# GRDC PODCAST TRANSCRIPT

**Data Partnerships Initiative**

[00:00:05] **Introduction** This is a GRDC podcast.

[00:00:12] **Sally Maguire** A virtual goldmine of data generated from thousands of research, development and extension projects, all supported by GRDC, will soon be discoverable, accessible and able to be used to accelerate research outcomes. Hello, I'm Sally Maguire. GRDC Data Partnerships Initiative is a $2.8 million investment over 18 months, and it brings together 12 Australian research organisations that are currently working on projects in which GRDC has invested. Decades of data has been collected on a range of topics, including yield, soils, genetics and climate. And this new initiative will collate existing data as well as provide a roadmap for improving data management in the future. Agriculture Victoria Crop Agronomy Research Leader, Associate Professor Glenn Fitzgerald is managing his department's involvement in the initiative, and he joins me now to explain more.

[00:01:12] **Associate Professor Glenn Fitzgerald** We're working with GRDC to collect data as part of our usual co-funded projects in science across all the different types of science we do and collecting measurements, recording data is always the first step in doing good science. So, as part of this initiative, what we're doing in particular is collecting the metadata, which is actually data that describes the data we collect. And that's key to understanding the data we do have as this allows us to know what types of data we are collecting in order to understand better all the data that's there to support our science.

[00:01:49] **Sally Maguire** So I guess then what is the nature of the data from your organisation that we're talking about?

[00:01:55] **Associate Professor Glenn Fitzgerald** Really what we're talking about is all that data were collected which can be from field laboratory experiments, could be growers paddocks, data from sensors, satellites, and it can span the range from genomics to crops, soils, animals and that goes from what we say, from genes to landscape. So it covers a whole range of all the science that we do. The data itself also covers a gamut from data that's fully open and accessible all the way to commercial and confidence type data, where there are different levels of privacy and restrictions for access depending on the data sensitivity. So with GRDC, it really covers the gamut of all the kind of data we click through the science we do.

[00:02:36] **Sally Maguire** Individual organisations like Agriculture Victoria have been maintaining and managing data collection in all sorts of ways over the years. So how will this initiative change that?

[00:02:47] **Associate Professor Glenn Fitzgerald** Good question. And one activity that this initiative is about was to document the key projects that we have had in the past decade or so with GRDC and that have been co-funded by both organisations. And what we've done there is to build the metadata descriptions for the types of data collected. And the second part of this looks forward rather than past and the legacy data to create new ways of collecting the metadata to document all the projects that we will have starting now into the future and begin setting up new ways to collect and store the data for future reusability. We've done that collection of legacy data, and then we're looking forward into how we might actually change the way we do our data collection. Working with GRDC then they've created what they're calling a metadata catalogue. So they want to be able to document clearly their investment and of course AVR does as well. What that does is collects sort of high level metadata around all the projects that they have funded over the years, and that is in the process right now actually being uploaded to their metadata catalogue. So they have a good sense of the breadth and the range of data that they have collected through these projects over the year, not just with AVR but across the entire country. So it gives that potential actually of because it's going to be public facing, other researchers then will be able to look at this range of data that has been funded by GRDC and start to use it to think about, well, how could we look across data and different projects and actually build something better and bigger than just an individual project?

[00:04:22] **Sally Maguire** And I think you kind of just referenced this then. But you know, as a researcher, what do you see as being the short term and the long-term benefits of having this sort of data collection?

[00:04:32] **Associate Professor Glenn Fitzgerald** I think the benefit of the partnership is including the development of tools that allow organisations, including AVR, to document the projects, like I said, that are with GRDC, you kind of have to start there. If you don't know what you've done and what you have, obviously you can't move forward. So it's a really good opportunity to do that and in the longer term it will create legacy systems to be able to manage and govern their data better and create data that's called more 'FAIR', which is an acronym meaning findable, accessible, interoperable and reusable. So that's a description that we have around metadata. And this project has really been focussed on the F and the A in that, in other words, findable and accessible. So can we find the data through this metadata and how accessible is it? What are the levels of privacy, etc.? Is it open, is it closed and whatever? But basically, where is it and can we access it? And what that does is it sets us up for the future for that next stage of what's called interoperable and reusable. So can we use the data to answer bigger and better questions, as it were, that go beyond just the funded projects that have been done in the past, which is typical funding cycle.

[00:05:40] **Sally Maguire** So when you improve your data collection, then that will flow through to better research methodology, maybe quicker outcomes, that sort of thing.

[00:05:49] **Associate Professor Glenn Fitzgerald** That's right. Obviously, science is based on data, but it's also the first step really in the science process. You start with your data, you build knowledge and information, and so scientific principles and analysis which allows then to answer those questions that are relevant to growers and to the broader industry. So in the past we've gone, as I've said, project by project, but if we can look across those projects and look at the bigger pictures - can we answer some of the bigger questions that are becoming more and more important in this very complex age of climate change and the environment and sustainability and all the rest of it?

[00:06:27] **Sally Maguire** Joining me now from the University of Adelaide is Alexis Tindall, who is manager of digital stewardship for the University Library. So, Alexis, tell us about your role and how you work with the data that's collected.

