

## Off the Mark - Unfortunate Trends in footyland

A few days ago an [AFL.COM.AU](https://afl.com.au) article was published based on data generated by Stats Insider, rather than Champion Data. Champion Data is the official data provider for the AFL, while Stats Insider has only just started recording similar match data by hand.

When I first saw it, I was, to be honest, a bit dumbfounded. Surely it would have been easier - and, let's be honest, more reliable - to use the official source of the numbers? These thoughts proved to be more popular (at least through twitter) than the AFL's official posting



**Harvey**

@roberto\_boberto



Hey imagine that, [@StatsInsider](#) provides freely available shot data and [@AFLcomau](#) writes an article using it? Why not use official data?

- Is the process 2 long?
- Is the data too hard to access?

Is this really the situation Aus biggest sport is now in?



**AFL**  @AFL · Aug 2

Who's the best shot from outside 50 in the league? [afl.com.au/news/2019-08-0...](https://afl.com.au/news/2019-08-0...)

11:27 PM · Aug 2, 2019 · [Twitter Web App](#)

 View Tweet activity

**8** Retweets **53** Likes



AFL   
@AFL



Who's the best shot from outside 50 in the league?



Not Buddy or JK? You won't guess who's the best shot outside 50  
Who do want to take the shot on goal from outside 50? It's probably not who  
you think  
[afl.com.au](https://afl.com.au)

3:04 PM · Aug 2, 2019 · [TweetDeck](#)

4 Retweets 19 Likes



**Marc McGowan** @ByMarcMcGowan · Aug 3

Replying to @roberto\_boberto @StatsInsider and @AFLcomau

Hi Harvey, this information is available for journalists from CD if requested. It's not a long process. The journo here, @JourdanCanil, simply used Stats Insider for this article and it was a good read. No conspiracy, nothing more to be made out of it.



1



2



**Harvey** @roberto\_boberto · Aug 3

if requested is the issue



1



**Marc McGowan** @ByMarcMcGowan · Aug 3

It's not a 'right' to have every bit of data. CD don't need me fighting its fight but there are people who work incredibly hard to gather this data, and it's a business. CD is on record saying it would support a media business buying the data and making it available.



**Marc McGowan**  
@ByMarcMcGowan

Replying to @TheArcFooty @roberto\_boberto and @SquiggleAFL

My podcast isn't purely about stats and where it is, it refers to how the clubs are using stats. The clubs deserve the data more than anyone. The media pays a large sum for it.

11:10 AM · Aug 5, 2019 · Twitter for iPhone

Let me say my part - and, of course people will chime in - but how can people clear up misconceptions without putting their thoughts out there.

Firstly let me point out a misconception here, I don't expect fans to be provided with all the data. I merely think that things like that get written about online like Jourdans using shot data (only kicked or missed) should be available for public consumption.

If an AFL journalist has an idea, requesting the data to do the article when it's really just a simple data cut isn't the situation AFL fans and journalists should be in. If Jourdan has an idea on shot accuracy from outside 50 (nothing around pressure, game state when kicking, distance they ran before the kick, etc) just a simple goal/missed should be able to present and write the article ASAP.

What the process of using Stats Insider indicated to me is that it was simply easier to go to Stats Insider rather than wait for an official person from within the walls of CD. This was confirmed by talking to people within the industry - with them noting that not only would there be a wait, but that the visual wouldn't have been provided, but rather only the summaries.

What fans are able to get from the box score is only the total summary at the end of the game. Not when the goal was scored, if the kicked missed everything or anything around the "pressure" of the kick when it takes place. So fans are left on the far left of the data access spectrum. This article and the data in it represents a happy middle ground. No information around pressure of the shot was given, no information about the speed of the player, the amount of defenders around the player when the shot was taken is given. Just in my opinion middle ground.

If we are to believe Marc - and I do - the data is available to journalists on request. But why can't it just be available?

All this misses really an answer to the question I had originally, which Marc has continuously refused to answer.





