INTERNATIONAL DESIGN CONFERENCE – DESIGN 2024
https://doi.org/10.1017/pds.2024.12

Operationalizing community-based open scientific design research benchmarks: application to model-based architecture design synthesis

Romain Pinquié [1,✉], Lionel Roucoules [2], Pierre-Alain Yvars [3] and Raphaël Chenouard [4]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, France, [2] Arts et Métiers ParisTech, France,
[3] ISAE-Supméca, France, [4] École Centrale de Nantes, France

✉ romain.pinquie@grenoble-inp.fr

Abstract

The point has repeatedly been made that validation is a crucial success factor in demonstrating the scientific contribution and ensuring the adoption of results. Still, researchers in design science validate their research findings too infrequently. We must all evaluate our claimed contributions on open benchmarks to improve validation quality and foster cumulative research. In this paper, we propose a meta-model to standardise and operationalise the concept of open scientific benchmarks in design science and to guide communities of researchers in the co-development of scientific benchmarks.

Keywords: *open science, design research, benchmarking, model-based systems engineering (MBSE), product architecture*

1. Introduction

1.1. Context: Where is the trust in engineering design research?

As seen through the media, it is hard to escape the conclusion that there is a growing mistrust of science among the public. The climate and COVID crises show that the acceptance of science and technology is on the wane. The distrust of science and technology affects not only large sections of the population but also propagates through the scientific and industrial communities for many reasons, including the lack of reproducibility - i.e., obtaining consistent results using the same input data, data collection method, and data analysis method - and replicability - i.e., getting consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data, data collection method and data analysis method -, failure to report failures, exaggeration and cherry-picking of results, etc. Although validation is one crucial success factor in demonstrating the scientific contribution and ensuring the adoption of results by companies, "*there is this concern that design research does not live up to the standards of science: it is creating in a sense too many theories and models, which jeopardises the coherence of the discipline and which indicates that design research does not yet have the means to test and refute design theories and models*" (Vermaas, 2014).

1.2. Problem: Lack of validation and cumulative research in engineering design

The point has repeatedly been made that researchers in design science validate their research findings too infrequently (Tomiyama, 2020; Isaksson et al., 2020; Cash, 2018; Barth et al., 2011; Eckert et al., 2003). Moreover, the lack of research data validation and openness hampers cumulative research. Indeed, we do not observe any means to operationalise cumulative science within the design research communities. The rapid disclosure of research results in design research papers does not support cumulative research as most papers are of poor quality compared to the quality criteria for reproducible and replicable research. What can we do to improve this situation?

Although no research method or empirical evaluation is perfect, it is far superior to merely asserting that a design theory, process, method or technology is valuable. In a nutshell, we are much better at proposing new results than at validating them. We can blame the lack of open data requirements and evaluation from government funding agencies and grant-givers, the intense pressure to publish academic - often incomplete - work, the use of bibliometric parameters to measure successful careerism, the competitive research grant programs and many other factors compromising the quality of research validation in engineering design. Nevertheless, in the end, it is the author's responsibility to make quantitative judgements with fair comparisons to the state of the art and be able to communicate their research methods and results in a very convincing and open way. We may argue that the responsibility is increasingly being shifted to the individual. Researchers have to be good, but it must also be tackled by research institutions such as government funding agencies, grant-givers, policymakers and communities of academics and industry practitioners, such as the Design Society, INCOSE, CIRP, etc.

The value of benchmarks has been thoroughly discussed (Sim et al., 2003). Using benchmarks results in a more rigorous examination of contributions. Benchmarking promotes collaborative open research, and replication is built into the method. Appropriately deployed, benchmarking is not about winning a contest but surveying a landscape. The presence of benchmarks in design research would state that contributions ought to be evaluated against clearly defined standards. The benchmarking process would require us to examine our understanding of design research. Still, excellent communication and collaboration will lead to a more substantial consensus on the fundamental problems, research goals and methods and encapsulate this knowledge in an evaluation. However, the lack of a standardised and operational definition of scientific benchmarks in design research stands as a research gap.

1.3. Statement of contribution

Two years ago, researchers belonging to the French national special interest group S.mart co-specified and co-developed a web-based version-control and collaborative platform (https://github.com/GIS-S-mart/Welcome) where communities of researchers can engage in the building of a sustainable evidence-based validation ecosystem (Pinquié et al., 2022). This collaborative space intends to enable a community of researchers to engage in an asynchronous collaboration for:

1. the co-definition of fundamental and practical research problems, goals, solutions, metrics, and
2. the fair and systematic evaluation of competing or complementary research proposals on open benchmarks with agreed-upon measurement protocols, metrics and benchmark exercises.

The contribution presented in this paper is twofold. First, a meta-model helps formally define, standardise, and operationalise the concept of scientific benchmarks to guide communities of researchers in design science in the sustainable co-development of new benchmarks. Second, an instance of the meta-model applied to the activity of concept finding for the model-based architecture design synthesis (MBADS) of engineered systems illustrates how to operationalise the meta-model.

