

- Adrian H.: [00:04](#) Hey. This is Adrian Hernandez, and welcome to the NIH Collaboratory Grand Rounds podcast. We're here to give you some extra time with our speaker, and ask them the tough and interesting questions you want to hear most. If you haven't already, we hope you'll watch the full Grand Rounds webinar recording to learn more. All of our Grand Rounds content can be found at rethinkingclinicaltrials.org. Thanks for joining.
- Adrian H.: [00:30](#) Hi. This is Adrian Hernandez, and I am a host for the Collaboratory Grand Rounds. And today I'm really pleased to have Rebecca Lee and Frank Rockhold, who just did a Collaboratory Grand Rounds on preparing for clinical trial data sharing and reuse, the new reality for researchers. Frank and Rebecca, thanks for joining us in this podcast.
- Frank Rockhold: [00:52](#) Glad to be here.
- Rebecca Lee: [00:52](#) Thank you, Adrian.
- Adrian H.: [00:53](#) Frank, let me start off with you. People have been talking about data sharing for, really, years, and I recall that you were one of the early champions for this. What's the problem we're trying to solve here?
- Frank Rockhold: [01:08](#) I think there's a high level issue around transparency and openness in science. I think we have a credibility problem in science in general and in clinical research in particular, but this has been going on for a number of years. If you think back to even protocol registration back in the '90s and the 2000s, and clinicaltrials.gov requiring protocol registration was one step on the way to try and create transparency. One way to help transparency or foster it is to make information available about research. That includes the design of the research protocol, information about the data that you've collected, so-called metadata, and then the data themselves to allow people to further investigate research findings. And if they choose to try and understand how you arrived at your conclusion, which then is all part of a greater problem around reproducibility. But all of those are pieces of a puzzle of transparency. They are actions to help achieve openness and transparency.
- Adrian H.: [02:27](#) Everyone agrees that open science and data sharing is really important. People believe in the principles here. And Rebecca, you have been really a champion in developing this, and you now have a platform that you're leading with Vivly to make this a reality. How does that work, and how has that been going?

Rebecca Lee: [02:47](#) At the core of my training, I'm a scientist, so to be part of the team that was all about sharing is really exciting to me. I was part of the original team at the MRCT Center of Harvard and Brigham and Women's that launched Vivly as a nonprofit, and Frank was there as well. And my career has been centered around developing drugs primarily, before then in the pharma and biotech industry at a CRO. And back then, I wish there was something like Vivly back then when as a scientist it would have made my work easier. As data sharing really allows scientists to see what clinical trials have been done before reducing duplication, and allows those trials in the planning stages to be reshaped as necessary. And data from clinical trials have really been locked away. Some in, as you probably know, Adrian, in some PIs personal computer, or a data bank, or maybe in a university's computer system. Others might be in a pharmaceutical company's personal cloud.

Rebecca Lee: [03:57](#) But for years we've really been talking about now the concept of big data, and as we think about these truly complex diseases in cardiovascular disease, diabetes, cancer, Alzheimer's, to really unlock those diseases, we really have to think about opening up those vaults and freeing those data. And the key's really turned now. Data sharing is the new reality, and many players have, individuals have stepped forward. But now major stakeholders like Duke, pharmaceutical companies, nonprofits have started to step forward. And now, well, some of the world's leading medical journals have stepped forward to require authors to have a data sharing plan. And we really hope that researchers hoping to publish in these journals will declare their data sharing plans in [ct.gov](#), and that will be a major step towards more sharing of individual participant-level data.

Adrian H.: [05:00](#) Let me ask... In another area around data sharing is that there've been a lot of worries about potential mishaps with data sharing, and a lot of concerns about so-called feeding research parasites. Frank, I'll start with you. Is that a concern about any potential harms around data sharing, and does that just enable research parasites?

Frank Rockhold: [05:34](#) Yes, Adrian. That has certainly been a concern from the very beginning, and there are two, at least two aspects to that that I could highlight. One is the concern that people will reanalyze their dataset. Someone else will reanalyze their dataset and get a different finding, and either publish that and either discredit them or get into a public debate about whose analysis is correct. And so I think that is a concern of many, that it will upset some other work they've done. To this date, that's not

been a big problem. People have been actively sharing large datasets now for six years, six or seven years, and I'd say there's maybe three or four or five cases where somebody has actually taken the time to reanalyze an individual dataset, and maybe one or two cases where there has been some debate. But on the whole, 95, 98 percent-plus of data reuse applications have really been to try and understand, answer new questions, and progress new knowledge. So I think that I would say on balance that fear has really been unfounded.