**Harvey** @roberto\_boberto · 2h

but why as a journo enter the debate without actually offering an answer to original question. Why would an internal journo decide to use outside data



1



**Ryan Buckland**  
@RyanBuckland7

Replying to @roberto\_boberto @CBCe8enn and 4 others

because it's:

- a) better
- b) easier to use/access
- c) interesting
- d) some combination of the above

2:02 PM · Aug 6, 2019 · [Twitter Web App](#)

That's the situation we are in, with only one provider of numbers for everything from a simple shot plot to the player ratings based on possession chains. I have also been informed, that CD wouldn't have actually provided what I think is the best part of the article (the visual).

Let's acknowledge that before we move on again, no one is saying that all the data should be provided. But if you write a piece online you must find it interesting and if you are paid for providing content at the very least you must think that other people find it interesting.

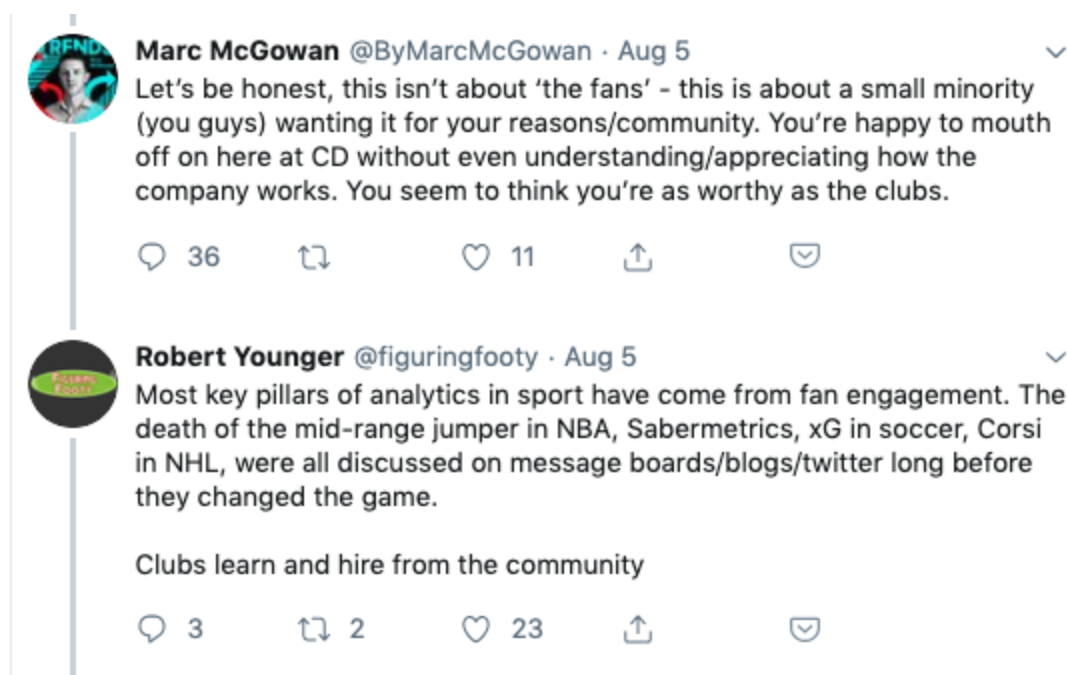
Take Marcs article here on Brisbane Lions

<https://www.afl.com.au/news/2019-07-30/stats-files-why-the-lions-might-be-the-afls-most-versatile-team>

No one thinks that this data is particularly advanced, or even the way its being used is particularly state of the art. But what if you liked the premise of the article but you are not a Brisbane fan, you can't investigate your team in the way that Marc did in his article. Lets say your a Brisbane fan, instead of the team summary you might be wondering does my favourite player change the way he plays using the metrics Marc has used. You can't do this. Good analytical pieces should promote debate among fans and analysts. If I was sitting down with my mate and mentioned that there's a cool article on Brisbane showing that how they

transition the ball changes week to week, her first question is probably whether the Eagles also do that. Unfortunately, because the data is locked up, and the journalist who used it didn't decide to cover that part, that's it - that's the conversation ending. What should be a jumping-off point instead becomes the end of the road. Is this the situation we want to be in?