1.4. Outline

In section 2, the paper begins by briefly reviewing the literature and discussing the lack of validation in design research with the activity of concept finding in an MBADS approach as an illustrative example. Section 3 will contribute to the improvement of validation in design research by proposing a meta-model, guidelines and a case study for operationalising the co-development of community-based open design benchmarks. Section 4 concludes the paper and suggests future works.

2. Literature review

2.1. Lack of validation in engineering design

The validation ("Does the method solve the problem?") and evaluation ("How well a method works and under which circumstances") of results is a scorching topic in design research. Indeed, although

https://doi.org/10.1017/pds.2024.12 Published online by Cambridge University Press

researchers developed guidelines that help validate design research (Isaksson et al., 2020; Hein and Lamé, 2020; Blessing and Chakrabarti, 2009; Carolyn C. Seepersad et al., 2006; Frey and Dym, 2006), design research often recommends loose sets of recommendations rather than reproducible and replicable procedures. Various reasons explain this lack of focus on validation (Gericke et al., 2022).

Moreover, design research is interdisciplinary, and using multiple research methods is difficult (Eckert et al., 2003). Literature reviews drew up extensive lists of research methods (Barth et al., 2011; Escudero-Mancebo et al., 2023) and design research objectives (Eckert et al., 2004; Cantamessa, 2003). When mixing research methods from multiple research areas, many challenges can arise due to the individual research cultures of each discipline involved.

Publishing or making our data available to others is not considered standard practice. As measured by the French Open Science Barometer, researchers in design science keep their results more confidential than other disciplines, with only 10 % of French publications in engineering which mention the sharing of their data and 15% that include a "Data Availability Statement" between 2013 and 2021, whereas 86 % of French publications in engineering mention the use of data[1]. This lack of openness is all the more regrettable, given that the opening up of data forces researchers to guarantee data quality. We may assume that it is mainly because we are primarily focused on getting grant money, and the influence of outside sponsors, such as industrialists, limits the openness of research data. Still, it is necessary to open our data to a scientific community that examines the same research question from multiple angles over time because one data collection effort does not lead to a definitive answer. Research methods and results should be well documented, with enough detail so that other teams can attempt to reproduce or replicate the findings and expand upon them. If they come up with the same general results over time, all of these efforts give evidence for the scientific truth of the findings.

We can also blame the competitive calls for research proposals and industrial situations that push us to make claims of contribution, which are often too broad and too complex to provide objective evidence justifying the claims (Isaksson et al., 2020) before the research project runs out of time and money (Blessing and Chakrabarti, 2009; Gericke et al., 2003).

Finally, among the objectives perused by researchers in design science, the development of new methods ranks at the top, and their validation is still a significant challenge. For instance, the influence of background knowledge (Lukas Paehler et al., 2023) hampers the evaluation of new design methods as it is practically impossible for the creators of the method to try out the method in numerous contexts as they need to be trained in the domains of application (Gericke et al., 2020). Conversely, when designers want to use a new design method, designers in the domain of application need to be trained on the method (Frey and Dym, 2006) before interpreting and adopting it, often in an organisational context that requires a collective adoption (Hein and Lamé, 2020). A recent study of the current practice to validate design methods in design research indicates a lack of standard metrics (Eisenmann et al., 2021).

Researchers can evaluate their proposals on open benchmarks to provide evidence of the claimed contributions. Scientific benchmarks advance a discipline by improving the science and increasing the cohesiveness of the community. In computer science, benchmarks serve as ground truth for a transparent evaluation using standard datasets, metrics and benchmarking protocols (Hirsch and Hofer, 2022; Thiyagalingam et al., 2022). The results of several evaluations can sometimes lead to a central collection of solutions that help to know what works, for whom, and under which circumstances and select solutions with ease. A benchmark emerges through a synergetic process of technical knowledge and social consensus proceeding in tandem. As a technical and social artefact, a scientific benchmark has a strong positive effect on research as it increases concern in the validation of the results, especially when used by a research community that is sufficiently well-established and has a culture of collaboration (Sim et al., 2003). Examples of benchmark exercises in engineering design include the landing gear system (Boniol et al., 2017), the thermostat and flight control system (Lempia and Miller, 2009) and the wheel brake system[2]. The welcome page of the open and community-based benchmarking platform[3] lists a collection of open benchmark exercises. However, most exercises are not characterised according to research goals. For instance, the landing gear case study cannot directly be used to benchmark candidate methods and

[1] https://barometredelascienceouverte.esr.gouv.fr/donnees-de-la-recherche/disciplines?id=disciplines.partage

