

## THE DATA FOR THE PROJECT

---

**T**his project started with a desire to understand how to make technology great and how technology makes organizations better. Specifically, we wanted to investigate the new ways, methods, and paradigms that organizations were using to develop and deliver software, with a focus on Agile and Lean processes that extended downstream from development and prioritized a culture of trust and information flow, with small cross-functional teams creating software. At the beginning of the project in 2014, this development and delivery methodology was widely known as “DevOps,” and so this was the term we used.

Our research design—a cross-sectional data collection<sup>1</sup> for four years—recruited professionals and organizations familiar with the word DevOps (or at least willing to read an email or social media post with the word DevOps), which targeted our data collection accordingly. Any good research design defines a target population, and this was ours. We chose this strategy for two primary reasons:

1. **It allowed us to focus our data collection.** In this research, the users were those who were in the business of

software development and delivery, whether their parent organization's industry was technology or was driven by technology, such as retail, banking, telecommunications, healthcare, or several other industries.

2. **It allowed us to focus on users who were relatively familiar with DevOps concepts.** Our research targeted users already familiar with terminology used by technology professionals who use more modern software development and delivery practices, whether or not they identified as DevOps practitioners. This was important, because time and space were limited, and too much time spent on background definitions and a long explanation of concepts, such as continuous integration and configuration management, could risk survey respondents opting out of the study. If a survey reader has to spend 15 minutes learning about a concept in order to answer questions about it, they will get frustrated and annoyed and won't complete the survey.

This targeted research design was a strength for our research. No research design is able to answer all questions, and all design decisions involve trade-offs. We did not collect data from professionals and organizations who were not familiar with things like configuration management, infrastructure-as-code, and continuous integration. By not collecting data on this group, we miss a cohort that are likely performing even worse than our low performers. This means our comparisons are limited and we don't discover the truly compelling and drastic transformations that are

possible. However, we gain explanatory power by limiting the population to those that fall into a tighter group definition. That increase in explanatory power comes at the expense of capturing and analyzing the behaviors of those that do not use modern technology practices to make and maintain software.

This data selection and research design did require some caution. By only surveying those familiar with DevOps, we had to be careful in our wording. That is, some who responded to our survey might want to paint their team or organization in a favorable light, or they might have their own definition of key terms. For example, everyone knows (or claims to know) what continuous integration (CI) is, and many organizations claim CI as a core competency. Therefore, we never asked any respondents in our surveys if they practiced continuous integration. (At least, we didn't ask in any questions about CI that would be used for any prediction analysis.) Instead, we would ask about practices that are a core aspect of CI, e.g. if automated tests are kicked off when code is checked in. This helped us avoid bias that could creep in by targeting users that were familiar with DevOps.

However, based on prior research, our own experiences, and the experiences of those who have led technology transformations in large enterprises, we believe that many of our findings are broadly applicable to teams and organizations undergoing transformations. For example, the use of version control and automated testing is highly likely to yield positive results, whether a team is using DevOps practices, Agile methodologies, or hoping to improve their lockstep waterfall development methods. Similarly, having an organizational culture that values

transparency, trust, and innovation is likely to have positive impacts in technology organizations regardless of software development paradigm—and in any industry vertical, since that framework is predictive of performance outcomes in different contexts, including healthcare and aviation.

Once we defined our target population, we decided on a sampling method: How would we invite people to take the survey? There are two broad categories of sampling methods: probability sampling and nonprobability sampling.<sup>2</sup> We were not able to use probability sampling methods because this would require that every member of the population is known and has an equal chance of participating in the study. This isn't possible because an exhaustive list of DevOps professionals in the world doesn't exist. We explain this in more detail below.

To collect the data for our research, we sent out emails and used social media. Emails were sent to our own mailing lists, which consisted of technologists and professionals who worked in DevOps (e.g., were in our database because they had participated in prior years' studies, were in Puppet's marketing databases because of their work with configuration management, were in Gene Kim's database because of their interest in his books and work in the industry, or were in Jez Humble's database because of their interest in his books and work in the industry). Emails were also sent to mailing lists for professional groups. Special care was also taken to send invitations to groups that included underrepresented groups and minorities in technology. In addition to direct invitations by email, we leveraged social media, with authors and survey sponsors tweeting links to the survey and

posting links to take the survey on LinkedIn. By inviting survey participation from several sources, we increased our chances of exposure to more DevOps professionals while addressing limitations of snowball sampling, discussed below.

