reflect the effort, so projects in the same domain with different efficiency need different amount of funds to reach the same point. The efficiency does not depend on the targeted TRL level, as the scores account for the additional efforts that are required at the higher stages. The start and end points of the developments, the durations of the

development, and the budget are usually available as objective data in all settings, public or private.

In terms of the definition of targets and the planning of efforts, the refined TRL scale helps to define more precisely which intermediate and final maturity levels a project should achieve at which point in time, and the scores provide a guideline for the assignment of resources. Certain refined TRLs can serve to define gates in the development process where decisions are taken about the continuation of the maturation, and the scores also help to indicate how much remaining effort can be expected in the average case after a certain maturity level has been reached. This can then be compared to the expected business value and revisions of the planning can be made early if needed.

Thirdly, when a project is ongoing, the project managers can check the project progress in a more objective way, helping to identify obstacles and problems early, and finding ways to address them. The project management can set internal goals in terms of funding and related advancements, using the refined maturity scale, analyse the progress at certain points in time, and perform corrective actions if needed. Difficulties can be identified early and addressed, they can result from technology specific factors, or e.g. contextual conditions. This helps to answer several questions, for example: (1) For every Euro or Dollar allocated to the project, how much progress on the 'innovation side' of the project are we getting, and why? (2) How does that compare to other comparable projects? (3) How could the progress be improved, what are the blocking points where the progress is low? Factors influencing the efficiency of the development processes will lead to different ratios of total scores to funding or financial investment.

The proposed method is applicable in different contexts: it does not depend on whether the development was funded publicly or not, and whether it is collaborative or not. When considering different domains, the scores attached to the increments of the maturity may need to be adapted based on domain knowledge, but the general principle can be transferred. Thus, the results of the study provide actionable intelligence and push to describe the progress and the final destination of the development in a more precise way. The method can help to anticipate critical steps and to adapt management strategies and focus the attention on these. In the interviews, the project actors from industry repeatedly highlighted such needs:

"The steps on the way within one of these levels can be quite large to implement and quite expensive ... it would be good to distinguish this in order to more accurately describe where you are, where you want to be at a certain point ... do we have enough efforts done within this project, how expensive is it up to now and how it could be, this helps". ID53_HERCULES_large_ind

"My experience is with anything that you do, whether it's software development or whether it's reaching your own goal, you can typically assess the smaller steps with more certainty than you can assess a large step ... So, breaking large steps down into small steps is sensible from an uncertainty assessment point of view ... you check where you are and at which level, and whether it was planned like this, or something shall be corrected". ID50_ORION_large_ind

8. Conclusions

8.1. Summary of the theoretical contributions

We performed an exploratory multi-case study of 54 market-upstream innovations developed in 5 technological RDI projects with

67 organisations. Responding to multiple calls for relevant research, the study enhances the understanding of the assessment of the progress of innovations at their development and validation stage. We consider the advancement of innovations *directly and objectively*, from the point of view of their *progress in maturity*. With this approach, the issue of the specificity of innovations is avoided, as the construct 'advancement of maturity' is not domain or technology specific. To assess the start and the end point of developments in a nuanced and unambiguous manner, we refined the conventional TRL scale (Mankins, 2009a), especially at the higher TRL, and, exemplarily for the technical domain of the considered projects, developed a scoring system for the progression from of the initial to the final level of maturity. The study found that although the path taken, the obstacles and difficulties, and thus the *absolute* amounts of efforts for maturation were different for each innovation in the studied projects, the *relative* increase of the efforts for maturation was comparable for the different innovations. The study also contributes to a better understanding of the innovation performance of RDI projects with multiple innovations, as it proposes a way to measure the innovation performance of a project in a similar domain, by the cumulated scores for individual innovations.

8.2. Limitations

We reported the results of an exploratory qualitative study that included some quantitative elements. The proposals derived from the qualitative study (the refined TRL scale and the scoring method) reflect the domain of the projects that we considered. One of the objectives of this exploratory work is to engage the discussion around the topic and stimulate further research developments.

