# Is ARR 2015 A Watershed Moment for How We Manage Hydrologic Data?

Mark Babister[1] Peter D M Brady[1, 2], Monique Retallick[1] and James Ball[2]

[1] WMAwater, Sydney, New South Wales, Australia
E-mail: peter.brady@wmawater.com.au

[2] School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, New South Wales, Australia

*Abstract*

*As network speeds have increased, online delivery of data has the potential to revolutionise the hydrologic industry not only in Australia but worldwide, for example the recent ARR Revision Projects on regional flood frequency estimation and the interaction of coastal and riverine flooding. There are a number of distinct advantages to centralising data stores, namely: quality control that can ensure practitioners are all using the same consistent data set as well as revision control in case changes are required in the future. These, and other, ARR projects, although small scale, have proved that the limitations are not technical but rather cultural, namely they fear of confidentiality when accessing services from remote third party servers as well as the institutionalised culture of downloading and archiving copies of data and software. Centralised data management confronts the established dogma of software licensing in that the concept of paying a single license fee for a particular version or data snapshot is, essentially, eliminated. We believe that this will be replaced with an alternate hybrid organisational structure similar to an open source software project combined with a not-for-profit business that would be run by the industry for the industry. This paper presents a brief overview of the already eliminated technical issues and, more importantly, challenges the hydrologic community to critically evaluate their view of data and software ownership, hydrologic engineering workflows and the direction of collaboration in a data centric workplace.*

## 1. INTRODUCTION

The current drivers of disruptive practices in industry appear to flow from internet and web-based technologies challenging established processes. For example, the shift within the music industry from consumers being required to purchase physical media at a store, to downloads from a virtual store and finally to the current situation where streaming services are rising to supplant downloads (Reynolds, 2015). As another example consider the recent friction between the taxi industry and the Uber ride-sharing platform (Spicer, 2015, McIlroy and Belot, 2015). In both these cases new, centralized technologies have challenged existing ideals and monopolies.

With both of these case studies the common ingredients are that there is an established industry mind-set with an essential monopoly and new tools developed and deployed via web based services. In the remainder of this paper we examine how these concepts relate to the upcoming edition of ARR and the fundamental changes this will drive within the industry. Because ARR is a de facto Australian standard its monopoly position will not change. However, we expect significant, disruptive changes within the online work flows surrounding the processes outlined in ARR. We continue this paper in Section 2 with an outline of the current state of internet connections within Australia compared key international markets and trends for the future – ARR is, after all, a multi-year project. Section 3 will outline what the current ARR editorial team understand is the current practice of use for ARR, while Section 4 will introduce the new delivery methods and systems. Finally, Section 5 will outline the ongoing challenges that the ARR team face beyond delivery of the "document".

## 2. INTERNET WITHIN AUSTRALIA

Internet bandwidth costs have long been a sore point for both consumers and businesses (Bingemann,

2015), with anecdotal reports backed by systematic industry reports. For example, industry leaders in content delivery and bandwidth peering such as Akamai[1] and CloudFlare[2] rank Australia as 46th and 47th worldwide, for the average and average-peak connection speeds respectively (Akamai, 2015a). Yet bandwidth costs in Australia are 20x higher than Europe (Prince, 2014), as shown in Figure 1. These rankings in speed are slowly dropping with Australia edging out of the top 50 worldwide. CloudFlare blames Telstra's monopolistic business practices as the cause of the extremely high cost of bandwidth in Australia (Prince, 2014).

