Experimentation as a Record: Building Software Under Technical Uncertainty and Documenting It While the Work Happens

Slayter Jameson

The Doer of Things, Pieter Inc.

The Doer of Things Record Series, Working Paper No. 2

November 2026

Author Note

Slayter Jameson is the founder of The Doer of Things, a service of Pieter Inc. The firm sells research and development services and related documentation support, so the author has a commercial interest in the subject of this paper. The author is an engineer and is not a tax advisor, certified public accountant, or attorney. No outside funding was received. This working paper has not been peer reviewed and has not been reviewed by a tax professional.

Correspondence concerning this paper should be addressed to The Doer of Things through https://iamthedoerofthings.com/contact.

Abstract

Most software work is delivery: the team knows what to build and roughly how. A smaller share is investigation, where nobody can say with confidence that the approach will work. This paper describes a working method for the second kind and shows that the method leaves behind, as a by-product, the kind of record the Treasury regulations describe when they define a process of experimentation. The method has five steps: state the question, list the alternatives before choosing one, build the smallest thing that can fail, fix the decision threshold before the test, and keep the failures. Each step is tied to engineering literature on risk-driven development and experimentation and, where the comparison is fair, to the wording of Treasury Regulation section 1.41-4(a)(5). The paper then reads the recordkeeping rule for the research credit, which asks for records in sufficiently usable form and detail and does not prescribe a format, and argues that records made during the work are stronger than reconstructions because they carry a date, state the uncertainty before the answer, and preserve the failures. It proposes a minimum record of four habits and a weekly three-line note. The correspondence between method and regulation is structural. It does not show that any work is qualified research, and nothing here is tax or legal advice. Case law was not surveyed. Each claim is graded by the kind of evidence behind it.

Keywords: process of experimentation, software engineering, technical uncertainty, research documentation, recordkeeping, Internal Revenue Code section 41

Experimentation as a Record: Building Software Under Technical Uncertainty and Documenting It While the Work Happens

Most software work is delivery. The team knows what to build and roughly how, and the job is to do it well. A smaller share is different in kind. Nobody on the team can say with confidence that the thing will work, and the plan is in truth a list of guesses. Brooks (1987) located the hard part of building software in the specification, design, and testing of the conceptual construct, as distinct from the labor of representing it in code. Work of the second kind is where that difficulty is concentrated.

The two kinds of work need different handling. A team that runs an investigation as though it were delivery can spend a quarter and finish with little to show for it. This paper describes a method for the investigative kind that the author's firm uses in its own engineering work. It then makes a second point. A team that works this way produces, without a separate effort, most of what a tax professional would need in order to evaluate the work under the federal research credit. The paper does not say that work done this way qualifies for that credit. That determination depends on facts this paper cannot know and belongs to a qualified tax professional.

Scope and Method

The paper combines three kinds of material and keeps them apart. The working method is practice: it reflects how the author's firm runs uncertain engineering work, and it is offered as such. The engineering and behavioral literature cited in support was checked against the published record of each source, and each is used only for the general proposition stated. The legal material is a reading of primary text by a non-lawyer. The regulation defining a process of

experimentation, the recordkeeping rule, the general records regulation, and the statutory definition of qualified research expenses were read in full for the provisions cited, in the versions current on November 28, 2026. Case law was not surveyed, and a reader who intends to rely on any legal statement here should have a certified public accountant or tax attorney review it first. The first paper in this series sets out the four statutory requirements for qualified research and is not repeated here.

Experimentation in the Engineering Literature

The idea that uncertain work should be organized around its risks is old in software engineering. Boehm (1988) proposed a spiral model in which each cycle of development begins by identifying objectives, alternatives, and constraints, then evaluates the alternatives and resolves the largest risks, often by prototyping, before committing to the next stage. The model is risk-driven: what the team does next depends on what it is least sure of.

