

Reviewing knowledgebase and database grant proposals in the life sciences: the role of innovation

Peter D. Karp *

SRI International, Artificial Intelligence Center, 333 Ravenswood Ave, AE206, Menlo Park, CA 94025, USA

*Corresponding author: Tel: +650-859-4358; Fax: +650-859-3735; Email: pkarp@ai.sri.com

Citation details: Karp, P.D. Reviewing knowledgebase and database grant proposals in the life sciences: the role of innovation. *Database* (2022) Vol. 2022: article ID baac106; DOI: <https://doi.org/10.1093/database/baac106>

Abstract

This article offers thoughts on reviewing grant proposals for biological knowledgebases and databases (KDs) in the hope of aiding grant reviewers and applicants in addressing the issue of innovation. Assessing such grant proposals involves a number of subtleties that are worthy of discussion, particularly for new reviewers and applicants. In part, this article is motivated by the release of two funding opportunity announcements by the US National Institutes of Health concerning KDs. We find that the amount of innovation required for different KD projects can vary significantly, particularly depending on where in its life cycle a given project is. Strong innovation is not necessarily required to have an impactful KD project. For example, PubMed has low innovation but high impact. The importance of innovation should be weighted differently for different KD projects depending on the challenges they face and their maturity. The score for the overall impact of a grant proposal might have little dependence on the innovation score, such as for a mature project that is already delivering strong impact.

Introduction

Evaluating innovation is a critical component of reviewing standard research proposals for agencies such as the US National Science Foundation, Department of Energy and US National Institutes of Health (NIH). Innovation is a review criterion for two funding opportunity announcements (FOAs) released in 2020 by the NIH concerning (NIH) knowledgebases and databases (KDs): ‘Biological Knowledgebase’ (PAR-20-097) and ‘Biomedical Data Repository’ (PAR-20-89). Here, I offer thoughts on reviewing grant proposals for biological KDs in the hope of aiding reviewers and applicants in addressing the issue of innovation for all the preceding funding agencies. Assessing and weighting innovation in such proposals involve a number of subtleties that are worthy of discussion and that have been unresolved for >20 years. Note that a broader discussion of criteria for evaluating life science data resources is available (1), although it does not address the criterion of innovation.

What is innovation?

Evaluation of innovation is a critical component of standard research grants such as NIH R01 (basic research) grants. For such grants, ‘innovation’ refers to novel scientific research results (i.e. novel findings or novel methodology) that are usually published in a scientific journal. Producing such innovations is the primary objective of research grants, and thus, assessment of innovation—the likelihood that a project will produce novel research results—is an essential criterion for evaluating scientific research grants.

However, producing novel research results is not the primary aim of KD projects. Their primary objective is to produce a collection of knowledge or data, and associated software tools, that has a high value for the scientific community by enabling and speeding innovation by that community. KD projects must make that collection available to the community, such as via a website that supports search and analysis operations related to the data. Often, informatics/computational research problems must be solved in the course of producing an impactful KD project. For example, in my group’s development of metabolic pathway KBs, we have solved research problems in the representation of metabolic pathways, pathway search, and visualization of entire metabolic networks.

The innovation review criteria listed in the NIH FOAs do not clearly list informatics research as the type of innovation they seek (although the phrase ‘...improvements...of theoretical concepts’ might be interpreted to include informatics research). For this article’s definition of innovation, we will indeed include publishable solutions to research problems (which I term ‘strong innovation’). Based on the statements in the two KD FOAs, more limited notions of innovation are also appropriate for these projects, such as ‘refinements, improvements or new applications of theoretical concepts, approaches or methodologies’. The database FOA also considers a project to be innovative if it simply makes use of ‘state-of-the-art methodologies, standards and practices’ (calling a project innovative because it makes use of existing standards is a surprisingly low bar). Presumably, the refinements/improvements criterion could be read as ‘refinements and improvements of

Received 26 September 2022; Revised 11 November 2022; Accepted 10 December 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

project methodologies', which could include new ways of marketing KD projects to scientists, engineering improvements that speed the operation of a KD website and user interface improvements to the website of a KD project that make it easier to use.