Exploratory studies have inherent limitations. The innovations considered here are limited to innovations developed within collaborative RDI projects funded by the European Union Research and Innovation Programmes in Enabling and Industrial Technologies, which have specificities such as a large number of autonomous organisations working together and having a strong application dimension. The projects from which the data was collected cover a range of topics and technologies, but all projects concerned engineering or software projects and aimed at incremental innovations that had to be integrated into larger technical and organizational systems. Nonetheless, the refined TRL scale in our view has a broad applicability. The 'labelling' of the refined maturity levels may have to be adapted to other domains, but the overall scheme is considered to be general, as the mechanisms of rollout and 'productisation', i.e. putting an innovation into continuous use are similar across many sectors of technology. We did not analyse projects funded by private firms or national funding programmes, but we expect the findings to be applicable here as well, as far as technological projects which are strongly applications-driven are concerned. The concrete scores illustrate the general approach but may have to be modified depending on the domain. E.g., the proposed scoring system reflects the situation in sectors of technology where safety and security are important aspects, but not as important as in aerospace or biomedical applications which require an extremely high level of reliability: in these sectors, the relative increase of the efforts between the TRLs is expected to be sharper. Due to the type of projects considered here, the scores are biased towards personnel efforts: developments that require a high level of material resources may need to be assessed using scores that grow faster with the maturity levels compared to those proposed.

We emphasize that we only consider the technological maturity of the innovations, not factors related to commercial readiness, which

depends on many non-technical factors. This is a different topic that is beyond the scope of this paper. We also did not include the issue of risk, which concerns several issues, as e.g. the risk of technical failure (the innovation does not perform as required) which is innovation-specific, the risk of insufficient return on investment and subsequently the termination of the development (the innovation is brought to the expected level of maturation, but the expected further effort is not in line with the benefits, or with the price that is paid for a new or improved product), or risks that result from competition. A contribution to this issue is that the scores provide an approximation of the additional efforts that can be expected for the further maturation of an innovation and can help to make the assessment of the expected benefits vs. total efforts in an objective fashion.

8.3. Future research directions

The limitations described above open several avenues for future research: theoretical and empirical studies are needed to better understand and assess the advancement of market-upstream innovations in various sectors. The need to assess the innovation performance of publicly funded projects will increase in the future, with the trend of the agencies to fund large RDI projects which reach high levels of maturity. Future research can use the findings of this study to investigate refinements of the TRL and adequate scores in other contexts, e.g. for other domains and different types of innovations (e.g. software with lower degrees of interaction with larger systems), or for innovations of different, more radical nature, with large investments needed on higher TRL scales, or in the context of projects aiming at obtaining patents, to better understand how to improve the advancement of innovations as well as the process behind the maturation of the innovations and the contextual conditions which have implications on the necessary efforts.

With regard to multi-actor RDI projects, better understanding is

needed on why the innovations developed in multi-actor RDI projects advance or do not advance after the project end. An important open question is to investigate how to quantify, anticipate and manage the risks in the process of the maturation of innovative technologies (Mankins, 2009b; Olechowski et al., 2020). Comparing the success of projects based on the scores and analysing the factors that led to superior or inferior performance can contribute to this analysis, by determining critical transitions on the maturity level (TRL) scale where the risk of failure is higher than on others and focusing management efforts and risk analysis on these.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

None.

Acknowledgments

We would like to thank the partners of the projects that were included as case studies who volunteered their time to contribute to this research. We are very grateful to the specialists who contributed to the validation of the proposed framework. A preliminary version of this paper was presented at the 81st Annual Meeting of the Academy of Management (AOM2021); comments and suggestions in conjunction with the presentation helped us to streamline the paper. The encouragements and the valuable and constructive comments of the two anonymous reviewers helped us to further improve the paper and to sharpen the conclusions.