[2] https://es-static.fbk.eu/projects/air6110/

[3] https://github.com/GIS-S-mart/Welcome (not all benchmarks implement the proposed meta-model yet)

tools that aim at improving the early verification of engineered systems designs because there are no metrics to compare competing solutions, no testing procedure, and it only deals with discrete behaviour that does not consider continuous physics-based hydraulic and electric power flows. We may think that the validation of engineering design research is much more complicated than the validation of computer science research, but arguments such as " *we would need several years to get evidence*", "*our design situation is really special*", "*we cannot replicate the simulation as we do not have the computing resources*" are no different from other disciplines. For instance, climate research, a discipline that requires more computing resources to solve research questions on highly complex situations and over a very long period, created the Coupled Model Intercomparison Project 6 (CMIP6), which has been internationally adopted as a shared infrastructure that serves as the provision of benchmarks against which to compare improvements in models and prediction quality. The same when researchers argue that "*the influence of designers' knowledge is too important to get comparable results*"; disciplines like Human-Computer Interaction are facing the same challenges but adopt the scientific method.

2.2. What is model-based architecture design synthesis?

Design synthesis (Kryssanov et al., 1999) is a problem-solving activity that involves exploring a design space to find solutions that satisfy a set of needs. This paper focuses on computational or model-based design synthesis, which is the algorithmic implementation and execution of design creation on a computer to leverage computational speed and depth of calculation to reduce the tedium for human designers and augment the process of searching the space of alternatives for a preferred solution (Cagan et al., 2005). However, this paper goes beyond existing model-based design synthesis definitions limited to the computation of design solutions (Rosich et al., 2016). Indeed, in this paper, MBADS is a design process which relies on computer-based automatic problem-solving techniques to generate architecture design solutions that are correct by construction. Examples include (Yvars and Zimmer, 2022, 2021; Ribeiro Dos Santos et al., 2019; Menu et al., 2018; Hartmann et al., 2018; Menu and Nicolai, 2017; Rosich et al., 2016; Nuzzo et al., 2014; Broodney et al., 2012). Based on this definition, MBADS is a 4-activity design process that is relatively similar to the one proposed in the contract-based design synthesis CONDEnSe (Ribeiro Dos Santos et al., 2019):

1. **Requirements development** aims to define a problem space based on functional requirements and design constraints. A functional requirement is an agreed-upon expectation for a system to perform a function at a specified level of performance under some conditions. A set of requirements is sometimes over-constrained by design constraints, which are limitations on the design or implementation of a system externally imposed by a stakeholder or external system. Constraints cannot be traded off and make certain designs “not allowed” whereas requirements make certain designs inappropriate for their intended use. The problem space includes all design solutions acceptable concerning the set of requirements and constraints, but all acceptable design solutions are not feasible for technical, economical, […], or legal reasons.
2. **Concepts finding** is an activity that intends to find incomplete sets of design variables, parameters and constants defining the set of acceptable and feasible conceptual under-defined design solutions. A design variable is an unknown variable whose value is computed at step 4 problem solving. A design parameter is a known design variable with a fixed value set after a design synthesis iteration concluded by an agreed-upon design decision; therefore, it is not calculated at step 4 problem solving. A constant has an immutable value which remains the same between two computations. Concept finding also partially defines the equations between the design variables, parameters and constants.
3. **Problem modelling** is the formal definition of a computational design problem that can be solved with computer-based automatic problem-solving techniques. To get a complete computable design problem, it is often necessary to complete the identification of the set of design variables, parameters and constants and their relationships.
4. **Problem solving** consists in automatically solving the computational design problem by providing solutions that are correct by construction with respect to the set of requirements. A solution is a set of values for the design variables that characterise the selected concepts.

https://doi.org/10.1017/pds.2024.12 Published online by Cambridge University Press

The problem-solving objective pursued by an MBADS process is not unique and often not well specified in existing contributions. So far, we have identified six MBADS objectives adapted from an existing taxonomy (Yvars and Zimmer, 2021). Still, a more systematic review of the literature and industrial tasks related to system architecting and the validation with a systematic case study is necessary to justify this taxonomy rigorously. The six objectives include MBADS for…

1. **Architecture generation**: generate system elements, which can be subsystem functions, abstract logical units or concrete hardware/software implementations, and their interfaces satisfying a set of requirements. See an example in (Hartmann et al., 2018).
2. **System configuration**: generate system architectures that are variants of a reference meta-architecture based on cardinalities and compatibility relationships between interchangeable system elements (i.e., subsystems or components). See an example in (Menu and Nicolai, 2017).
3. **Resource allocation**: allocate hardware and/or software system elements to system functional requirements. See an example in (Yvars and Zimmer, 2022).
4. **Space allocation**: space allocation defines the bounding box of the subsystems and their spatial position and orientation within a system. See an example in (Rosich et al., 2016)
5. **Interfaces routing:** define the path and wrap for each physical interface between two subsystems within the system bounding box. See an example in (Formentini et al., 2022).
6. **System sizing:** to determine the value of unknown design variables of a system element. See an example in (Yvars et al., 2009).