Champion Data (CD) do great work - let me state that point blank. Let's take one of the more popular metrics used, the Player Ratings. Did you know: They came from [Dr Karl Jackson's PhD](#). Tim Bedin, one of their Data Scientists, has a masters in physics. The people crunching the numbers are highly qualified and do great work. My question is why we must limit the research and understanding of football to only a select few.



I want to address this idea that there is no money in making this data available for fans.

The national football league - yep, the American one, with the SuperBowl - recently ran a [big "data bowl."](#) whereby they allowed ANYONE access to player tracking data.

When I interviewed [Michael Lopez](#) on my podcast, "Chilling with Charlie," one of the points he made was that people at clubs were unsure as to the full value of this data - they themselves want to see what people could come up with, or would approach problems facing the game.

This competition was won not by insiders, but outsiders: university students! That's right, folks, they entered and they won! They were able to showcase their talents to the league and to clubs and many were then able to get employment within clubland.

The data providers didn't go under because a subset of data was released. I'd argue they may even be better off, because clubs were able to see the value in the data - they wanted to try out some of the presentations themselves and they were able to invest in people that have shown they have the skills.

Would there be less pressure on the AFL around scoring and congestion? Should important questions like this only be addressable by a small subset of people? Can the best possible solutions only come from them? Or could the AFL benefit from a more open data policy like the NFL did? But hiring an analyst isn't making money for CD so how does that work?

It's a little complex, so I'll break down a few things I find quite weird.

First, the cap. AFL clubs have a salary cap on players as well as something called a soft cap. This soft cap applies to money spent on data (including from Champion Data) and club analysts.

The premium product from Champion Data, I am informed, is around \$150,000 per season. Only about five clubs pay for it, currently. (This figure is from me as a fan asking around, rather than investigating as a journalist, but I believe it's roughly correct.) The AFL sets the soft cap, forcing clubs to make decisions about the amount of money they spend on data vs hiring their own analysts and even spend on AFLW.

If you were an organisation being asked to pony up \$150,000 for some data, you'd want to know what you're going to get out of it. And what you get out of it has two elements: the data quality itself and the people in place to extract the value from the data.

The data quality is certainly there: CD hire numerous people at each and every game to ensure the collection of data. But good analysts cost money, and great analysts even more - getting these people to extract value from complex datasets costs is expensive business.

So some (or most) clubs choose not to build their own expertise in analysis, instead relying on the product from Champion Data. When they do work in-house, the soft cap pressures them to keep costs down. The soft cap limits football department spending, restricting clubs' ability to hire the best data scientists, which means they must rely on CD's products.

This arrangement is a little questionable to me in light of the fact that CD is both the exclusive provider of stats and 49% owned by the AFL. It means that each year, clubs are essentially forced to funnel money into the pockets of private individuals, and strongly disincentivized, if not outright prevented, from taking steps to wean themselves off CD by hiring their own experts. Is this really appropriate?

<https://www.afr.com/property/afl-numbers-man-chris-hume-becomes-gillon-mclachlans-neighbour-with-66m-buy-20181104-h17hk0>

So by giving more people the opportunity to show that they can get value in the data - maybe clubs might be more willing to invest in the higher tier of data from champion.

Now this data cost thing: What is going on here regarding access to data? I'm told from Marc that clubs are the most worthy of data, which I agree with.

Now let's examine what clubs get for their money. Why are clubs writing their tools like web scrapers to get some of their data?



**Darren O'Shaughnessy**

@RankingSw



Replying to [@roberto\\_boberto](#)

The clubs don't get this information either. 18 clubs each writing a Telstra app scraper 😞

11:42 · 03 Jul. 19 · [Twitter Web Client](#)

What is going on here? The Telstra data comes from CD. If clubs value it, why wouldn't they just purchase it directly?