Appendix A. Supplementary materials

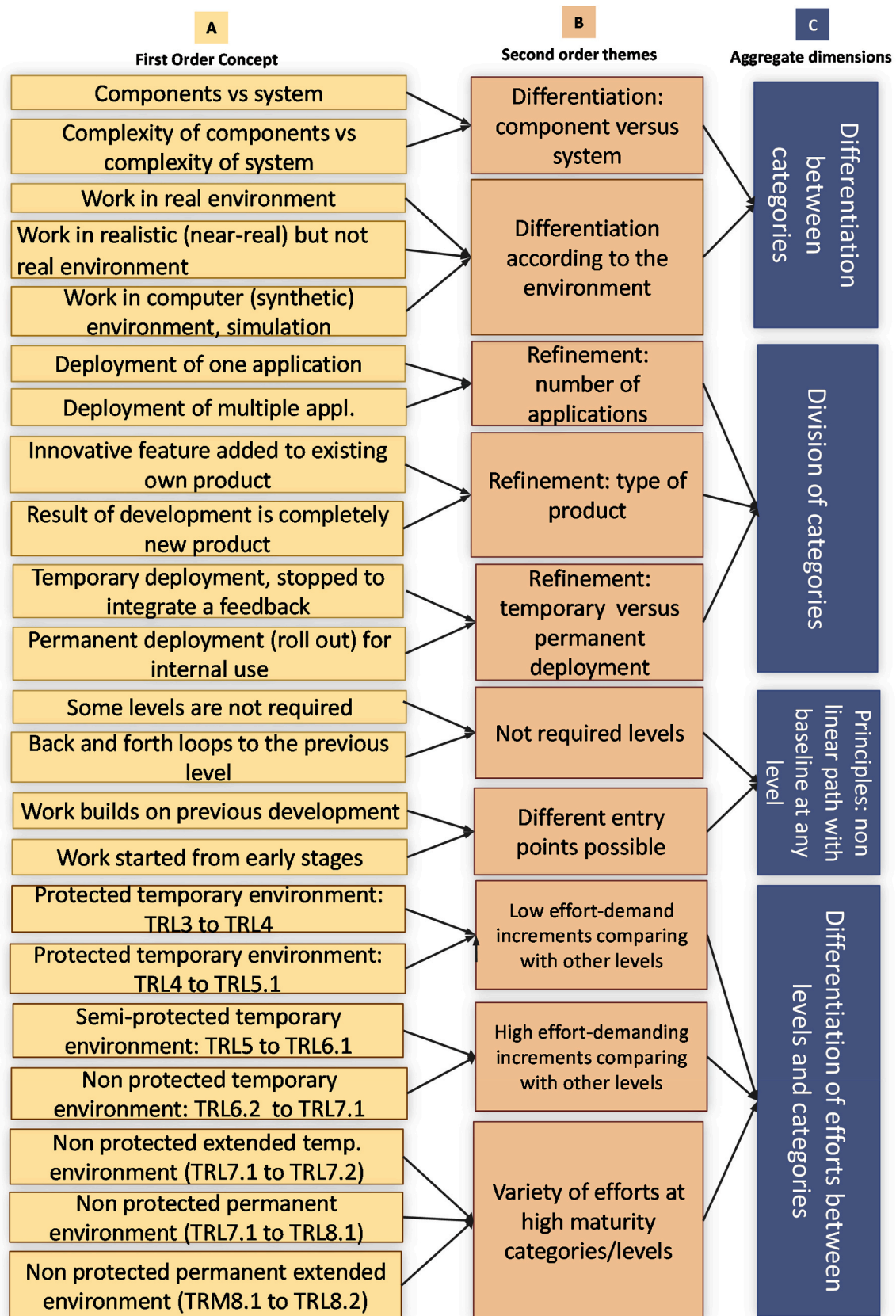
Supplementary materials to this article can be found online at <https://doi.org/10.1016/j.technovation.2022.102495>.

Appendices.

Appendix 1. Overview of the studied projects and maturity level of the innovations

	Project name	Thematic	EC Programme and funding scheme	Project duration months	Developed innovations total number	Their maturity level (end of the project, conventional TRL scale, see Table 1)			
						TRL5. Technology validated in relevant environment	TRL6. Technology demonstrated in relevant environment	TRL7. System prototype demonstrated in operational environment	TRL8: System completed and qualified
1	HERCULES	Energy efficiency, ICT	FP7, STREP	37	7	2	2	2	1
2	PEGASUS	System management, ICT	FP7, STREP	36	7	0	0	4	3
3	GEMINI	ICT	H2020, IA	34	9	0	4	5	0
4	PERSEUS	Sensor technology, ICT	H2020, RIA	36	10	0	0	7	3
5	ORION	Production, ICT	H2020, RIA	44	21	5	3	7	6