To expand our reach into the technologists and organizations developing and delivering software, we also invited referrals. This aspect of growing our initial sample is called referral sampling or snowball sampling because the sample grows by picking up additional respondents as it spreads, just like a snowball grows as you roll it through the snow. Snowball sampling was an appropriate data collection method for this study for several reasons:

- **Identifying the population of those who make software using DevOps methodologies is difficult or impossible.** Unlike professional organizations like accounting or civil engineering, which in the US have national certifications such as CPA (Certified Public Accountants) or PE (Practice of Engineering), there is no central accrediting board that could give us a list of professionals to reference. Beyond this, we could not scour organization charts (even if they were publicly available) for job titles as not everyone has “DevOps” or other important keywords in their job title. In addition, many technologists, especially at the beginning of the research project, had nontraditional job titles. Even if organization charts were public, many job titles are too generic to be useful for recruitment in the study (such as “software

engineer,” which can include developers working in teams using waterfall or DevOps methods). Snowball sampling is a method well suited for studying specific groups whose populations cannot be easily identified.

- **The population is typically and traditionally averse to being studied.** There is a strong (and unfortunate) history of organizational studies of technical workers leading to “Lean transformations” which really just mean a significant workforce reduction. Snowball sampling is a method that is ideal for populations that are often averse to being studied; by referring others to the study, they can vouch for the questions (reassuring the new participant that the questions are not propaganda) or even for the reputation of the researchers.

There are some limitations inherent in snowball sampling. The first limitation is the potential that the initial users sampled (in our case, emailed) are not representative of the communities they belong to. We compensated for this by having an initial set of invitations (or informants) that was as large and as diverse as possible. We did this by combining several mailing lists, including our own survey mailing list, which had a diverse set of respondents covering a large variation from company size and countries. We also reached out to underrepresented groups and minorities in technology through their own mailing lists and organizations.

Another limitation of snowball sampling is that the data collected is strongly influenced by the initial invitations. This is a

concern if only a small group of people are targeted and then asked for referrals, and the sample grows from there. We addressed this limitation by inviting a very large and diverse group of people to participate in the study, as described above.

Finally, there may be a concern that findings will not be representative of what is actually happening in the industry, that we may have blind spots that we do not see in our data. We address this in a few ways. First, we do not simply rely on the research results each year to inform our conclusions; we actively engage with the industry and the community to make sure we know what is happening, and triangulate our results with emerging trends. That means we actively seek feedback on our survey, through the community at conferences, and through colleagues and the industry; we then compare notes to see what trends are emerging, never relying on only one data source. If any discrepancies or mismatches occur, we revisit our hypotheses and iterate. Second, we have external subject matter experts in the industry review our hypotheses each year to ensure we are current. Third, we explore the existing literature to look for patterns in other fields that may provide insights into our study. Finally, we ask for input and research ideas from the community each year and use these ideas when we design the research.

---

<sup>1</sup> A cross-sectional design means the data was collected at a single point in time. However, it precluded us from longitudinal analysis because our responses are not linked year over year. By repeating the study over four years, we were able to observe patterns across the industry. While we would like to collect a longitudinal data set-that is, one where we sample the *same* individuals year over year-this could reduce response rates due to privacy concerns. (And what happens when those people change teams or jobs?) We are currently pursuing research in this

area. Cross-sectional research design does have its benefits: data collection at a single point in time reduces variability in the research design.

<sup>2</sup> Probability sampling is any method of statistical sampling that uses random selection; by extension, nonprobability sampling is any method that does not use random selection. Random selection ensures that all individuals in a population have an equal chance of being selected in the sample. Therefore, probability sampling is generally preferred. However, probability sampling methods are not always possible because of environmental or contextual factors.