Research on product development reaches a similar view from another direction. Thomke (1998) treated experimentation as a form of problem solving carried out in repeated cycles, and studied how the choice of experimental mode, such as computer simulation or physical prototypes, and the cost and speed of each cycle affect development. The implication for a software team is that an experiment is a unit of work with a cost, and that cheap, fast experiments placed early are worth more than expensive ones placed late.

Neither source was written about tax, and neither is evidence about any legal standard. They are cited to show that the method described next rests on established engineering thinking and was not designed backward from a regulation.

A Working Method for Uncertain Software Work

State the Question

A feature is a statement of intent, such as adding real-time collaboration to an editor. A question is narrower: can two clients be kept consistent with less than 200 milliseconds of delay on the current data model? The question has an answer. It can be yes, no, or yes with conditions. Writing it down turns an open-ended build into an investigation that can end.

List the Alternatives Before Choosing One

There is usually more than one way in. The method asks the team to write down at least two approaches, with what each would cost and what result would lead the team to abandon it. Teams skip this step because the first idea feels right. Six weeks later they change approach, and nobody can say why the first was chosen or what was learned from it. This is the step Boehm (1988) places at the start of every cycle.

Build the Smallest Thing That Can Fail

The purpose of a first prototype is to find out whether the approach is dead. The team builds the piece that carries the risk and leaves out everything else: sign-in, styling, and error handling beyond what the test needs. If the piece works, the rest of the build is justified. If it fails, the team has spent days where it might have spent months.

Fix the Decision Threshold Before the Test

Before running a test, the team writes down what it expects and what measured value would change its mind. Without that, almost any result can be read as encouraging. Scientists call the same discipline preregistration. Nosek et al. (2018) describe it as a way to keep prediction separate from postdiction, the explanation of results after they are seen, since the two are easily

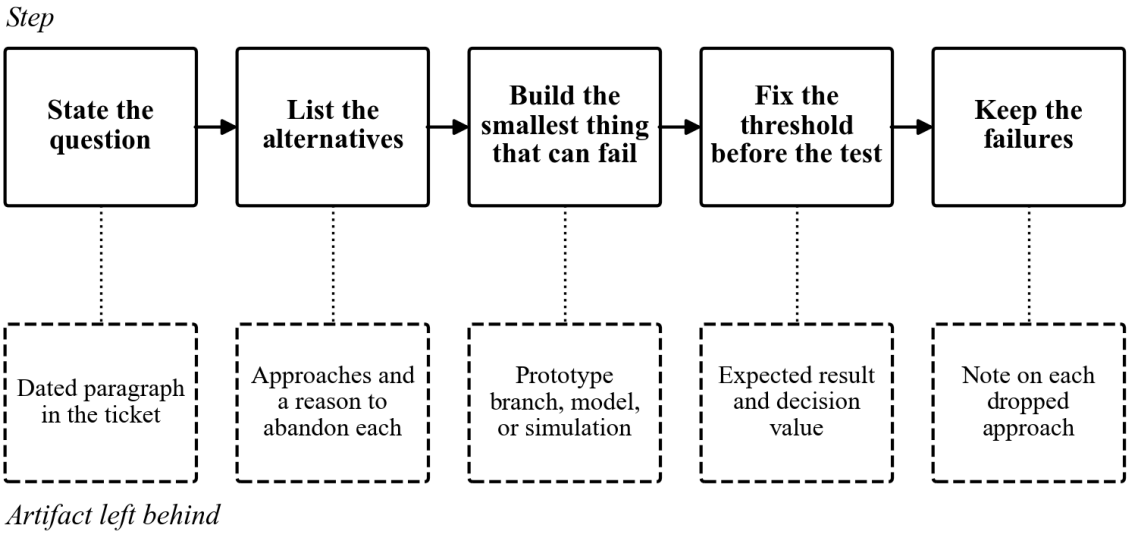
confused once the data are known. An engineering benchmark is not a scientific study, and the comparison is an analogy. The underlying point carries over: a threshold set in advance separates an experiment from a demonstration.