Appendix 2. Data structure overview



References

- Aguiar, L., Gagnepain, P., 2017. European cooperative R&D and firm performance: evidence based on funding differences in key actions. *Int. J. Ind. Organ.* 53 (C), 1–31.
- Aiello, F., Mannarino, L., Pupo, V., 2020. Innovation and productivity in family firms: evidence from a sample of European firms. *Econ. Innovat. N. Technol.* 29 (4), 394–416.
- Arranz, N., Arroyabe, M.F., Fernandez de Arroyabe, J.C., 2018. Network embeddedness in exploration and exploitation of joint R&D projects: a structural approach. *Br. J. Manag.* 31 (2), 421–437.
- Arroyabe, M., Arranz, N., Fernandez de Arroyabe, J.C., 2015. R&D partnerships: an exploratory approach to the role of structural variables in joint project performance. *Technol. Forecast. Soc. Change* 90B, 623–634.
- Baldwin, C., Clark, K., 1997. Managing in an age of modularity. *Harv. Bus. Rev.* 84–93. Sept–Oct 1997.
- Bendoly, E., 2014. System dynamics understanding in projects: information sharing, psychological safety, and performance effects. *Prod. Oper. Manag.* 23 (8), 1352–1369.
- Bican, P., Brem, A., 2020. Managing innovation performance: results from an industry-spanning explorative study on R&D key measures. *Creativ. Innovat. Manag.* 29 (2), 268–291.
- Birchall, D., Tovstiga, G., 2005. Capabilities for Strategic Advantage: Leading through Technological Innovation. Palgrave Macmillan, Basingstoke, ISBN 978-0-230-52249-7, p. 320.
- Birchall, D., Chanaron, J.-J., Tovstiga, G., Hillenbrand, C., 2011. Innovation performance measurement: current practices, issues and management challenges. *Int. J. Technol. Manag.* 56 (1), 1–20.
- Chiesa, V., Masella, C., 1996. Searching for an effective measure of R&D performance. *Manag. Decis.* 34 (7), 49–57.
- Chiesa, V., Frattini, F., Lazzarotti, V., Manzini, R., 2009. Performance measurement in R&D: exploring the interplay between measurement objectives, dimensions of performance and contextual factors. *R&D Management* 39 (5), 487–519.
- Collinson, S., Liu, Y., 2019. Recombination for innovation: performance outcomes from international partnerships in China. *R&D Management* 49 (1), 46–63.
- Cooper, R., 1990. Stage-gate systems: a new tool for managing new products. *Bus. Horiz.* 33 (3), 44–54.
- Cooper, R., 1999. From experience: the invisible success factors in product innovation. *J. Prod. Innovat. Manag.* 16 (2), 115–133.
- Cooper, R., 2007. Managing technology development projects. *Res. Technol. Manag.* 49 (6), 23–31.
- Cooper, R., 2017. *Winning at New Products: Creating Value through Innovations*, fifth ed. Perseus Publishing, Cambridge, MA, ISBN 978-0465093328, p. 448.
- Cooper, R., Kleinschmidt, E., 2007. Winning businesses in product development: the critical success factors. *Res. Technol. Manag.* 50 (3), 52–66.
- Cornford, S., Sarsfield, L., 2004. Quantitative methods for maturing and infusing advanced spacecraft technology. *IEEE Aerospace Conference Proc.* 1 <https://doi.org/10.1109/AERO.2004.1367652>.
- de Oliveira, J., Escrivão Filho, E., Nagano, M., Ferraud Sergio, A., Rosim, D., 2015. What do small business owner-managers do? A managerial work perspective. *J. Glob. Entrepr. Res.* 5 (19), 1–21.
- Dewangan, V., Godse, M., 2014. Towards a holistic enterprise innovation performance measurement system. *Technovation* 34 (9), 536–545.
- Dzialis, M., 2020. How to evaluate innovative ideas and concepts at the front-end? A front-end perspective of the automotive innovation process. *J. Bus. Res.* 110, 502–518.
- Dzialis, M., Blind, K., 2019. Innovation indicators throughout the innovation process: an extensive literature analysis. *Technovation* 80–81, 3–29.
- EC, 2014. Horizon 2020 – Work programme 2014–2015. General Annexes. Annex G. Technology Readiness Levels (TRL). Available at : https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-rl_en.pdf. (Accessed 6 December 2021).