An MBADS process usually combines several objectives and can be applied successively with the results of a first iteration (e.g., architecture generation) that serves as input for a second iteration pursuing a new objective (e.g., system configuration). Moreover, the inputs and outputs of each activity depend on the objective of the MBADS process. For example, the output of the activity "concepts finding" is different if the objective is "system configuration" or "system sizing". The former provides architecture alternatives, whereas the latter gives a single configured architecture. To illustrate our definition of an open scientific benchmark, this paper evaluates competing and/or complementary claimed contributions related to the MBADS process for system sizing.

2.3. Model-based architecture design synthesis for system sizing

MBADS for system sizing is a 4-activity process starting from developing a set of requirements leading to a first problem space within which one can search candidate concepts of solution. The preferred concepts are then formally defined as a computable design problem for which automatic solving techniques can be applied to find a list of solutions that are correct by construction.

Table 1. Activities of a model-based architecture design synthesis process for system sizing

Requirements development	
Inputs:	- Set of expected system functions. - A system architecture with a given configuration.
Outputs:	- Set of functional requirements. - Set of design constraints.
Concepts finding	
Inputs:	Outputs of the activity "requirements engineering".
Outputs:	- Set of typed design variables with their range of acceptable values. - Set of design parameters, each with a fixed value. - Set of constants, each with an unmodifiable value. - Preliminary set of relationships between design variables, design parameters, and constants.
Problem modelling	
Inputs:	Outputs of the activity "concepts finding".
Outputs:	- Complete set of design variables, design parameters, and constants. - Complete set of relationships between design variables, design parameters, and constants.
Problem solving	
Inputs:	Outputs of the activity "problem modelling".
Outputs:	At least one value for each design variable.

https://doi.org/10.1017/pds.2024.12 Published online by Cambridge University Press

2.4. Claimed contributions of design synthesis, focus on concepts finding

In our scope, which is limited to the process of design synthesis for system sizing, we will focus on the activity of concept finding as defined in Table 1. Several design methods supported by a toolset claim to support concept finding (Albert et al., 2009; Pailhès et al., 2011; Suh, 2001). They were not explicitly developed for a model-based approach but are broad and claim to be generic enough to support engineers willing to solve any architecture design problem.

3. An illustrated definition of an open design research benchmark

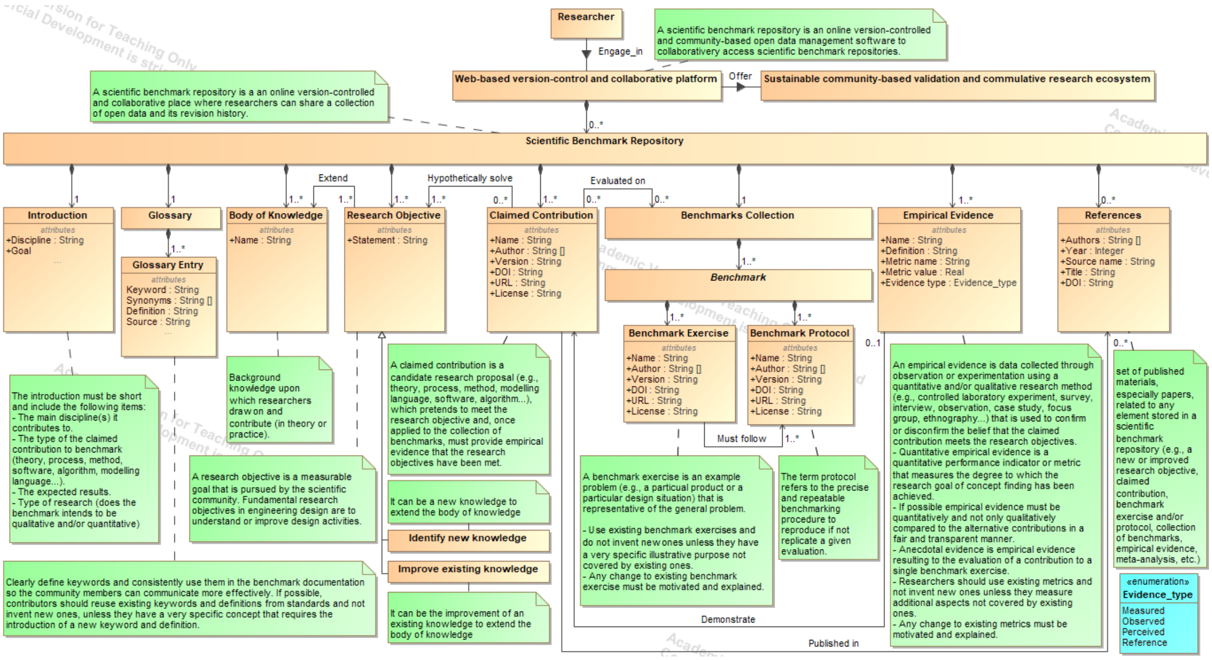

Figure 1. Meta-model formally defining the concept of benchmark in design research