[00:06:40] **Alex Tindall** I'm the manager of digital stewardship at the University of Adelaide Library, and the library supports all of our research community in various ways, including helping them look after their data and make the best use of it and meet the best practices for their field and for the research community widely. We a host data repository and we make it possible for people to expose their data online, which makes it more discoverable. It gives people another avenue into the research that we're creating, and it ensures that we're being transparent and accountable in the work that we're doing. So we advise the researchers on how to look after their data well, and how to make the best use of it. Then we also play a role in taking care of that data. In certain circumstances that needs to be disposed of or destroyed at a certain point, particularly when it's something that is sensitive so we can take care of those matters for them as well. We have an academic liaison team that supports researchers in many ways, including planning their projects, advice and publishing, ensuring that their research is having the widest impact. And we work hand in hand with researchers from all disciplines along that journey.

[00:07:47] **Sally Maguire** So this Data Partnerships Initiative is focusing on data discovery and organisational alignment. What exactly does that mean?

[00:07:54] **Alex Tindall** We have partnered with GRDC on projects over many years and so we actually have collected quite a lot of data from these projects. Over the last ten years or more, the research community has recognised the value of keeping that data, making it discoverable and making it available for re-use. So with this project, we've been looking back at past research projects, making sure that we're looking after that data well, but also making it more discoverable for other people. So that means it's available for re-use if it has potential to expand the scope of another project or we're making sure that we know what we've done already. So we're reducing duplication with covering the same ground helps us be more focussed and make the best value out of the investments that we make in this area. In terms of organisational alignment, the library works with many research disciplines and over the last number of years we are recognising that best practice in terms of data is making it 'FAIR', which is findable, accessible, interoperable and reusable. Different disciplines are at different stages upon that journey and some are really engaged with data sharing, really engaged with making their data discoverable, really publishing their data well. Some are just starting on this pathway. So in terms of organisational alignment activities, the GRDC has resourced us to be able to focus on the agricultural research community and look at how they can apply best practice in their field, what the opportunities are by thinking about 'Fair' in terms of managing their data and making it discoverable. What the challenges and barriers might be and ensuring that they're meeting the expectations of both GRDC as the body that's investing in this research, but also the broader agricultural and other research communities.

[00:09:44] **Sally Maguire** So I understand that the data discovery work involves the locating and the cataloguing, but it could go back sort of ten years. So how big a job is this going to be? And tell us a little bit about your methodology.

[00:09:56] **Alex Tindall** This process is looking at data that goes back ten years. And so more recently, the university supports our researchers with shared network drives with a data repository with standards for publishing that data. But really, that's something that's only come into play over the last five years and really is only been commonly used over the last two or three years in various disciplines. Before that, many researchers were managing their own data on their own computers, on portable hard drives, and some data that is not even in digital form. We have had the opportunity through this process to look back at those collections and make sure we're backing them up and looking after them. According to the community expectations, helping those researchers who are managing some of those things themselves and looking at how the institution can actually support the work by keeping that data in appropriate locations. In terms of methodology, it's been a lot of legwork, so we've been lucky to have a data discovery person who's been reaching out to researchers that we know have led these GRDC investments. And we've also been doing a process of looking back through the publications, the academic journal publications we've made in agricultural research and seeing if we can find further projects that have acknowledged the investment of GRDC that may not have been on our radar. In those circumstances, it's a matter of getting in touch with the researcher and ensuring that they have that data available, talking to them about this opportunity to make it discoverable, working with them on the challenges around that, so whether the data is in a form that it is easy or safe to share, ensuring that any sensitivities in those data are protected and ensuring that we are working in the best interests of everybody who's collaborated with us on that project. In many cases it's about writing metadata for the project, and metadata is information about data. So at the simplest level, it's like what you use to find a book, who created it with the author is what the title is, any relationships it might have and preparing a brief description about it. But there are other pieces of information we can add to that metadata, including where the trial took place when the research took place. That can make it easier for a future user to work out whether they want to access that data or not, or whether it's going to be useful to them.

[00:12:20] **Sally Maguire** That's great. So ultimately, what do you see as being the benefits of this project?

[00:12:25] **Alex Tindall** I see extraordinary benefits in thinking about data in this way. Simply put, we are very lucky to be able to work with the limited resources that we have that exist in the research sector. And so we want to make the best use of those things that we can. We're being transparent and accountable when maximising the value of that investment. So ensuring that if there is a possible future use of this data, that we have it available and ready for that future use, which could then either expand other projects or ensure that a better focus, we're making sure that we're available for checking to ensure that our research partners in the community can ensure we can back up what it is that we say. But I also see this as a really useful way to face the future challenges in our sector, you know, as we have novel technologies emerging, things like machine learning, things like gathering data, using sensors and drones, things like artificial intelligence, the data is only going to become larger and more complex, but also the tools that we can use to access that become more powerful. And so some of these foundations we are laying and good data management and good data discoverability will help us be prepared for that future. In terms of the specific benefits of this investment, it has helped us take that time and do that close work with those researchers to help them understand best practice in this area to see the opportunities that it provides for their particular type of research and to overcome any myths or any fears or any concerns that this may raise and to really dig down on what this kind of practice means in agricultural research. Because while we are working to widely accepted principles, each research discipline that we apply them to throws up something new challenges and it's worth taking some time to think through that we don't always have that opportunity, we service a very wide range of research disciplines and so it's been a real blessing to be able to take this time to work closely with the agricultural research community on this.

[00:14:30] **Sally Maguire** That was Alexis Tindall, manager of digital stewardship at the University of Adelaide Library. And earlier I spoke to Crop Agronomy Research Leader, Associate Professor Glenn Fitzgerald from Agriculture Victoria to find out more information about GRDC's Data Partnership initiative, you can visit their website, grdc.com.au. I'm Sally Maguire, this has been a GRDC podcast. Thanks for listening.