- EC, 2015. Horizon 2020 Indicators. Assessing the Results and Impact of Horizon 2020, ISBN 978-92-79-49476-5. <https://doi.org/10.2777/71098>.
- EC, 2020. Horizon 2020 Programme ‘Leadership in enabling and industrial technologies’. Available at <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/leadership-enabling-and-industrial-technologies>. (Accessed 2 December 2021).
- EC, 2021a. Horizon 2020 Interactive Dashboard. Data retrieved from <https://webgate.ec.europa.eu/dashboard/hub/>. (Accessed 11 March 2022).
- EC, 2021b. Innovation Radar. <https://digital-strategy.ec.europa.eu/en/activities/innovation-radar>. (Accessed 6 December 2021).
- Edison, H., Bin Ali, N., Torkar, R., 2013. Towards innovation measurement in the software industry. *J. Syst. Software* 86 (5), 1390–1407.
- Eisenhardt, K.M., 1989. Building theories from case study research. *Acad. Manag. Rev.* 14 (4), 532–550.
- Eisenhardt, K.M., 1991. Better stories and better constructs: the case for rigor and comparative logic. *Acad. Manag. Rev.* 16 (3), 620–627.
- Eisenhardt, K.M., 2021. What is the Eisenhardt Method, really? *Strat. Organ.* 19 (1), 147–160.
- Eisenhardt, K.M., Graebner, M., 2007. Theory building from cases: opportunities and challenges. *Acad. Manag. J.* 50 (1), 25–32.
- Ellwood, P., Williams, C., Egan, J., 2020. Crossing the Valley of Death: Five Underlying Innovation Processes. *Technovation*, p. 102162.
- Enkel, E., Javier, P.-F., Gassmann, O., 2005. Minimizing market risks through customer integration in new product development: learning from bad practice. *Creativ. Innovat. Manag.* 14 (4), 425–437.
- Enzing, C., Mahieu, B., Poel, M., Potau, X., Beckert, B., Gotsch, M., Som, O., Thielmann, A., Reiss, T., 2015. Ex Post Evaluation and Impact Assessment of Funding in the FP7 NMP Thematic Area. Main Report. European Commission, ISBN 978-92-79-45413-4. <https://doi.org/10.2777/844400>.
- Fernandes, G., Pinto, E., Araújo, M., Magalhães, P., Machado, R., 2017. A method for measuring the success of collaborative university-industry R&D funded contracts. *Procedia Comput. Sci.* 121, 451–460.
- Garud, R., Berends, H., Tuertscher, P., 2017. Qualitative approaches for studying innovation as process. In: *The Routledge Companion to Qualitative Research in Organization Studies*, pp. 226–247. <https://doi.org/10.4324/9781315686103-15>.
- Gault, F., 2018. Defining and measuring innovation in all sectors of the economy. *Res. Pol.* 47 (3), 617–622.
- Gehman, J., Glaser, V., Eisenhardt, K., Gioia, D., Langley, A., Corley, K., 2018. Finding theory-method fit: a comparison of three qualitative approaches to theory building. *J. Manag. Inq.* 27 (3), 284–300.
- Gioia, D., Corley, K., Hamilton, A., 2013. Seeking qualitative rigor in inductive research. *Organ. Res. Methods* 16 (1), 15–31.
- Grimaldi, R., Von Tunzelmann, N., 2002. Assessing collaborative, pre-competitive R&D projects: the case of the UK LINK scheme. *R&D Management* 32 (2), 165–173.
- Hagedoorn, J., Link, A., Vonortas, N., 2000. Research partnerships. *Res. Pol.* 29 (4), 567–586.
- Héder, M., 2017. From NASA to EU: the evolution of the TRL scale in public sector innovation. *Innovat. J.: Publ. Sector Innovat. J.* 22 (2) article 3.
- Henttonen, K., Ojanen, V., Puimalainen, K., 2016. Searching for appropriate performance measures for innovation and development projects. *R&D Management* 46, 914–927.
- Hittmar, S., Varmusa, M., Lendela, V., 2015. Proposal of evaluation system for successful application of innovation strategy through a set of indicators. *Procedia Econ. Finance* 26, 17–22.
- Hung, C.-L., 2017. Social networks, technology ties, and gatekeeper functionality: implications for the performance management of R&D projects. *Res. Pol.* 46 (1), 305–315.
