

Stanford SOCIAL INNOVATION^{Review}

Next Generation Nonprofits Supplement

Sponsored by AWS

Data Lakes

By Richard Kostro

Stanford Social Innovation Review
Winter 2023

Copyright © 2022 by Leland Stanford Jr. University
All Rights Reserved

Stanford Social Innovation Review
www.ssir.org
Email: editor@ssir.org

Data Lakes

A powerful tool for collecting and analyzing data requires nonprofits to rethink not just their technology, but their culture.

BY RICHARD KOSTRO

Technology has afforded nonprofits access to a tremendous amount of data to raise revenue, run their operations, and pursue their mission. But access is not the equivalent of understanding the full story that data can tell.

It is easy for organizations to make sense of their data when they only have to deal with a few data sets. However, the process becomes more difficult when additional data sets are added, particularly when those additional data sets come from tools purchased from different vendors and that have their own unique data structure. When this happens, traditional spreadsheets and databases are insufficient. Simple databases can crash under the weight of too many users, and large, complex spreadsheets with thousands of lines of embedded calculations can freeze or run maddeningly slow.

Fortunately, data lakes allow users to store, manage, clean, transform, and analyze different data sets in one place. To be clear, a data lake is not a data warehouse—the main difference being the order in which data is loaded and transformed. Traditional data warehouses require data to be transformed first, and only data that adheres to a specific format and schema are accepted. This generally means that not all data can be stored, especially newly acquired or unstructured data sets, leading organizations to lose access to that data. Data warehouses are also typically just storage facilities. They don't contain the tools that allow for data cataloging, reporting, governance, advanced analysis, and machine learning.

Data lakes, on the other hand, allow data to be stored immediately and create or change schema over time. They also come with a complete complement of modular tools for access control, cataloging, and analytics, which together enable organizations to easily customize their data lake for their specific needs.

At Share Our Strength, a US nonprofit committed to ending childhood hunger, we addressed the challenges of managing our many discrete fundraising data streams by initiating the process to implement a data lake in 2021. This has put us on a path to replace the organization's

numerous independently managed spreadsheets with a single, accessible, shared system of record. It has reduced the burden on our database administrators by providing automated, real-time reporting. It has further prepared us to use advanced concepts in AI and machine learning to create better predictive models of donor behavior to improve our fundraising. Also, the data lake has made us more cyber-secure by giving us the capability to grant or restrict access to sensitive data sets on a granular level.

The technical journey from identifying the need for a data lake to implementing one was relatively simple. The cloud-based tools for storing, transforming, cataloging, and analyzing our data as well as controlling access to it were readily available from AWS. Far more difficult than the data engineering, however, was the cultural engineering required to bring new teams together, build trust in a new way of doing business, and agree on a new set of standards for managing our data.

CULTURAL REENGINEERING

During the process we learned that the real work in building a data lake is in reengineering the way that the organization thinks about its data. Our team estimates that we spent 80 percent of our time building



RICHARD KOSTRO is the senior vice president and CIO at Share Our Strength and serves on the board of directors of The Nonprofit Alliance Foundation.

consensus across teams, and only 20 percent of the time designing and implementing the actual infrastructure.

One of the first areas of cultural reengineering was to make the IT department a better partner to the fundraising department. Prior to when I joined Share Our Strength as its CIO in 2017, the IT department focused almost exclusively on server maintenance and desktop support, and—although we had technical skills to offer—we were never part of the conversation of how to better manage data. Therefore, my first step was to expand the services the IT department offered to include building and maintaining the organization's data lake. In doing so, my department offered a new source of creativity, capability, and perspective to our business units' efforts to use data and technology wisely.

A second aspect of our cultural reengineering was building trust between the IT and fundraising departments. This effort took time and patience for people coming from different disciplines to learn each other's language and to trust each other's expertise. For example, when the IT department demonstrated the data lake's capabilities, we were met with some skepticism because the analyses looked different from the Excel spreadsheets the fundraisers were familiar with. At the same time, we in the IT department also needed to guarantee that we were demonstrating capabilities germane to the fundraisers' current priorities, such as automating data-cleaning tasks that occupied an outsize part of their day, instead of demonstrating capabilities that would only be potentially useful in the future, such as finding unique trends in donor behavior. It took several iterations and rounds of clarifying questions before we began to understand each other's perspectives and were then able to establish a prioritized list of problems to solve.

Finally, our organization embarked on a third area of cultural reengineering: developing a shared understanding of how our data was captured and structured. Often, the rules for storing and structuring data made sense for one fundraising team, but those same rules did not easily translate and apply to the work of other teams. This led to irrelevant data being stored in some structured fields, a lot of important data being stored in comment fields that did not easily lend themselves to analysis, and a realization that sometimes our analyses were incomplete because they did not fully take into consideration the nuances of how or where our data were stored. Data cataloging, which is a central feature of data lakes, gave us the tools and opportunity to identify and resolve these irregularities and ultimately improve our analyses.

Share Our Strength is only a year into using a data lake to improve our fundraising, but we are already seeing tremendous gains. We are beginning to measure returns on various investments with greater specificity because we are including more and better data streams, which in turn allows us to make smarter decisions. We are also working to automate complicated yet repetitive data transformations and data pulls—saving our donor operations team time as they support major- and mid-tier fundraising officers in multiple markets. Finally, our more consistent data standards are helping us to better understand donor behaviors, which has allowed us to have more meaningful communications with them.

The data lake contributed to each of these successes, yet the lake itself was only possible thanks to our nonprofit's cultural transformation. For organizations looking to take a similar journey to improve their fundraising capabilities, our advice is to build trust within the organization and then *iterate, iterate, iterate*, as there are likely several unknown opportunities for innovation within your own data sets. ♦

Accelerating Mission Impact

Cloud technology has fueled Moneythink's ability to better serve students in real time.

BY JOSHUA LACHS

A few years ago, my nonprofit organization, Moneythink, went through a digital metamorphosis. Since 2008, we have worked to increase the number of historically marginalized students graduating with college degrees while carrying little to no financial burden. Early on, we identified college matriculation and affordability as critical inflection points that were prime for innovation. In 2016, we committed to finding the most strategic intervention for the largest number of students who could benefit from our services.

In partnership with the behavioral science firms Ideo.org and Ideas42, we embedded a human-centered design approach into our product development. Meeting with students, their families, and student-advisors helped us to understand and think through what successful digital experiences could look like for our primary beneficiaries—current and future generations of education consumers.

Cloud technology proved to be the right solution for translating confusing financial aid letters, thus allowing users to compare their college affordability options.

Between 2016 and 2019, we ran a virtual college financial-coaching program with one-on-one college and financial guidance for Pell Grant-eligible high school seniors in Chicago and the San Francisco Bay Area. Our students were supported throughout the college application and matriculation processes, including FAFSA completion, college fit and matching, financial aid award letter comparison, and personal financial planning. Our coaches used an in-house prototype to have more effective advising conversations, which led our students to make more informed

decision about which college to attend. We tested several tools within our program that expedited previously laborious tasks such as collecting financial award letters and translating their information by hand. And we discovered that our technology enabled students to quickly understand the affordability of college options, so that our advisors could then focus on the conversations that really matter.

Despite our successful virtual coaching program, we still faced barriers to large-scale growth and impact. Our in-person student

JOSHUA LACHS is the CEO of Moneythink. For more than 20 years, he has served in leadership positions within the college success and career-readiness fields.