In summary, we will make use of a broad definition of innovation that includes publishable solutions to informatics research problems, refinements or improvements to existing methodologies and making use of state-of-the-art practices and standards.

The role of innovation

Imagine a proposal for a new KD where the KD clearly meets an unsatisfied need: it collects an important type of data or knowledge (such as that generated by a new experimental technique) that fills a critical void in our scientific information infrastructure. No other KD exists in this area, and we as reviewers anticipate high demand for the KD by the user community. However, imagine that from an informatics viewpoint, it is quite straightforward to create such a KD. Existing informatics techniques will suffice for capturing, searching, presenting and analyzing the data, and no research advances are required to bring this KD to fruition.

How should we review a proposal for such a KD?

One viewpoint is that strong innovation (solving research problems) is always essential, and therefore, a proposal lacking innovation should be dinged substantially.

I advocate a different viewpoint: nothing is wrong with satisfying the KD need using existing technology. The ultimate goal of KD proposals is to satisfy some information need, period. If the investigators can satisfy the need without solving informatics research problems, they should not be penalized for their ability to solve the problem using existing computational methods. Why prevent funding of a project that solves a critical information need without solving a research problem? Indeed, proposals that do not require critical path innovation are at lower risk and have a higher probability of success and impact than proposals for which research problems must be solved in order to solve the ultimate information need. If we weight each proposal by its likelihood of success, a low-risk proposal should be considered more impactful than a high-risk proposal, with all other things being equal.

Consider also that new KD proposals are probably more likely to require strong innovation than proposals for mature KD projects. The longer a KD project has operated, the more likely that it has solved the research informatics problems that have blocked its progress. At some point in its life cycle, it is quite possible that all relevant informatics problems have been solved, and no further strong innovation is required. Should a reviewer ding a proposal for a mature KD that contains no strong innovation? I say no: If the project can fulfill its mission and satisfy the needs of its user community without solving research problems because no significant research problems remain, there is nothing wrong.

We should expect that some mature KD projects will reach a point at which engineering and other improvements continue (e.g. user interface improvements and continuation of curation efforts), but no major research challenges are addressed. For that matter, even if some research challenges remain in the overall field in which the KD project operates, the KD project needs not address those research problems

because future solutions may well be published by other researchers in the field.

Let us consider PubMed as a case study. Introduced in 1996, PubMed now contains 34 million publications and processed 2.57 billion searches in 2021. PubMed is a highly used resource that fulfills an essential role in the biomedical information infrastructure. As far as I am aware, PubMed has performed very little strong innovation over the years, at least from a user perspective (one exception is its use of synonyms to enrich searches (2)); another innovative feature is the 'similar articles' operation (3). For that matter, PubMed has performed little weak innovation as well; for example, its search interface, query language and result presentation have undergone little change in the past 10–15 years. While there may have been some technical changes to its backend, such as using cloud computing to process PubMed searches, I would judge its innovation level as low.

If PubMed were funded by an NIH institute through its extramural program, its funding would likely have been discontinued years ago. Reviewers would likely have criticized it for its lack of both strong and weak innovation and given the proposal a poor score. Other higher scoring proposals would have taken precedence for funding, and PubMed would be no more.

Yet, PubMed has an extremely high usage level and provides high value to the biomedical research community—many researchers and clinicians use it daily and would be hard-pressed to replace it if it disappeared. Without PubMed, doctors would have no access to the newest research and clinical advances, and the quality of patient care would suffer.

How then should reviewers score innovation, both individually and as a component of the overall impact score? The innovation score and the overall impact score should be decoupled. A grant proposal could receive a poor innovation score (e.g. for PubMed) but a high overall impact score when innovation is not required for the success of the project. This is a valid approach for (i) a new project that requires little or no innovation for a high impact or (ii) an already impactful mature project that has many years of innovation behind it but few if any remaining research problems to solve. Sometimes, the anticipated impact of the project may depend strongly on innovation; however, the contribution of the innovation criterion to the overall impact score should be weighted to reflect the nature and maturity of the project.

For example, PubMed might score low on innovation but receive a high impact score because it is highly used and fills a unique role in the biomedical information space.

