

A Double-Sided Marketplace for Orphaned Research Projects

Samuel A.W. Klein and Evan A. Warfel

Every lab we've worked in or with has a collection of orphaned projects metaphorically gathering dust. Some are described by the researchers involved as close to being finished. Others have sat untouched for years. Nearly all of them were not abandoned because the findings were non-significant or inconvenient, but rather because the momentum to finish them evaporated.¹

We believe that the accumulation of abandoned, partially complete projects is both substantial and largely invisible, and it points to a phenomenon unaddressed by the broader scientific community. The gap between what a lab can reasonably complete and what the field recognizes as a publishable contribution has been quietly widening for decades. Evidentiary standards have tightened through the increase of multi-study designs, pre-registration, within-paper replications, other responses to the replication crisis, and rightly so. In addition, the volume of scientific literature has drastically increased, often leading to more involved meta-analyses and systematic reviews compared to decades past. But research capacity within labs has remained largely unchanged. Across a wide range of project types — including individual studies that need two to four sibling studies to be publishable, meta-analyses that outlast the students working on them, analyses of archival datasets that stall when a visiting researcher leaves, and so on — competent but incomplete work piles up, out of sight from other researchers. Not only does this represent wasted effort, we believe the scientific record is meaningfully poorer for it due to unsurfaced knowledge and unanswered questions.

The problem we highlight is largely a coordination failure, as it is something no individual lab can solve on its own. Labs and researchers sit on partially completed projects that represent real scientific assets: collected data, potentially analyzed or even turned into interpretable findings. Other researchers have the capacity, motivation, or need to see those projects through to completion (i.e., publication). But there is no mechanism connecting the two sides of this potential market, and so both sit idle.

The reasons projects become orphaned vary but follow recognizable patterns. Many are simply casualties of capacity: the graduate students and postdocs who would have carried the work forward moved on before it could, and the collective motivation to finish left with them. Some orphaned projects conceptually belong to a paper that never materialized, or which would have required a handful of follow-up studies the lab did not have the resources to run. Other projects were exploratory and when the findings didn't point clearly enough toward a follow-up worth pursuing, the project fulfilled its purpose but had nowhere to go.

One of us has multiple studies examining how exposure to counter-stereotypes reduces racially biased decision-making, and how cognitive load moderates the effect of such exposure. Despite yielding interpretable findings, the graduation of undergraduate and graduate-student co-authors have indefinitely stalled the project. One of us has a project at roughly 85% completion, which involves a high-quality dataset of qualitative and quantitative self-esteem measures analyzed in a way that hasn't been done in the field, and which has been stalled for three years since the visiting graduate student collaborator returned to their home institution. The orphaned projects we know about includes large-scale replications and extensions, an abandoned meta-analysis, and analyses of archival data with new modeling techniques, and a variety of

¹ This is distinct from the classical account of why nearly complete studies never see the light of day, more commonly known as the file-drawer problem (where researchers consciously suppress or self-censor projects with non-significant findings)

exploratory studies calling for further investigation. To borrow a phrase from Qiu and Nielsen (2022), orphaned projects are a form of intellectual dark matter: work that is invisible to other researchers, with a near-zero chance of entering the scientific record, and thus near-zero expected value.

Notably, the growing gap between lab capacity and publishable unit of work may strongly discourage exploratory research from being attempted. When an orphaned project has near-zero expected value, a small lab is incentivized to commit only to work that fits cleanly into a planned multi-study paper. Dedicating time and effort to see what happens is likely turning into a luxury fewer labs can afford these days.²

Why Existing Metascientific Practice Does Not Catch Orphaned Projects

Metascientific reforms made in the past two decades have markedly improved the rigor and reproducibility of scientific output. But existing tools don't address the growing volume of orphaned scientific assets sitting outside the scientific record. *Registered Reports* (RRs) address the classical file drawer problem: journals commit to publication before data are collected, reducing the incentive to suppress inconvenient findings. But orphaned projects are typically already collected; RRs can't retroactively rescue them. PsyArXiv and OSF both lower the cost of posting projects to nearly zero, and newer platforms like Octopus.ac and ResearchEquals.com allow researchers to publish individual paper components (e.g., a dataset, an analysis, a hypothesis) as standalone credited outputs. But as the social currency of science remains the published paper, there is little incentive to post work on platforms that lack a mechanism for converting the work into forms that have a meaningful chance of influencing the scientific discourse. The bottleneck is not a shortage of places to deposit scientific work, but an absent connection between orphaned projects and those who could complete them.