Keep the Failures

The branch that did not work is evidence that an alternative was real and was evaluated. A short note is enough: what was tried, what happened, and why it was dropped. Deleting it to tidy the repository throws that evidence away. Nygard (2011) makes a related argument about architecture. He observes that the motivation behind a decision is among the hardest things to track over the life of a project, and proposes keeping a short dated record for each significant decision that states its context and consequences. Figure 1 shows the five steps and the artifact each one leaves.

Figure 1

Five Steps of the Working Method and the Artifact Each Step Leaves



Note. The steps are the author's practice. The artifacts listed are examples. The figure does not represent a legal test.

How the Method Corresponds to the Regulation

The Treasury regulations define a process of experimentation as a process designed to evaluate one or more alternatives to achieve a result where the capability or the method of achieving that result, or its appropriate design, is uncertain as of the beginning of the taxpayer's research activities. The process involves identifying the uncertainty, identifying one or more alternatives intended to eliminate it, and identifying and conducting a process of evaluating the alternatives, for example through modeling, simulation, or a systematic trial and error methodology (Treas. Reg. § 1.41-4(a)(5)(i)). The regulation adds that the process must be evaluative and generally should be capable of evaluating more than one alternative. Table 1 sets the steps of the method beside these elements.

Table 1

Steps of the Working Method Beside the Elements Named in the Regulation

Step in the method	Element in Treas. Reg. § 1.41-4(a)(5)(i)	Artifact the step leaves
State the question	Identification of uncertainty concerning the development or improvement of a business component	Dated paragraph in the ticket or design document
List the alternatives	Identification of one or more alternatives intended to eliminate that uncertainty	Short list of approaches with a reason to abandon each
Build the smallest thing that can fail	Conduct of a process of evaluating the alternatives	Prototype branch, test harness, model, or simulation
Fix the threshold before the test	No direct counterpart. Supports the requirement that the process be evaluative	Expected result and decision value, written before the run
Keep the failures	No direct counterpart. Shows that more than one alternative was evaluated	Note on each dropped approach, with the branch or data retained

Note. The middle column quotes or closely paraphrases the regulation. The pairing of rows is the author's and is structural only. It does not establish that any activity is qualified research.

The correspondence should not be pressed further than it goes. A process of experimentation is one of four statutory requirements, each applied separately to each business component, and the statute also excludes several categories of activity outright. The regulation itself warns that uncertainty concerning the development of a business component does not establish that all activities undertaken to achieve it constitute a process of experimentation (Treas. Reg. § 1.41-4(a)(5)(i)). The fraction of activities that must do so is fixed at 80% or more, measured on a cost or other consistently applied reasonable basis (Treas. Reg. § 1.41-4(a)(6)). A team that follows the method has a clearer account of what it did. Whether that account satisfies the law is a separate determination.

The Recordkeeping Standard

The regulations state the recordkeeping rule for the research credit in one sentence. A taxpayer claiming the credit must retain records in sufficiently usable form and detail to

substantiate that the expenditures claimed are eligible for the credit (Treas. Reg. § 1.41-4(d)). The same paragraph refers to the general rule on records, which requires a person subject to income tax to keep such permanent books of account or records as are sufficient to establish the amount of gross income, deductions, credits, or other matters required to be shown in a return (Treas. Reg. § 1.6001-1(a)). It also provides that the Internal Revenue Service and taxpayers may agree to guidelines for the keeping of specific records.

Two features of that text are worth stating plainly. First, the paragraph does not prescribe a format, a tool, or a list of required documents. The standard is one of usability and detail. Second, the word contemporaneous does not appear in the regulation. The argument made in the next section, that records made during the work are stronger than those assembled afterward, is an argument about evidence and is not a quotation of a legal requirement. What a given examiner or court expects in practice is a subject on which case law and professional experience govern, and this paper did not survey either.