- Hurmelin-Laakkanen, P., Nätti, S., Pikkariainen, M., 2021. Orchestrating for lead user involvement in innovation networks. *Technovation* 108, 102326.
- Janger, J., Schubert, T., Andries, P., Rammer, C., Hoskens, M., 2017. The EU 2020 innovation indicator: a step forward in measuring innovation outputs and outcomes? *Res. Pol.* 46, 30–42.
- Kenley, R., El-Khoury, B., 2012. An analysis of TRL-based cost and schedule models. In: *Proceedings, 9th Annual Acquisition Research Symposium. Report NPS-AM-12-C9P21R02-086*. Available at: <https://calhoun.nps.edu/handle/10945/51339>. (Accessed 26 March 2021).
- Kim, E., Kim, S., Kim, H., 2017. Development of an evaluation framework for publicly funded R&D projects: the case of Korea’s Next Generation Network. *Eval. Progr. Plann.* 63, 18–28.
- Klessova, S., Thomas, C., Engell, S., 2020. Structuring inter-organizational R&D projects: Towards a better understanding of the project architecture as an interplay between activity coordination and knowledge integration. *Int. J. Proj. Manag.* 38 (5), 291–306. <https://doi.org/10.1016/j.ijproman.2020.06.008>.
- Kock, A., Heising, W., Gemünden, H., 2015. How ideation portfolio management influences front-end success. *J. Prod. Innovat. Manag.* 32 (4), 539–555.
- Kostopoulos, K., Spanos, Y., Soderquist, K., Prastacos, G., Vonortas, N., 2019. Market-, firm-, and project-level effects on the innovation impact of collaborative R&D projects. *J. Knowl. Econ.* 10 (4), 1384–1403.
- Laursen, M., Svejvig, P., 2016. Taking stock of project value creation: a structured literature review with future directions for research and practice. *Int. J. Proj. Manag.* 34 (4), 736–747.
- Lazzarotti, V., Manzini, R., Mari, L., 2011. A model for R&D performance measurement. *Int. J. Prod. Econ.* 34 (1), 212–223.
- Maietta, O.W., 2015. Determinants of university-firm R&D collaboration and its impact on innovation: a perspective from a low-tech industry. *Res. Pol.* 44 (7), 1341–1359.
- Mankins, J.C., 2009a. Technology readiness assessments: a retrospective. *Acta Astronaut.* 65 (9), 1216–1223. <https://doi.org/10.1016/j.actaastro.2009.03.058>.
- Mankins, J.C., 2009b. Technology readiness and risk assessments: a new approach. *Acta Astronaut.* 65 (9), 1208–1215.
- Mannak, R., Meeus, M., Raab, J., Smit, A., 2019. A temporal perspective on repeated ties across university-industry R&D consortia. *Res. Pol.* 48 (9), 103829.
- Markham, S., 2013. The impact of front-end innovation activities on product performance. *J. Prod. Innovat. Manag.* 30 (S1), 77–92.
- Markham, S., Lee, H., 2013. Product development and management association’s 2012 comparative performance assessment study. *J. Prod. Innovat. Manag.* 30, 408–429.
- Martin, B., 2016a. Twenty challenges for innovation studies. *Sci. Publ. Pol.* 43 (3), 432–450.
- Martin, B., 2016b. Introduction to discussion paper on ‘The sciences are different and the differences are important’. *Res. Pol.* 45, 1691.
- Martinsuo, M., Klakegg, O.-J., van Marrewijk, A., 2019. Delivering value in projects and project-based business. *Int. J. Proj. Manag.* 37 (5), 631–635.
- Martinsuo, M., Suomala, P., Kannianen, J., 2013. Evaluating the organizational impact of product development projects. *Int. J. Manag. Proj. Bus.* 6 (1), 173–198.
- Messeni Petruzzelli, A., Ardito, L., Savino, T., 2018. Maturity of knowledge inputs and innovation value: the moderating effect of firm age and size. *J. Bus. Res.* 86, 190–201.
- Mills, A., Durepos, G., Wiebe, E., 2010. *Encyclopedia of Case Study Research*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412957397>.
- Nakamura, H., Kajikawa, Y., Suzuki, S., 2012. Multi-level perspectives with technology readiness measures for aviation innovation. *Sustain. Sci.* 8 (1), 87–101.