This section aims to illustrate how to apply the benchmark meta-model (Figure 1) to define an open benchmark related to concepts finding for MBADS[4]. The proposed community-based open science platform is concerned primarily with benchmarks created and used by a technical research community, especially the engineering design research community. The benchmark seeks to 1) improve validation quality; 2) facilitate reproducibility if not replicability; 3) foster cumulative research in engineering design; and 4) rapidly learn the essence of background knowledge with an interest in finding its leading edge. The benchmark community gathers the four co-authors, each with his background leading to a mix of expertise. Although members of the engineering design research community may not want to open and share the dirty details of their research data, this proposal does not force them to share all their data, especially input data provided by industrial partners. However, we must all evaluate our claimed contributions using the same tasks and experimental materials to improve validation quality and foster cumulative research.

3.1. Introduction

The introduction should briefly discuss the primary discipline(s) of the research and the goal of the benchmark. In this particular case, the scientific benchmark concerns mainly the engineering design, systems engineering, and product lifecycle management of industrial and scientific communities, and its goal is to collect and compare competing or complementary contributions claiming to improve the activity of concept finding in an MBADS process for system sizing.

[4]https://github.com/GIS-S-mart/Benchmark-7_Model-Based-Architecture-Design-Synthesis/blob/main/Benchmarks/Benchmark-MBADS_System_Sizing-Concept_Finding.md

https://doi.org/10.1017/pds.2024.12 Published online by Cambridge University Press

3.2. Glossary

The definition of the glossary is crucial to establish a shared conceptualisation amongst the members of a domain of discourse, especially in design, where the community defining or using the benchmark gathers contributors from different disciplines. Defining the vocabulary (e.g., design synthesis, design concepts, system architecture, problem modelling, etc.) was our first activity within our benchmark community. To share a common understanding of ambiguous keywords, contributors shall agree upon the meaning of the keywords by systematically providing a definition and synonyms. Contributors should reuse existing terms and definitions from existing standards and not invent new ones unless they have a particular concept that requires the introduction of a new keyword and definition.

3.3. Body of knowledge

The body of knowledge is the background knowledge upon which researchers will draw on and contribute in theory or practice. This benchmark aims to increase the knowledge related to the activity of concept finding (Section 2) in the MBADS approach for system sizing introduced (Table 1).

3.4. Research objective

A research objective is a measurable goal pursued by the benchmark community. Design research objectives belong to two fundamental objectives (Eckert et al., 2004). On the one hand, one can identify new design knowledge by better understanding design as a complex human-centric activity. In this paper, the fundamental research objective is to improve the activity of concept finding in an MBADS process for system sizing. The benchmark community had to agree upon fundamental, co-designed, measurable research objectives with objective evidence transparently shared online. One can create knowledge by proposing a new measurable research objective, a new design method that helps engineers find architecture design concepts or a new means to evaluate the claimed contributions. On the other hand, one can improve existing design knowledge by increasing the level with which a measurable research objective has been met. For instance, one can enhance an existing design method to increase the number of design variables identified during the concept finding activity. Improving new knowledge is inherent to creating new one since a subtle understanding of how a design activity works often suggests improvement opportunities. In design research, we often believe our research objective is new and claim a unique contribution with evidence based on our own example design problem and evaluation metrics. Consequently, cumulative research is minimal without shared fundamental research objectives.

3.5. Claimed contributions

This section is the collection of candidate research proposals, which pretends to meet the research objective and, once applied to the collection of benchmarks, must provide empirical evidence that the research objectives have been met. Comparisons of contributions claiming to improve the concept finding activity are often unfair because they mix five elements: (1) the design philosophy, including the axioms, postulates, assumptions, and convictions taken to be true to serve as a premise or starting point for further reasoning and arguments, (2) the design method in line with the design philosophy and prescribing the concepts to be captured and the design procedure to be followed by the practitioner; (3) the modelling language for encoding the concepts; (4) the modelling software that supports the modelling of the concepts found by the practitioner using the preferred modelling language; and (5) the modelling method, that is, the procedure to be followed by the practitioners to model the concepts using the preferred modelling language and software. As an example, we benchmarked a proposal that 1) adopts a lean engineering design philosophy, 2) combines Function, Physical Principle and Technology (FPPT) and Skins Skeletons (SK2) as a design method, 3) uses the FPPT and SK2 Domain Specific Languages (DSL) based on Ecore, 4) uses Eclipse as modelling software, and 4) follows the modelling method to model FPPT-SK2 concepts with FPPT and SK2 DSLs in Eclipse.

https://doi.org/10.1017/pds.2024.12 Published online by Cambridge University Press

3.6. Benchmark

A benchmark is the combination of a benchmark exercise and a benchmark protocol. However, evaluating a claimed contribution on a benchmark exercise can require following several benchmark protocols, or a protocol can serve as an evaluation procedure for several benchmark exercises.