A Double-Sided Marketplace for Bundling “Orphaned” Science Research Projects

So far, we've established that labs have research assets but lack the momentum to finish them. We also note that there are groups of researchers that can supply the momentum, but lack the assets. These groups include students looking for projects, postdocs rounding out their CVs, and researchers seeking additional studies for a paper already in progress. Our proposed solution is a platform that could connect the supply of orphaned projects with the demand for incorporating them into other research or otherwise seeing them through to completion. Labs upload short, standardized writeups (1- to 2-page templates) of orphaned projects along with relevant materials (e.g., data, code). Write-ups capture methods, IRB documentation, CRediT contributions, and a standardized characterization of the project's current state. The platform should perform two functions: open browsing and automated matching.

Regarding open browsing, the platform should function as a searchable corpus of available orphaned projects, initially anonymized and only deanonymized following an inquiry. A researcher writing a paper on, say, the relationship between cognitive load and moral judgment could discover three other orphaned studies touching the same construct, request access, and with credit and consent, integrate them together, or work as a new collaborator with the original authors.

² It's worth noting that the incentive for p-hacking likely arose partly from this same dynamic: when the only way to extract scientific value from an exploratory effort was to produce a clean publishable result, the pressure to manufacture one was real.

Regarding automated matching, the platform should actively surface candidate bundles of studies when a critical mass of conceptually related projects accumulates. Using LLM-based parsing of uploaded writeups, datasets, and analytic pipelines, the platform can identify a gap in what's already present across contributing projects, and what might be missing (including potential additional studies, which analyses remain to be run, and what a realistic path to submission looks like). It can then flag researchers who might be well-positioned to take the bundle forward.

Why now? Five years ago, organizing and matching hundreds of project writeups would have required expensive human curation. Now, LLM agents can parse writeups, materials, methods, and code to produce a structured picture of what an orphaned project contains and what it still needs; they can surface candidate matches across projects at low cost; and they can proactively reach out to researchers about potential collaborations, reducing the burden of monitoring the platform actively. The platform we're describing wasn't feasible a few years ago; it is now.

A Two-Stage Pilot

To test both the scale of the problem and the feasibility of the platform, we propose a two-stage pilot designed to be informative regardless of outcome.

Stage 1: a short, anonymous survey targeting 200–400 PIs across social, cognitive, and personality psychology, designed to estimate the orphaned-project population. The survey should count the number of orphaned projects and why they were abandoned, self-assessed methodological soundness, willingness to share under varying conditions of anonymity and credit, and whether the existence of such a platform would potentially change their willingness to undertake exploratory work in the future. The last question probes whether the platform would merely recover stranded work or also inspire additional exploratory research. If median lab counts come back near zero, or PIs report no interest in contributing or changing their exploratory behavior, our premise is weaker than we assumed, this finding should be made public.

Stage 2: a minimum-viable matching exercise, recruiting 10-20 labs to upload 3–5 orphaned projects each in the standardized format, before applying LLM-based matching and light human curation. The success criterion is producing at least 1–2 actual bundled papers end-to-end while characterizing the friction points encountered. Stage 2 is not trying to demonstrate that the platform scales but rather testing whether the proposed mechanism can produce real publications from previously stranded work, and what further development is still needed before a larger version would be worth building. On the supply side, labs may initially offer their weakest stranded projects rather than their best ones, and even well-matched bundles may ultimately prove too heterogeneous in measures or populations to cohere into a publishable paper. Characterizing these risks is what Stage 2 is designed to do.

Conclusion

Competent, incomplete research is sitting in labs around the world, invisible to the researchers who could finish it. The coordination failure we've identified may be large enough and addressable enough to justify investing in a solution. A two-staged pilot study can confirm and potentially offer a proof of concept for a genuinely new layer of scientific infrastructure that recovers stranded knowledge, lowers the barrier to exploratory research, and widens the set of labs that can meaningfully contribute to the scientific record.