Thanks for letting me give a quick overview of the place Trove has in the Australian Institutional repository landscape.

Let's get going with some quick facts -
Trove has four basic functions

- It’s an aggregator of collections from research, cultural heritage and community organisations.
- It’s a digital collections host – most known for digitised newspapers but more recently expanded into hosting other types of digitised and born digital content.
- Platform to research and build – some basic tools in the interface and a number of APIs to make the data freely available.
- Community end users. They’re the ones who do the research, make connections, correct and enrich the data for others.

The reason I’m here is to talk about that role as an Australian aggregator of metadata.
There are different ways to get metadata into Trove. Some are bulk transfers like the tram, and some are steadier streams like the pedestrians:

MARC records via Libraries Australia.

Select databases hosted by vendors GALE and Informit.

A few foreign fulltext repositories including Hathi Trust and Open Library.

Direct-to-Trove – this is harvesting from institutional repositories and data stores. It grew out of predecessor aggregators services like Australian Research Online, Picture Australia and Music Australia – subject based aggregators.

In the eight years since those separate services started to disappear Aggregation looks completely different - much wider group of organisations we work with, wider spectrum of collections, very different back room processes.

Trove harvests 400 collections from Australian repositories, representing about 250 organisations, in this Direct-to-Trove pathway.
In the university sector, the institutional repositories give us lots of different collections. Every Australian university contributes something. What they contribute varies from uni to uni. Could be:

theses
datasets
published academic output including research papers, journal articles, book chapters and pre-prints
coursework
open access books they publish
open journals they're hosting.

Might get other bits of data not directly from the IR but managed by the same team in some way.
Example of that include:

Data set descriptions that come via ANDS RDA.

Biographical records. In the uni sector we try to source biographies from ORCID.

Also get a smattering of small RedBox MINT collections of identities or names.
It might seem like Institutional Repositories are the only direct-to-Trove metadata that is coming from Universities.

In reality other parts of universities are contributing to Trove in their own right.
Examples include:

- museum archaeology and specimen type teaching collections
- subject specific repositories with working papers
- university archives
- heritage collections managed by the Library or someone else
- research collections like AusStage.

Overall, universities contribute 1.9 million works via the Direct-to-Trove harvesting pathway.
This is the breakdown of the types of content that come from universities repositories, based on the Trove zone it ends up in:
The vast majority – 84%, is in our journals, articles and data sets zone. 6% ends up in the books zone, of that 6% theses are about half. 5% in the pictures zone. 4% in the diaries, archives and letters zone. 1% in music. And less than 1% are biographical or people records.

That gives you an idea of the kind of material university IRs are putting into Trove –1.9 million works, the bulk of it is articles.
These are the kinds of stats that are pretty easy to get from Trove:

You can do other kinds of statistics, depending on what you’re interested in. Eg. This is the breakdown by state, instead of item type -

On the left is the number of items in Trove, based on which state the University is in. On the right is the breakdown, by state, of theses. It can highlight things like NSW gives 20% of all works on the left, but 27% of all theses, on the right.

Part of Trove’s value is this ability to pull sector-level statistics.
When people think of Trove, their most immediate thoughts around value are some of the very basic things, like:

Large # of users per day, particularly Australians.
It’s a single place to search all Australian repositories.
These are valuable as a single user doing single searches.
Trove also has value as a more sophisticated tool.
Has the potential to do comparisons and produce statistics at a National level, or a sectoral level on the aggregated data from Australian repositories.
The slides before on how many works are contributed, what states they come from, and what type of content it is are all based on freely available Trove data.

There's a lot more we could be doing, and aren’t, part of that is resourcing. Part of that is because of variability in the data we get from repositories.
I think there’s a perception that everyone starts out on the same starting line, that repository data is quite homogenous across sites. We’ve got lots of great standards, we must all be sharing the same data.

We’re not.

ORCIDs, DOIs, access and licence indicators and so on are great for linked data. They’re being implemented at most university repositories.
Often not added to data output feeds that come to services like Trove. So even though the data might be quite rich locally within the repository, it can be quite basic in Trove.

No longer is everyone at the same starting line.

Even basic data has to be reviewed by the Trove team because some sites differ on which fields they use, and we want to provide a consistent user experience.
Here’s a list of the kind of work the Trove team does on every collection, every time there is a system change, to ensure we’re getting standardish data across repositories.

It’s a big and ongoing commitment.
In terms of finding University resources in Trove, you can do a search limited to records coming from University repositories.

There’s a unique collection identifier called a NUC symbol that’s used in Trove. Differentiates each collection including the repository.

Which means it’s possible to construct a search just limited to records from Australian Institutional Repositories.

This particular search is for items from NSW and Qld university institutional repositories.
Aggregated metadata is not the only way universities are getting resources into Trove. IRs are also doing some collaborative digitisation with Trove. Items like back issues of student newspapers - Tharunka and Woroni – as well as resources to service research communities like the early 20th century Australian art magazine – Art in Australia. Keep in mind that digitisation partnerships are another avenue that’s being used to get data into Trove.
There are a few Trove APIs where people can pull back the data:
An older OAI-PMH feed for Australian theses.
Trove API, which people like ExLibris are using to pull out a similar data set of theses, but de-duplicated, and there's everything else from Trove in there as well.
And a separate set of APIs to draw back People data.
So I hope that’s given you a decent overview.
Lots of paths for data to come into Trove.
Lots of university collections being added with the IR as the jewel in the crown.
Big variation in the way data comes to Trove.
Normalisation is a significant activity undertaken at the Trove end.
Opportunities for national statistical reporting based on Trove data.
Ability for anyone to pull data back in a machine-readable form.