3.6.1. Benchmark exercise

A benchmark exercise is an example problem that is representative of actual problems. It is essential to build confidence in the appropriateness of the benchmark exercise that will be used to validate the claimed contributions (Carolyn C. Seepersad et al., 2006). Demonstrate the usefulness of claimed contributions for some example design problems require establishing a minimum level of similarity between the benchmark exercises and the actual problem for which the claimed contributions are intended. So far, we have proposed a coupling bolts benchmark exercise and plan to create new ones.

3.6.2. Benchmark protocol

A benchmark protocol is the precise and repeatable benchmarking procedure to reproduce or replicate a given evaluation. By following the benchmark protocol, someone should be able to reproduce, if not replicate, the evaluation of a claimed contribution on the related benchmark exercises.

3.7. Benchmarks collection

The usefulness of a claimed contribution to a benchmark exercise may be anecdotic. All research methods have a flaw or two, and one data collection effort does not lead to a definitive answer to a research question. It is necessary to demonstrate the usefulness of a claimed contribution beyond an example problem. Nevertheless, defining a benchmark exercise that serves as an example problem that encompasses all properties of the actual problem is difficult. Thus, to build confidence in genericity and identify limits, we can use a set of benchmark exercises to confirm or disconfirm that a claimed contribution meets the research objectives. A benchmark collection is a set of benchmark exercises whose union covers all characteristics of the actual problem. The current version of the benchmark collection includes only one benchmark with the coupling bolt benchmark exercise and its benchmark protocol. However, to provide objective evidence on the genericity and limits of each claimed contribution, it is necessary to develop new benchmarks with different characteristics. Continued evolution of a benchmark collection is also essential to prevent researchers from making changes to optimise the performance of their contributions on a particular set of tests.

3.8. Empirical evidence

Empirical evidence is data collected through observation or experimentation using a quantitative and/or qualitative research method to confirm or disconfirm the belief that the claimed contribution meets the research objectives. Quantitative empirical evidence is a quantitative performance indicator or metric that measures the degree to which the research goal of concept finding has been achieved. If possible, empirical evidence must be quantitatively and not only qualitatively compared to the alternative contributions fairly and transparently. Anecdotal evidence is empirical evidence resulting in the evaluation of a contribution to a single benchmark exercise. Researchers should use existing metrics and not invent new ones unless they measure additional aspects not covered by existing ones. Therefore, any change to existing metrics must be motivated and explained.

3.9. References

References are published materials for any open data stored in a benchmark repository.

4. Conclusion and future works

This paper proposed a meta-model to formally define benchmarks to improve validation quality and foster cumulative research in engineering design. The expected minimum viable content for each

https://doi.org/10.1017/pds.2024.12 Published online by Cambridge University Press

element of the meta-model is specified with guidelines and illustrated on the activity of concepts finding in a model-based architecture design synthesis (MBADS) process for system sizing.

In future works, we will mature the current benchmark and invite researchers in design science to contribute by evaluating their research proposals or participating in the co-development of the benchmark content. Involving external researchers will enable us to validate the utility and usability of the benchmarking ecosystem. We will also develop a sound taxonomy of systems architecting problems and extend the benchmark to cover the other activities of the MBADS process for system sizing and deal with different MBADS objectives. Finally, we plan to create a new interface to help non-experts in GitHub contribute and explore an ecosystem of benchmarks by navigating via separate entries.

References

Albert, A., Korkiat, S., Christian, S. and Wolfgang, B. (2009), “A Method to define a Product Architecture Early in Product Development Using the Contact and Channel Model”, Stanford University, Stanford, CA, USA, pp. 241–252.

Barth, A., Caillaud, E. and Rose, B. (2011), “How to validate research in engineering design?”, 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, , Vol. 2: Design Theory and Research Methodology, Copenhagen, Denmark, pp. 41–50.

Blessing, L.T.M. and Chakrabarti, A. (2009), DRM, a Design Research Methodology, Springer London, London, https://dx.doi.org/10.1007/978-1-84882-587-1.

Boniol, F., Wiels, V., Aït-Ameur, Y. and Schewe, K.-D. (2017), “The landing gear case study: challenges and experiments”, International Journal on Software Tools for Technology Transfer, Vol. 19 No. 2, pp. 133–140, https://dx.doi.org/10.1007/s10009-016-0431-4.

Broodney, H., Dotan, D., Greenberg, L. and Masin, M. (2012), “Generic Approach for Systems Design Optimization in MBSE”, INCOSE International Symposium, Vol. 22 No. 1, pp. 184–200, https://dx.doi.org/10.1002/j.2334-5837.2012.tb01330.x.