The form on which the credit is claimed has moved toward reporting by business component. The Internal Revenue Service (2025) states that the section of the form that collects business component information is optional for tax years beginning before 2026 and will be required for tax years beginning after 2025, subject to exceptions described in the instructions. The instructions change, and the current version should be consulted. The direction is consistent with a record that is organized by component from the start.

Why Records Made During the Work Are Stronger

The phrase sufficiently usable form and detail is where many companies fall short. The work happened, and the evidence that it involved uncertainty and experiment is scattered across

chat threads, closed tickets, and the memory of an engineer who has since left. A record written while the work happens has three properties that a reconstruction lacks.

The first is a date. A dated note shows what was known on that day. The second is the uncertainty, stated before the answer. A problem written down in March and solved in May is hard to characterize as something invented afterward. The third is the failures. People who write after the fact tend to tell a clean story, and people who write during the work record what went wrong. The classic experiments on hindsight found that people told the outcome of an event judged it to have been more likely than people who were not told, and were largely unaware of the effect (Fischhoff, 1975). That research was not done on engineers, and it does not show that any particular reconstruction is wrong. It gives a reason to expect that accounts written after the answer is known will understate how open the question was.

Reconstruction also costs more, in the author's experience. It pulls senior engineers away from current work to remember old work, and the result is weaker than a note written at the time would have been. That is a practitioner's observation and has not been measured.

A Minimum Record

An adequate record takes no new tool. It takes four habits. Table 2 lists them with the reason for each.

Table 2

Four Habits That Make Up a Minimum Record

Habit	What is written	Why
Name the component	A tag on tickets and design documents that ties uncertain work to one product, process, software, technique, formula, or invention	The statute applies its requirements separately to each business component (26 U.S.C. § 41(d)(2))
Write the uncertainty at the start	One dated paragraph: what the team is trying to make work, what it does not know, and why the answer is not already available to it	Uncertainty is defined as of the beginning of the research activities
Record alternatives and results	What was tried, what happened, what changed, with links to commits, test output, benchmark runs, lab notes, and design reviews	These are the elements the regulation names in a process of experimentation
Note who did the work	Names and an approximate share of time per component, kept monthly	A tax professional has to connect people and costs to activities

Note. The habits are the author's recommendation. The regulations do not prescribe them.

The fourth habit reflects how the statute builds up the amounts involved. Qualified research expenses include in-house research expenses, and those include wages paid or incurred to an employee for qualified services performed by that employee (26 U.S.C. § 41(b)(1), (b)(2)(A)). Qualified services are services consisting of engaging in qualified research or in the direct supervision or direct support of research activities that constitute qualified research (26 U.S.C. § 41(b)(2)(B)). The definitions tie cost to what a person was doing. A monthly estimate made at the time is, in the author's view, easier to stand behind than an annual estimate made from memory. How time is measured and allocated for a given company is a matter for its tax professional, and this paper does not address the computation of any amount.

The Weekly Note

The cheapest form of the record is three lines written by each engineer at the end of the week: what the team was trying to make work, what it tried, and what happened. The note is

dated and kept in the repository or a shared document. Fifty-two of those in a year amount to a real record of the work, written by the people who did it, in their own words.

Practices to Avoid

Three practices weaken a record. The first is backdating. A record should carry the date on which it was written, and a note written later should say so. The second is writing toward a conclusion. The engineer's job is to record facts. Whether the work qualifies is for the tax professional to decide, and notes that read like advocacy are less convincing than notes that read like engineering. The third is documenting everything. Routine work is routine, and the effort belongs on the projects where something was unknown at the start.

Evidentiary Status of the Claims in This Paper

Table 3 states the kind of evidence behind each principal claim and what would strengthen it.