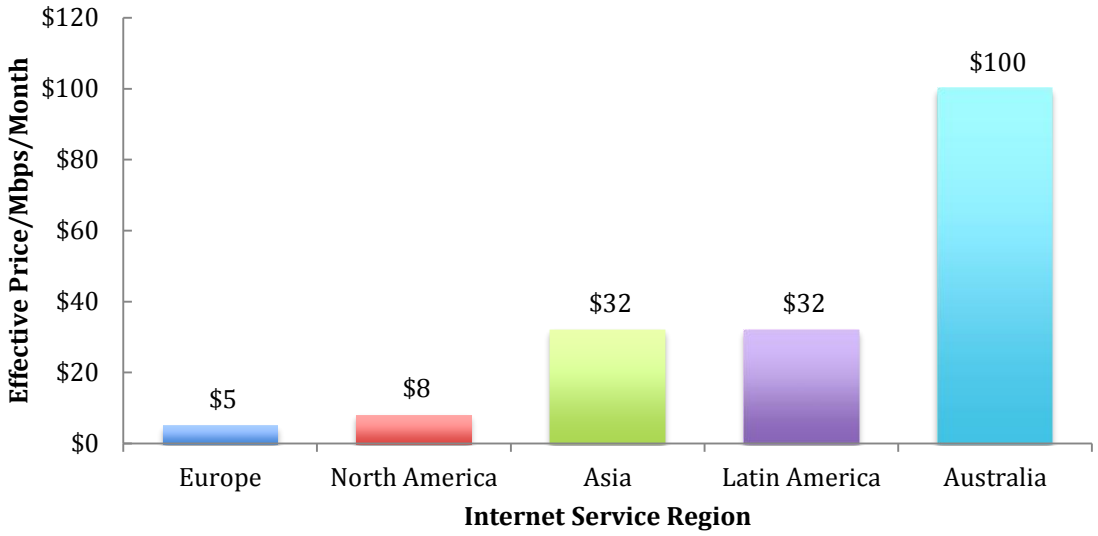


**Figure 1 – Relative cost of network bandwidth by geographic region (Prince, 2014).**

However, simple rankings such as this neglect the fact that internet speeds, while expensive, are steadily increasing within Australia, as shown in Figure 2. Further, it must be noted that business internet connections run at significantly faster speeds than residential connections (because residential connections outweigh the number of business connections and bias the average speed shown in Figure 2 down).

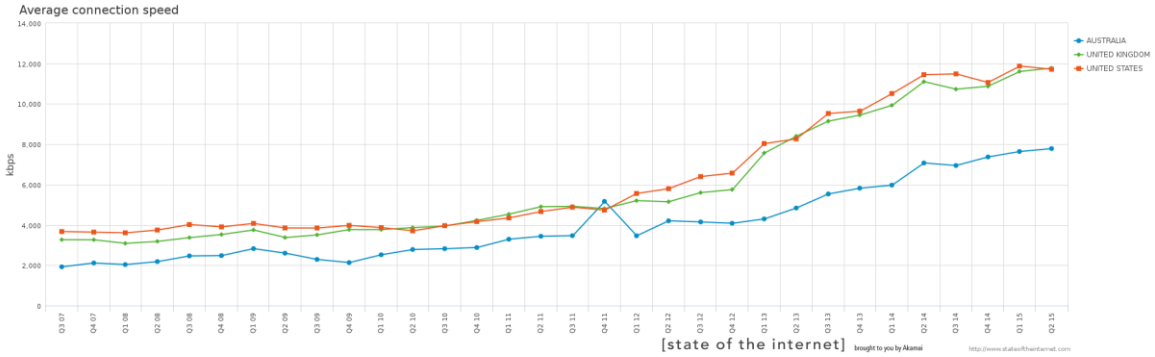


**Figure 2 – Average network connection speeds as a function of time for Australia, the United Kingdom and the U.S.A (Akamai, 2015b).**

There is further potential for increase in Australian internet speeds due to the National Broadband Network (NBN). As at 2nd December 2015 the NBN's stated target is to provide:

> "*…25 megabits per second to all premises and at least 50 megabits per second to 90 percent of fixed line premises as soon as possible.*" (Turnbull and Cormann, 2015)

[1] https://www.akamai.com/

[2] https://www.cloudflare.com/

As an example of current pricing structures iiNet is charging in the order of $120 per month for existing fiber 100 Mbps connections. Therefore, it is highly likely that the NBN's 25 and 50 Mbps lines will be less that the $120/month for 100Mbps. Given that a quick survey of business grade high-speed DSL connections are in the range of $300/month to over $1000/month for connection speeds up to 100Mbps, the NBN will be a positive development for business-to-business connectivity. These lower costs for higher speeds have the potential to drive significant disruptive change within the hydraulics and hydrology sphere. However, to properly examine the potential for disruption it is necessary to examine the current operation of industry based on ARR87.