Cagan, J., Campbell, M.I., Finger, S. and Tomiyama, T. (2005), “A Framework for Computational Design Synthesis: Model and Applications”, Journal of Computing and Information Science in Engineering, Vol. 5 No. 3, pp. 171–181, https://dx.doi.org/10.1115/1.2013289.

Cantamessa, M. (2003), “An empirical perspective upon design research”, Journal of Engineering Design, Vol. 14 No. 1, pp. 1–15, https://dx.doi.org/10.1080/0954482031000078126.

Carolyn C. Seepersad, Pedersen, K., Emblemsvåg, J., Bailey, R. and Allen, J.K. (2006), “The Validation Square: How Does One Verify and Validate a Design Method?”, Decision Making in Engineering Design, ASME Press, pp. 303–313, https://dx.doi.org/10.1115/1.802469.ch25.

Cash, P.J. (2018), “Developing theory-driven design research”, Design Studies, Vol. 56, pp. 84–119, https://dx.doi.org/10.1016/j.destud.2018.03.002.

Eckert, C., Clarkson, P.J., and Stacey. (2003), “The spiral of applied research: A methodological view on integrated design research”, Proceedings of the 14th International Conference on Engineering Design (ICED’03), Stockholm, Sweden.

Eckert, C., Clarkson, P.J. and Stacey, M. (2004), “The lure of the measurable in design research”, presented at the International Design Conference - DESIGN, Dubrovnik, Croatia.

Eisenmann, M., Grauberger, P., Üreten, S., Krause, D. and Matthiesen, S. (2021), “Design method validation–an investigation of the current practice in design research”, Journal of Engineering Design, Taylor and Francis Ltd., Vol. 32 No. 11, pp. 621–645, https://dx.doi.org/10.1080/09544828.2021.1950655.

Escudero-Mancebo, D., Fernández-Villalobos, N., Martín-Llorente, Ó. and Martínez-Monés, A. (2023), “Research methods in engineering design: a synthesis of recent studies using a systematic literature review”, Research in Engineering Design, Vol. 34 No. 2, pp. 221–256, https://dx.doi.org/10.1007/s00163-022-00406-y.

Formentini, G., Bouissiere, F., Cuiller, C., Eric, P. and Claudio, D. (2022), “CDFA method : a way to assess assembly and installation performance of aircraft system architectures at the conceptual design”, Research in Engineering Design, Springer London, Vol. 33 No. 1, pp. 31–52, https://dx.doi.org/10.1007/s00163-021-00378-5.

Frey, D.D. and Dym, C.L. (2006), “Validation of design methods: lessons from medicine”, Research in Engineering Design, Vol. 17 No. 1, pp. 45–57, https://dx.doi.org/10.1007/s00163-006-0016-4.

Gericke, K., Eckert, C., Campean, F., Clarkson, P.J., Flening, E., Isaksson, O., Kipouros, T., et al. (2020), “Supporting designers: moving from method menagerie to method ecosystem”, Design Science, Vol. 6, p. e21, https://dx.doi.org/10.1017/dsj.2020.21.

Gericke, K., Eckert, C. and Stacey, M. (2022), “Elements of a design method – a basis for describing and evaluating design methods”, Design Science, Cambridge University Press, Vol. 8, p. e29, https://dx.doi.org/10.1017/dsj.2022.23.

Gericke, K., Meissner, M. and Paetzold, K. (2003), "Understanding the context of product development", Proceedings of the 19th International Conference on Engineering Design (ICED13) Design For Harmonies, Vol. 3: Design Organisation and Management., Seoul, Korea.

Hartmann, C., Chenouard, R., Mermoz, E. and Bernard, A. (2018), "A framework for automatic architectural synthesis in conceptual design phase", Journal of Engineering Design, Vol. 29 No. 11, pp. 665–689, https://dx.doi.org/10.1080/09544828.2018.1532494.

Hein, A.M. and Lamé, G. (2020), "Evaluating engineering design methods: taking inspiration from software engineering and the health sciences", Proceedings of the Design Society: DESIGN Conference, Vol. 1, pp. 1901–1910, https://dx.doi.org/10.1017/dsd.2020.317.

Hirsch, T. and Hofer, B. (2022), "A systematic literature review on benchmarks for evaluating debugging approaches", Journal of Systems and Software, Vol. 192, p. 111423, https://dx.doi.org/10.1016/j.jss.2022.111423.

Isaksson, O., Eckert, C., Panarotto, M. and Malmqvist, J. (2020), "You need to focus to validate", Proceedings of the Design Society: DESIGN Conference, Vol. 1, pp. 31–40, https://dx.doi.org/10.1017/dsd.2020.116.

Kryssanov, V.V., Tamaki, H. and Ueda, K. (1999), "Synthesis and emergence in engineering design problem solving", Proceedings First International Symposium on Environmentally Conscious Design and Inverse Manufacturing, presented at the Proceedings First International Symposium on Environmentally Conscious Design and Inverse Manufacturing, IEEE, Tokyo, Japan, pp. 690–695, https://dx.doi.org/10.1109/ECODIM.1999.747699.