Logically, it's pretty obvious that whatever price point it's at, the clubs feel it's too much, given what they know about it's value. Knowing its value is key here, as it's very hard to estimate a dataset's potential without the ability to explore it. And that ability has two parts: access and capability, both subject to decisions that fall under the soft cap.

I know of at least four clubs who pay for raw data from CD, not to use the advanced CD metrics based on that data, but to instead create their own metrics from it. This is partly because of a lack of trust in metrics they can't actually reproduce, and partly because clubs value different

inputs and they want control over these inputs in their models. This seems wasteful, requiring clubs to spend soft cap money on something they don't actually want - money that could have gone toward their own in-house analyst, or perhaps the AFLW.

Some clubs have found ways around this. A common one, which you might have noticed, is that clubs develop relationships with universities. This essentially provides them with a workforce outside the soft cap: university students working on honours, masters or PhD projects.

This is incredibly stifling for the industry - when I went to the [World Congress of Science and Football](#) I watched numerous AFL presentations trying to answer things around game style, congestion and a whole heap of other interesting things.

I felt almost embarrassed but questions to the presenters consisted of:

- This is only one sides data how does it generalise
- This data has both sides but only one match (data only available from match simulations)
- These kinds of questions are already been looked at in sport x,y,z what's next
- 

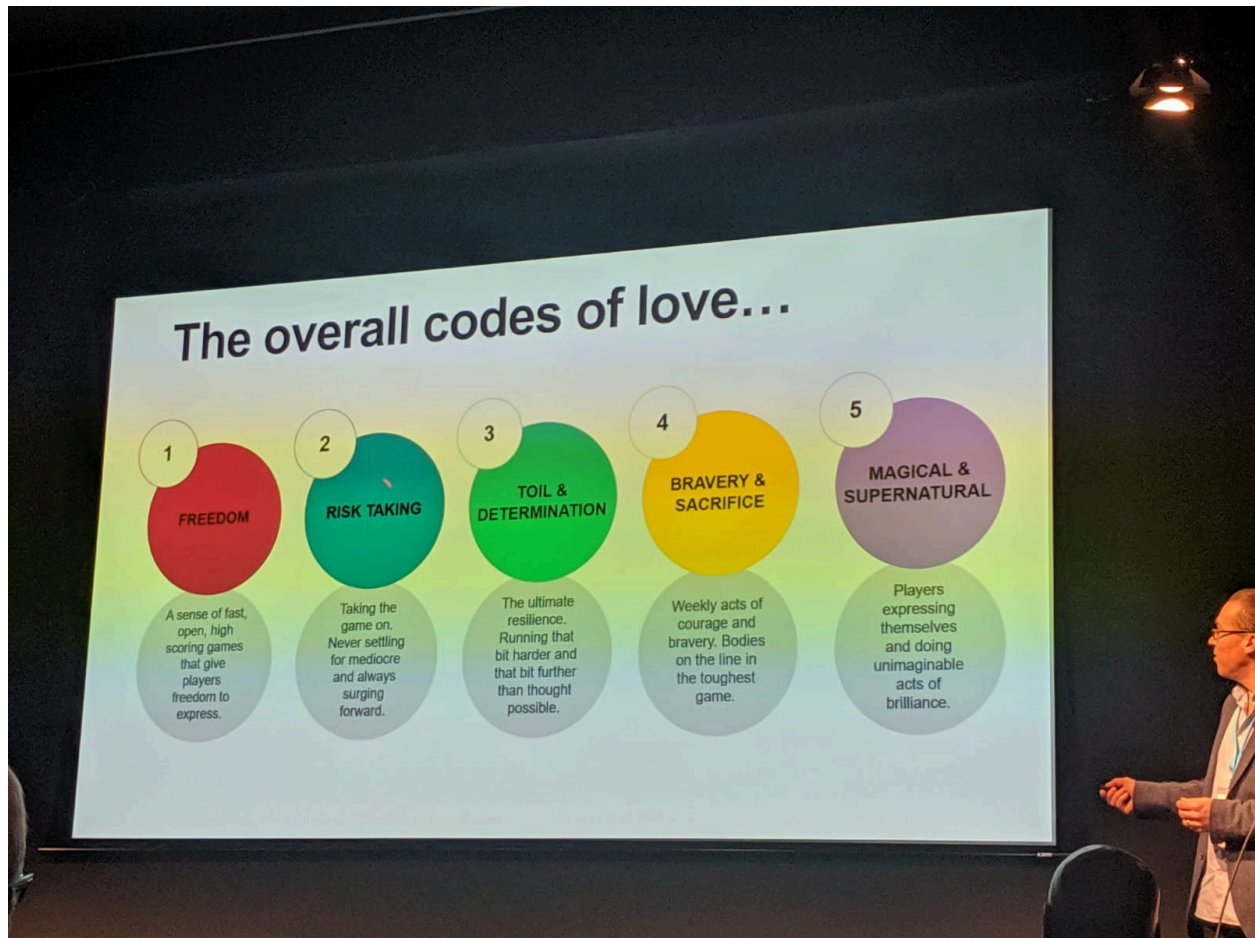
The common answer - we need more data. But it's not available. It exists, but is locked up by the complex relationship that exists between clubs, the AFL and CD, stifling research into the game.