- OECD, 2018. Oslo Manual: Guidelines for Collecting, Reporting and Using Data on Innovation, fourth ed. The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris/Eurostat, Luxembourg. <https://doi.org/10.1787/9789264304604-en>.
- Ojanen, V., Vuola, O., 2006. Coping with the multiple dimensions of R&D performance analysis. *Int. J. Technol. Manag.* 33, 2–3.
- Olechowski, A., Eppinger, S., Joglekar, N., Tomaschek, K., 2020. Technology readiness levels: shortcomings and improvement opportunities. *Syst. Eng.* 23, 395–408.
- Pandza, K., Wilkins, T., Alfoldi, E., 2011. Collaborative diversity in a nanotechnology innovation system: evidence from the EU Framework Programme. *Technovation* 31, 476–489.
- Perkmann, M., Neely, A., Walsh, K., 2011. How should firms evaluate success in university–industry alliances? A performance measurement system. *R&D Management* 41 (2), 202–216.
- Peters, W., Doskey, S., Moreland, J., 2017. Technology maturity assessments and confidence intervals. *Syst. Eng.* 20 (2), 188–204.
- Prater, J., Kirytopoulos, K., Ma, T., 2016. Optimism bias within the project management context: a systematic quantitative literature review. *Int. J. Manag. Proj. Bus.* 10 (2), 370–385.
- Ritala, P., Hurmelinna-Laukkanen, P., 2009. What's in it for me? Creating and appropriating value in innovation-related coopetition. *Technovation* 29, 819–828.
- Schwartz, L., Miller, R., Plummer, D., Fusfeld, A., 2011. Measuring the effectiveness of R&D. *Res. Technol. Manag.* 54 (5), 29–36.
- Schwartz, M., Peglow, F., Fritsch, M., Günther, J., 2012. What drives innovation output from subsidized R&D cooperation?—project-level evidence from Germany. *Technovation* 32, 358–369.
- Science/Business, 2021. Let the Party Begin: EU Commission Plans 'Symbolic Launch' for Horizon Europe. January 26, 2021. <https://sciencebusiness.net/frameworkprogrammes/news/let-party-begin-eu-commission-plans-symbolic-launch-horizon-europe>. (Accessed 16 November 2021).
- Soh, P.-H., Subramanian, A., 2014. When do firms benefit from university–industry R&D collaborations? The implications of firm R&D focus on scientific research and technological recombination. *J. Bus. Ventur.* 29 (6), 807–821.
- Spanos, Y., Vonortas, N., 2012. Scale and performance in publicly funded collaborative research and development. *R&D Management* 42 (5), 494–513.
- Strauss, A., Corbin, J., 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage, Thousand Oaks, CA.
- Tan, W., Ramirez-Marquez, J.-E., Sauser, B., 2011. A probabilistic approach to system maturity assessment. *Syst. Eng.* 14 (3), 279–293.
- Torrisi, S., Gambardella, A., Giuri, P., Harhoff, D., Hoisl, K., Mariani, M., 2016. Used, blocking and sleeping patents: empirical evidence from a large-scale inventor survey. *Res. Pol.* 45, 1374–1385.
- vom Brocke, J., Lippe, S., 2015. Managing collaborative research projects: a synthesis of project management literature and directives for future research. *Int. J. Proj. Manag.* 33, 1022–1039.
- von Hippel, E., 1986. Lead users: a source of novel product concepts. *Manag. Sci.* 32, 791–805.
- von Hippel, E., 1994. "Sticky Information" and the locus of problem solving: implications for innovation. *Manag. Sci.* 40, 429–439.
- Webster, A., Gardner, J., 2019. Aligning technology and institutional readiness: the adoption of innovation. *Technol. Anal. Strat. Manag.* 31 (10), 1229–1241.
- Yin, R., 2017. *Case Study Research and Applications: Design and Methods*, sixth ed. SAGE Publications, Inc, p. 352.