Table 3

Kind of Evidence Behind Each Principal Claim

Claim	Kind	What would strengthen it
The elements of a process of experimentation and the 80% measure	Regulatory. Read from the current regulation.	Case law on how the elements have been applied to software development.
The recordkeeping rule asks for usable form and detail and prescribes no format	Regulatory. Read from the current regulation.	Case law and IRS guidance on what records have been accepted or rejected.
Wages for qualified services are part of qualified research expenses	Statutory. Read from the current text of section 41(b).	Review by a tax professional. Not developed further here.
Uncertain work is best organized around risks and alternatives	Established engineering literature.	None needed for the general proposition.
Thresholds set in advance keep a test honest	Analogy to scientific preregistration.	Evidence from engineering teams specifically.
The five-step method leaves the elements of a process of experimentation as a by-product	Structural comparison by the author.	Review of real project records by a tax professional.
Records made during the work are stronger than reconstructions	Argument from the nature of the evidence, supported by hindsight research in general populations.	Case law on reconstructed records. Studies of engineering recall.
Reconstruction costs more and yields less	Practitioner observation. Not measured.	Time and outcome data across engagements.

Limitations

The method described here is one firm's practice and has not been tested against alternatives. The examples are illustrative and were written by the author. The legal reading was done by an engineer from the text of the statute and regulations as they stood on November 28, 2026, and it has not been reviewed by a tax professional. Case law and administrative guidance, apart from one set of form instructions, were not surveyed, and both bear heavily on what records are accepted in practice. The paper says nothing about the other three statutory requirements, the statutory exclusions, the treatment of software developed for internal use, or

the computation of any amount. The author's firm sells engineering and documentation services, and the paper's recommendations should be weighed with that interest in mind.

Conclusion

A team that states its question, lists its alternatives, builds the smallest test that can fail, fixes its threshold in advance, and keeps its failures is doing sound engineering. It is also writing down, in passing, an uncertainty, a set of alternatives, an evaluation, and a decision. Those are close to the elements the Treasury regulations name when they define a process of experimentation. The tickets, design notes, code changes, and test logs already tell the story. Someone has to gather them and tie each to the business component it belongs to, and a qualified tax professional has to decide what they show. The habit of working in a way that leaves the trail is one any team can start this week.

Disclaimer

The Doer of Things provides research and development services and related documentation support. We are not a tax advisor, CPA, or attorney. Eligibility for the federal R&D tax credit under Section 41 (including any payroll tax election under Section 41(h)) depends on the specific facts and circumstances of each company and must be determined by qualified tax professionals. We do not guarantee any tax credit amount or outcome. Companies should consult their own CPA, tax attorney, or other qualified advisors.

References

Boehm, B. W. (1988). A spiral model of software development and enhancement. *Computer, 21*(5), 61-72. https://doi.org/10.1109/2.59

Brooks, F. P., Jr. (1987). No silver bullet: Essence and accidents of software engineering. *Computer, 20*(4), 10-19. https://doi.org/10.1109/MC.1987.1663532

Credit for Increasing Research Activities, 26 U.S.C. § 41 (2026). https://www.law.cornell.edu/uscode/text/26/41

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 288-299. https://doi.org/10.1037/0096-1523.1.3.288

Internal Revenue Service. (2025). *Instructions for Form 6765: Credit for increasing research activities* (Rev. December 2025). U.S. Department of the Treasury. https://www.irs.gov/instructions/i6765

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

Nygard, M. (2011, November 15). *Documenting architecture decisions*. Cognitect. https://cognitect.com/blog/2011/11/15/documenting-architecture-decisions

Qualified Research for Expenditures Paid or Incurred in Taxable Years Ending on or After December 31, 2003, 26 C.F.R. § 1.41-4 (2026). https://www.ecfr.gov/current/title-26/section-1.41-4

Records, 26 C.F.R. § 1.6001-1 (2026). https://www.ecfr.gov/current/title-26/section-1.6001-1

Thomke, S. H. (1998). Managing experimentation in the design of new products. *Management Science, 44*(6), 743-762. https://doi.org/10.1287/mnsc.44.6.743