Lempia, D. and Miller, S. (2009), "Requirements Engineering Management Handbook, U.S. Department of Transportation, Federal Aviation Administration".

Lukas Paehler, Olga Sankowski, Selin Ureten, Matthias Eisenmann, and Dieter Krause. (2023), "Impact of Method Users on the Application of Design Methods - Assessing the Role of Method-related Background Knowledge", DS 125: Proceedings of the 34th Symposium Design for X (DFX2023), presented at the The 34th Symposium Design for X, The Design Society, pp. 055–064, https://dx.doi.org/10.35199/dfx2023.06.

Menu, J. and Nicolai, M. (2017), "Ast 2017 a Framework for Automated Design, Verification , and Simulation of Electrical Power Systems for Aircraft", International Workshop on Aircraft System Technologies, pp. 1–10.

Menu, J., Nicolai, M. and Zeller, M. (2018), "Designing Fail-Safe Architectures for Aircraft Electrical Power Systems", 2018 AIAA/IEEE Electric Aircraft Technologies Symposium, EATS 2018, pp. 1–14, https://dx.doi.org/10.2514/6.2018-5032.

Nuzzo, P., Xu, H., Ozay, N., Finn, J.B., Sangiovanni-Vincentelli, A.L., Murray, R.M., Donzé, A., et al. (2014), "A contract-based methodology for aircraft electric power system design", IEEE Access, Vol. 2, pp. 1–25, https://dx.doi.org/10.1109/ACCESS.2013.2295764.

Pailhès, J., Sallaou, M., Nadeau, J.-P. and Fadel, G.M. (2011), "Energy Based Functional Decomposition in Preliminary Design", Journal of Mechanical Design, Vol. 133 No. 5, p. 051011, https://dx.doi.org/10.1115/1.4004193.

Pinquié, R., Duigou, J.L., Grimal, L. and Roucoules, L. (2022), "An open science platform for benchmarking engineering design researches", Procedia CIRP, Vol. 109, pp. 472–477, https://dx.doi.org/10.1016/j.procir.2022.05.280.

Ribeiro Dos Santos, C.A., Hany Saleh, A., Schrijvers, T. and Nicolai, M. (2019), "CONDEnSe: Contract Based Design Synthesis", Proceedings - 2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems, MODELS 2019, pp. 250–260, https://dx.doi.org/10.1109/MODELS.2019.00009.

Rosich, A., Berx, K. and Pinte, G. (2016), "Model-based design synthesis: application to optimal air-compressor system design", IFAC-PapersOnLine, Vol. 49 No. 21, pp. 416–422, https://dx.doi.org/10.1016/j.ifacol.2016.10.604.

Sim, S.E., Easterbrook, S. and Holt, R.C. (2003), "Using benchmarking to advance research: a challenge to software engineering", 25th International Conference on Software Engineering, 2003. Proceedings., presented at the 25th International Conference on Software Engineering, 2003. Proceedings., IEEE, Portland, OR, USA, pp. 74–83, https://dx.doi.org/10.1109/ICSE.2003.1201189.

Suh, N.P. (2001), Axiomatic Design : Advances and Applications, Oxford University Press.

Thiyagalingam, J., Shankar, M., Fox, G. and Hey, T. (2022), "Scientific machine learning benchmarks", Nature Reviews Physics, Vol. 4 No. 6, pp. 413–420, https://dx.doi.org/10.1038/s42254-022-00441-7.

Tomiyama, T. (2020), "How to write a good STC Dn paper", CIRP Annals Manufacturing Technology.

Vermaas, P.E. (2014), "Design Theories, Models and Their Testing: On the Scientific Status of Design Research", An Anthology of Theories and Models of Design, Springer London, London, pp. 47–66, https://dx.doi.org/10.1007/978-1-4471-6338-1_2.

Yvars, P.A., Lafon, P. and Zimmer, L. (2009), "Optimization of mechanical system: Contribution of constraint satisfaction method", 2009 International Conference on Computers & Industrial Engineering, presented at the Industrial Engineering (CIE39), IEEE, Troyes, France, pp. 1379–1384, https://dx.doi.org/10.1109/ICCIE.2009.5223678.

Yvars, P.-A. and Zimmer, L. (2021), "A Model-based Synthesis approach to system design correct by construction under environmental impact requirements", Procedia CIRP, Vol. 103, pp. 85–90, https://dx.doi.org/10.1016/j.procir.2021.10.013.

Yvars, P.-A. and Zimmer, L. (2022), "Towards a correct by construction design of complex systems: The MBSS approach", Procedia CIRP, Vol. 109, pp. 269–274, https://dx.doi.org/10.1016/j.procir.2022.05.248.