Frank Rockhold: [06:51](#) The second concern, which I think speaks to the research parasite part, is if I spend six or seven years running a clinical trial and I release the data, is someone else going to get to publish a number of papers and go get a grant that rightfully should have been mine? Because I should keep the data. Because I, and I use this term loosely, but I own the data, and I put that in quotes. I think that's the concern. It's certainly been highlighted around some of the ICMJE debates. And I don't know, it might be too early days to know whether that's a reality or not. My view is that people can include that in their plans when they design a trial. If they know they're going to be sharing the data, then they plan ahead of time for what questions they think are legitimate.

Frank Rockhold: [07:46](#) And I think as has been published by some noteworthy authors, the key at the end is to hope for collaboration. So if somebody has a question they want to answer using your data, that you open up the data to them, but encourage a collaborative approach as opposed to a competitive approach. So, too early days to know whether that's a reality, but I would say as an editorial comment, patients expect their data to be shared, so I don't think academics or anybody having an argument about who should or shouldn't be answering questions from my data should ever be more important than what the expectations of the patients are.

Adrian H.: [08:31](#) That is very true. And I'll just go on the record and say I am a research parasite. I have used other data and published from it. And also agree fully with the notion that patients' expectations are for, if they're volunteering for something, that they want as much knowledge to come out from that. Rebecca, as things have developed for Vivly and the platform here, how are you intending to have incentives for its use. Certainly, making the data available is an important aspect. Are there any plans in the future to help recognize those who donate or share data on the platform, or even use it from the platform?

- Rebecca Lee: [09:25](#) Absolutely. Those ideas have really been a core part of our development as Vivly has evolved and as we've created the platform. Digital object identifiers and how Vivly has been structured is already baked in to the development in how Vivly is structured. DOIs are digital object identifiers. This is already a part of the platform. You can already see how incentives can be utilized to track and provide credit for those that contribute data to Vivly, and also for those that utilize Vivly. You can see for collaborators how they can also credit those that have donated the data by using these DOIs, these enduring links to the secondary datasets. I think those are really things that we as a platform can utilize, and we believe that is clearly important.
- Adrian H.: [10:41](#) That's terrific. Well, last question for both of you. Right now, I'm just going to hazard to guess, but you can correct me, that it's a relatively small percentage of clinical trial data that's shared now. In five years, what do you estimate the percentage will be?
- Frank Rockhold: [11:03](#) I would say, if you think... And let's go back to 2003 and '04, when we had the debate about registering protocols and the initial uptake was slow. Of course it had the force of the journal editors and then ultimately the US Congress behind it, and so it did take off relatively rapidly. And I think in spite of the early concern that being forced to register protocols was going to slow the advance of science, that certainly didn't occur. I think if you then move forward to the individual patient data sharing, it is a lot more complicated and it is a lot more costly, so the uptake is naturally going to be slower. But my prediction is that once we get to a place where 25 to 30 percent of people are routinely sharing data, what will happen is that the population of people who want data and the population of people who are willing to share data will start to coincide.
- Frank Rockhold: [12:07](#) And I think that will, once we hit that critical mass, it will become more common. I think it will always be an issue of enabling and funding. I do think there are some practical issues that are going to require some leadership in the academic and government community to make it clear when you're planning a research project, allowing funds and logistics for sharing are important. And I think Gates Foundation and NIH, for instance, are already doing that. I think once that takes hold, that sort of a nudge will increase the uptake.
- Adrian H.: [12:47](#) Okay, great. And Rebecca, any predictions for five years in terms of the state of affairs for data sharing?

- Rebecca Lee: [12:56](#) Yeah. I mean, what I've seen just in the last 18 months, I would say, when we've been very much engaged, is we've seen individual stakeholders take bold steps to foster in data sharing. But now we've seen more a broader range of stakeholders move together to build this ecosystem, to both reuse data and step forward to contribute data. So I would say give or take maybe another three years for us to see kind of this inflection point, where we would see the majority of data shared in the IPD space. Maybe that's a bold projection, but hopefully within three years' time, that we would see the vast majority of data in the clinical trial space, and hopefully other types of data like real world data also shared. Which I think would cause some rapid advances in some of these diseases that impact all of patient lives, which is really where we're all... Of course, the mission of Vivly, Duke, and others will become the reality of our mission, will be closer to a reality, and hopefully fulfilling that reality as well.
- Adrian H.: [14:22](#) Well, great. Well, I look forward to achieving those lofty goals. And Frank and Rebecca, thanks for spending time with us on this podcast here, and really appreciate all your efforts in terms of driving data sharing and open science for the community.
- Rebecca Lee: [14:43](#) Thank you.
- Frank Rockhold: [14:44](#) Thank you for the opportunity, Adrian.
- Adrian H.: [14:47](#) Please join us for our next podcast as we continue to highlight fascinating, informative changes in the research world, lessons learned from Collaboratory Grand Rounds as we go forward. Thank you.
- Adrian H.: [14:58](#) Thanks for joining today's NIH Collaboratory Grand Rounds podcast. Let us know what you think by rating this interview on our website, and we hope to see you again on our next Grand Rounds, Fridays at 1:00 PM Eastern time.