I watched David Rath present at the conference on issues around congestion, including by what metrics those at AFL house use to judge if they are heading in the right direction. It was a great talk, and by the end, you could see how the process worked - as a fan, you felt comfortable that at the very least, the process was right. But I was lucky - I was able to go because I am a student and I presented at this conference, which meant the university paid my way. If I'd just wanted to hear about how the AFL went about thinking about rule changes and how to assess their impact, it would have cost \$600 for one day.

Should information like David presented - the process in which changes were decided on and the evaluation of those rules - should only people who have paid \$600 or only people who are unlucky/lucky to be presenting at this conference know?

For those of you wondering, I took photos - here are photos

We were run through a focus group



Which set the groundwork for things about the game that fans love.

Then we were run through what parts fans thought were vulnerable

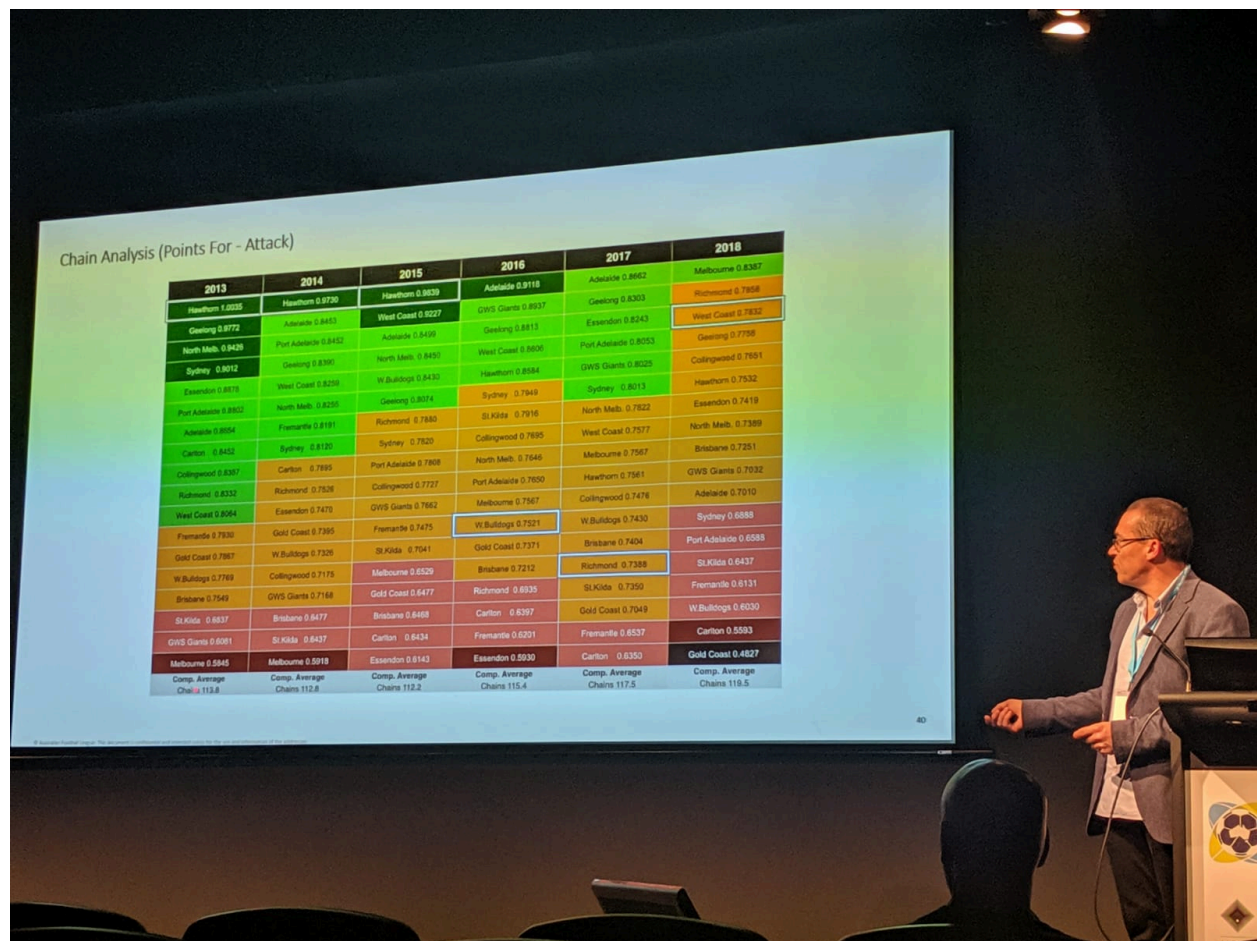




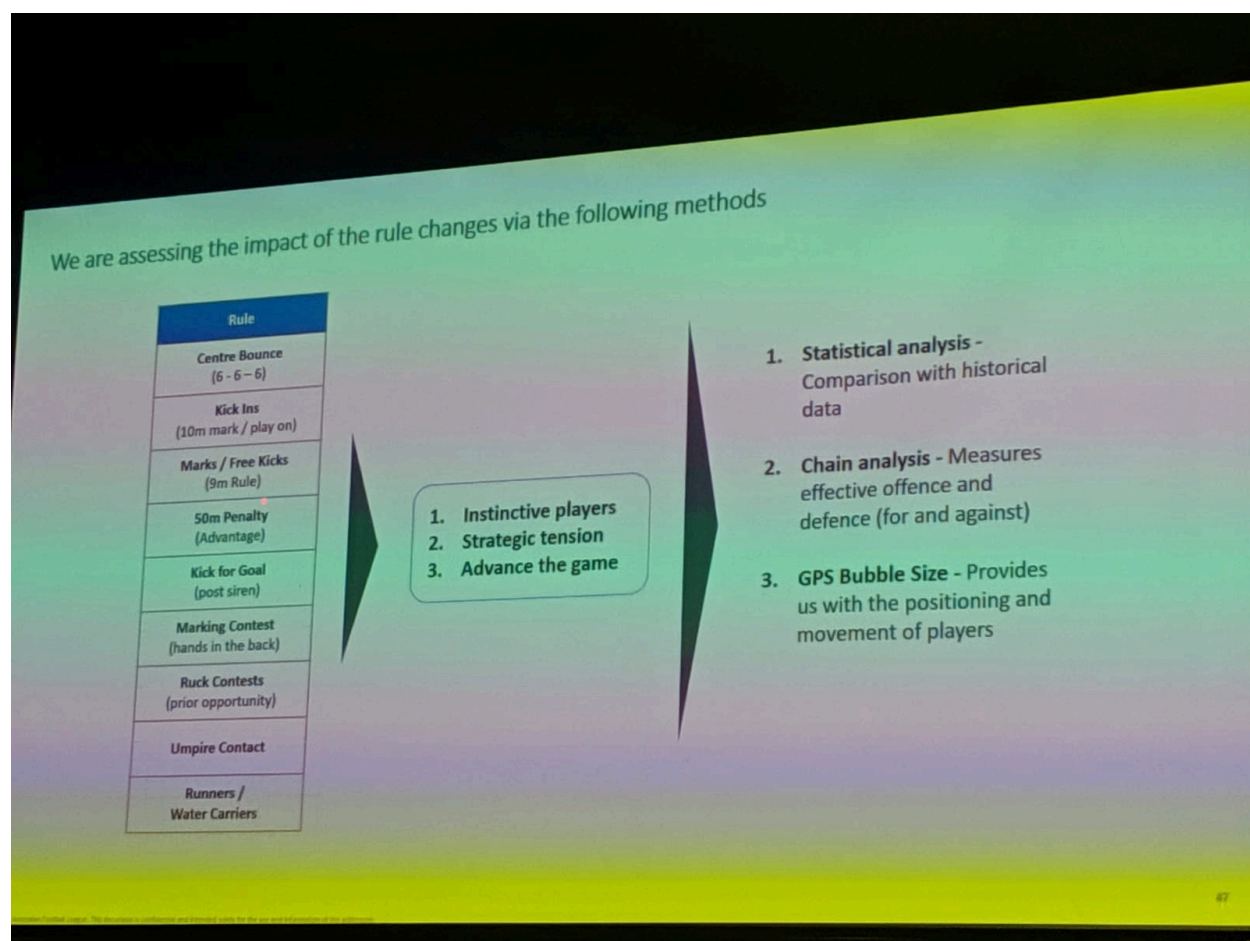
We were given data driven explanations about what was happening in the game.







Then we were told how the impact of rule changes was going to be assessed.



