

# Health System Informatics Leader Uses Flywheel to Create *AI-Ready Data Sets*

*Platform serves annotation and analysis needs for both pathology and radiology*

The University of Texas Medical Branch (UTMB) is an academic health science center located in Galveston, Texas. Founded in 1891 as the country's first unified public medical school and hospital, UTMB has a long history of being forward-thinking, and its faculty and staff share that mission.

Peter McCaffrey, MD, serves as Director of Pathology Informatics, Director of the newly-founded Division of Bioinformatics and Artificial Intelligence, and Director of Laboratory Information Services as well as Assistant Professor of Pathology. While his role began in 2018 with a focus on pathology, it soon became clear that UTMB's radiology department was facing many of the same challenges as the department of pathology. Dr. McCaffrey and

others wanted to realize greater value from UTMB's imaging assets, and collaborate more efficiently inside and outside the system's walls. As a result, Dr. McCaffrey's role shifted to leading the informatics effort in both departments, helping shape UTMB's approach to bioinformatics and artificial intelligence.

When the COVID-19 pandemic hit, the need to collaborate efficiently and remotely on imaging research became more urgent than ever. As Dr. McCaffrey sought tools that would help the hospital share chest x-rays, a colleague within the UT system introduced him to Flywheel.

"I looked at it and I said, 'This quite directly solves the problem we face,'" he recalls. "The problem is always infrastructure—how we're going to put this data in a place we can share it and organize work around the assets. It became clear to me that we probably needed this for both path and radiology."

## A “Living Repository” Fuels Research *within UT*

Dr. McCaffrey explains the traditional challenge of collaboration: “If we have a bunch of chest x-rays, and I want to work with someone at UT El Paso, how do I give them that data set? I don’t want to just send the literal files into their OneDrive, which was the way of doing it,” he says. “So the original appeal to me was, here’s a single place we can put this information. Everyone comes there to fetch it, rather than me sending everyone data. I don’t have to update it. It’s just a living repository.”

That repository is now fueling UTMB’s COVID research, which is a collaboration between the health system and UT El Paso (UTEP). When a provider at UTMB marks an x-ray in Epic as a suspected COVID case, the file is automatically uploaded to Flywheel. The image repository there can be accessed by computer science researchers at UTEP in their work developing segmentation algorithms. As the data set grows, the researchers are able to easily re-run and refine their AI model, and Flywheel tracks extensive details on all processing to support reproducibility. The platform is also being used for research on traumatic brain injury (TBI).

## Building Compelling, Curated Data Sets *Ready for Sharing*

While Flywheel initially appealed to Dr. McCaffrey as a solution for image conveyance, he quickly realized it could help address a larger strategic challenge he faced in his role. As UTMB went live with digital pathology, and as its radiology archive grew, Dr. McCaffrey aimed to more fully leverage the health system’s imaging assets.

As he recalls thinking, “This is a million gigabytes of data that we’re going to be creating, at significant cost. What are we going to do with this? How do we make it valuable?”

Dr. McCaffrey recognized the unique position the hospital was in to create data sets and make them available for AI researchers. “We feel our special capability as a hospital is not really to build models,” he says, “but to make data sets that are useful for building models. And to make data sets that aren’t just big, but that have complicated things, multimodal things, things that have nuanced clinical phenotypes.”

Flywheel’s ability to support annotation and image analysis “in the loop” is now enabling UTMB to create these AI-ready data sets in both pathology and radiology. “It allows

us to coordinate efforts—not just around fetching images, but also looking at images and annotating them, putting a bounding box around inflammation and tagging it, for example. This work adds structured metadata to images and adds value and utility to those assets. That value in annotation is the product of work by expert clinicians and the resulting datasets carry a higher value both implicitly and explicitly. That’s really useful because when you’re trying to do AI work and to obtain the financial support needed to do that work, you really are searching for a rational business case tied to some base revenue that you can quantify,” he says.

He notes that when the health system creates an image, it has already invested in that image, through its equipment and the provision of care. “And so now we’re just doing a marginal extension and adding value to those data sets by creating more annotations and metadata about them. But without Flywheel, there’s not a clear way to organize that process.”

As he summarizes, Flywheel provides “a full structure around this activity of curation, which is really the bread and butter of what our AI group is doing—even more so than creating any actual models. We want to make the data set so we can work with others and compel them to work with us to create their models.”

In addition to the revenue-generating potential of these data sets, Dr. McCaffrey points to several other ways that Flywheel is providing a return on investment. “One is just not having a complicated internal infrastructure,” he says. “We can run our entire image organization and image brokering activities with far fewer people than if we were doing it any other way.” He also points out how Flywheel has streamlined grant applications by simplifying the processes of creating registries and working with collaborators, “which of course is much easier to do with a single tool.”

Moving forward, Dr. McCaffrey sees Flywheel as a large part of his tech stack, a position that isn’t easily earned. “I want the most parsimonious set of tools possible,” he says. “Then the onus is on us to build the last-mile features on top of them. So now, if it’s imaging research, we’re doing it in Flywheel. It’s clear to me that it’s a broad service, a robust service, one that meets path and radiology needs, and one that also handles interoperability questions.

“If it’s imaging research,  
we’re doing it in Flywheel.”

So I’m embracing it and centralizing it more, and turning off things that conflict with it.”

# Solving Pervasive *Collaboration* *Challenges*

UTMB's work with UTEP has already provided a proof of concept for the potential of leveraging the hospital's data in AI research. Dr. McCaffrey explains that "one of the holy grails for hospitals is, 'How do we work with other people who are AI experts and give them access to work on our data? And the corollary to that for an AI person is, 'How do I get access to interesting healthcare data?' So we see Flywheel as our solution for this."

Dr. McCaffrey notes that Flywheel's de-identification makes it easy to share data in a compliant way with UTEP's computer scientists. "They're building computational models by using batches of data that we curate as a hospital, whereas we don't have any computer science PhDs. That's the model we want to continue to implement."

Flywheel's tools for access and provenance have been extremely well received by the academic researchers. "They've been very delighted with how this has worked," he says. "When turnover happens with postdocs or PhD students, all we do is just modify the access controls in the projects. New people are then granted access to it; they're able to use the

## Flywheel's ROI *at UTMB*

Cost Reduction	Revenue Generation
Staffing efficiency	Streamlined grant submissions
Simplified infrastructure	AI-ready data sets
Reduced IT costs	Ability to collaborate with commercial innovators

Flywheel CLI and the pipeline SDK to pull the data in their modeling projects even as a direct API call. From what we understand this is an unprecedentedly easy way for them to get access to healthcare data. The alternative is just to initiate an incredibly complex IT project every single time."

Dr. McCaffrey sees potential for Flywheel to serve as a platform for future academic and clinical partnerships beyond specific studies. He envisions using it for a UT system colloquium, and also for collaborations like "AI meetups" with other universities.

Beyond purely academic collaborations, Dr. McCaffrey also sees great potential—and low competition—for working with commercial AI entities. As a former biodesign fellow at Texas Medical Center, he has seen firsthand the gaps between healthcare institutions and the innovation ecosystem.

“You can walk in the door at a major research hospital and it’ll be three years before you have a data collaboration,” he says. “So we’ve identified this as a great opportunity to say to innovators, ‘We have a way to work with you.’ Having the ability to create data sets with PHI protection and access controls solves so many of these really pervasive problems.”

“We’ve identified this as a great opportunity to say to innovators, ‘We have a way to work with you.’”

Visit [flywheel.io](https://flywheel.io)  
for more on how Flywheel is  
enabling research in high gear.