Innovation as a fraction of impact

Another facet of understanding the relationship between innovation and project impact is that in a research project, the innovation contributes the vast majority of project impact. If the innovation fails and the project produces no research results, it will have essentially no impact. If the innovation succeeds, it constitutes close to 100% of the project's impact. Even a successful research project may have no impact once the innovation has been completed.

In contrast, mature KD projects achieve a massive impact on Day 1 of a new funding cycle. Consider a KD project that has steadily added new data and knowledge and new

informatics tools to its repertoire for a number of years. Imagine that, each year for 10 years, the project added two new informatics operations and tools to its website, for a total of 20 tools. On Day 1 of its new funding period, the KD project is providing those 20 tools to n users, producing an overall impact of $20n$ (for simplicity, we are not considering the impact of the data or knowledge within the resource, and we assume that each tool has an equal impact). If no new tools are added to the project for the next 3 years, the project is still providing a yearly impact of $20n$. If the project instead continued at its previous rate of innovation, adding two more tools per year, it would have an impact of $20n$ in the first year, $22n$ in the second year and $24n$ in the third year (an average of $22n$ per year). Indeed, the project impact increases due to the new innovation, but the average impact is only 10% greater over the course of the 3 years because most of the project impact is the result of work performed in previous grant years—from prior innovations.

Another reason why innovation is a fraction of the overall impact for KD projects is that for many KD projects, innovation is a small fraction of the budget. Most of the budget (e.g. 70–90%) for a KD project will be spent on operational tasks such as curation, website operations, quality assurance, user support, outreach to user communities and software maintenance. The remaining 10–30% supports research and innovative developments. Furthermore, the budget available for innovation tends to decrease over time for a given project because funding agencies are repeatedly decreasing KD project budgets. Although increasing automation can decrease costs, the overall project costs often rise because of inflation and because operations usually become more complex (and hence more expensive) each year due to increases in system complexity (such as larger data volumes and larger amounts of software due to innovations from previous years). Thus, operational costs tend to grow over time and displace the funds available for innovation, since most of the value of the overall project is lost if the project becomes non-operational.

Summary

Innovation in KD projects is a blanket term that covers several areas including informatics research (strong innovation), refinements of existing methods (medium innovation) and utilization of state-of-the-art methods (weak innovation).

The amount of innovation required for different KD projects can vary significantly, particularly depending on where in its life cycle a given project is.

Strong innovation is not necessarily required to have an impactful KD project. For example, PubMed has low innovation but high impact.

The importance of innovation should be weighted differently for different KD projects depending on the challenges they face and their maturity.

The score for overall impact might have little dependence on the innovation score, such as for a mature project that is already delivering a strong impact.

It would be beneficial if the NIH KD FOAs more clearly described the desired types of innovation. Is research important? Are engineering improvements or new outreach methods considered innovation? Is the use of modern methods an innovation? Is the use of standards considered innovation?

It would be beneficial if the NIH KD FOAs clarified the innovation review criterion and stated that its weight in KD projects should be far less than that for basic research projects and that its weight will vary across different KD projects.

The preceding changes will empower investigators to craft projects that produce high-quality KDs, will yield more consistent reviewing by different review panels, and will lessen the chance of poor grant reviews because reviewers and proposers have different understandings and assumptions regarding innovation. I urge reviewers to be conscious of the different mindset needed to review KD proposals as they contribute to building a stronger research community.

Acknowledgements

I thank Ida Sim for contributing the PubMed example.

Funding

SRI International.

Conflict of interest

None declared.

References

1. Drysdale, R., Cook, C.E., Petryszak, R. *et al.* (2020) The ELIXIR core data resources: fundamental infrastructure for the life sciences. *Bioinformatics*, **36**, 2636–2642.
2. Yeganova, L., Kim, S., Chen, Q. *et al.* (2020) Better synonyms for enriching biomedical search. *J. Am. Med. Inform. Assoc.*, **27**, 1894–1902.
3. Lin, J. and Wilbur, W.J. (2007) PubMed related articles: a probabilistic topic-based model for content similarity. *BMC Bioinform.*, **8**, 423.