Later this month, [Karl Jackson will present “The Story of Scoring”](#) - how data has changed the narrative. Unfortunately, fans can’t hear this unless they pony up \$500.

So this is the situation we are in. Research is being stifled, while fans and journalists don’t have easy access to data.

This is where I think the online community - the wonks, the nerds - can and want to play a role. People are itching to be able to show a club or the league that they have the technical ability and the thinking capacity to ask interesting questions in ways that clubs can utilize. No one thinks for a second that people will start collecting the data themselves and CD will go out of business. But what is the harm in someone being the next “Figuring Footy,” who went from “hey, let me show you I can do expected scores” to being hired by an AFL club? Do examples like that really undermine CD value proposition?

What is the harm, in articles like Jourdan’s, instead of providing goals and misses, being able to show us fans whos the best kick under pressure? The best kick when sprinting into an open goal? This kind of information isn’t new in club land. What’s the harm in the public knowing?

Opening up data just a little bit more data wouldn't do the league or CD any harm. By allowing data savvy fans to write better content than what is currently available should only create more appetite for it.

My very last point is about the online community and why it has to come from the online community under the system in place.

While I have no interest in wanting to work for an AFL club, I have numerous friends that do. The sad reality is that while analyst jobs are possible the pay rate for them is severely under that which they could get elsewhere - especially for data science roles. To paint a bleak picture, I have heard of club analysts who are paid the equivalent of 3 days work a week and have to find another role. But they do so simply because of a love of the industry and or club. I have heard stories about issues getting an analyst for a job in AFL land. The feedback to the hiring manager was that people are requesting too much money. The response? Find someone who will take a pay cut to be in the AFL system.

The online community operate outside of their normal 9-5 that pays the bill simply because they genuinely love the sport. They would like to grow the analysis behind their sport for free, like other bloggers have overseas. That's the situation we are in, unfortunately. A bunch of people who want to grow the sport analytically, being denied the opportunity.