## 3. CURRENT USAGE PATTERNS

Australian Rainfall and Runoff 1987, aka ARR87, (Pilgrim, 1987) has been in continuous use for nearly 30 years. It was written based on work performed in the late 1970s and early 1980s with a large emphasis on simplified computational techniques for hand calculations and map based contour calculations. Since publication, numerous computer programs have been developed both in-house by engineering practitioners for their own use and, potentially, for commercial sale. Critically, all these software programs are written in either compiled or scripted languages that are run locally on a licensed computer directly by the practitioner interested in the results.

This locality mindset is driven by the book itself: until recently you bought a printed copy of volumes one and two and you kept and used your personal copies. You retained physical ownership of the product. This ownership framework is similarly apparent with additional software and data services such as the LPI ALS data set for NSW in that you would purchase a copy that is shipped to you and it would be available locally.

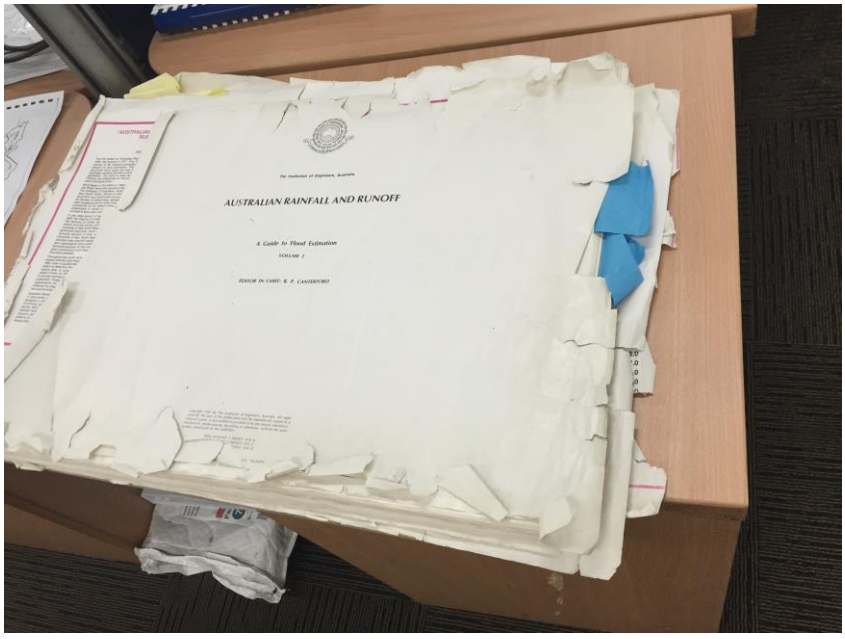


**Figure 3 – A well-loved, some would say typical, copy of Volume 2 of ARR87.**

There is not necessarily anything wrong with this approach per se and, as with most workflows, there are pros and cons. The single largest positive with this approach, particularly for large data sets, is that having a copy stored locally reduces the bandwidth demand on an already overpriced, undersized internet connection. Further, most purchase schemes allow for lower ongoing maintenance costs after the initial purchase. The major negative under this framework is the generally large capital cost at the initial purchase. For a reductionist example, consider the purchase of a book: there is the capital cost to obtain the book in the first place but then there are essentially no ongoing costs after that point in time.

Finally, the largest issue with this framework is the provision of updates or, more specifically, how we

expect that it happens in practice in that no or minimal updates occur. For example, consider the problem of a data set that is distributed via a DVD-ROM. As with any data set there will be errors that must be corrected as well as changes that grow naturally – such as changing roads on a topographic data set. Under the present purchase and retain model, every licensee must be contacted to be told that there is a data update after which the licensee must decide if they wish to purchase the update and have the item shipped to them. Given the number of steps in this chain there is a very high probability that there are islands of isolated data, and programs, books etc, currently in use within industry.

It is the view of the current editorial team of ARR that the negative benefits outweigh the positive benefits of ARR87. Therefore, the key issue for the new edition becomes one of how to properly utilize the advances in 30 years of technology to deliver a relevant, useful publication for practitioners.

## 4. DELIVERY OF THE NEW ARR

The new ARR will be a significant departure from the approach described in the previous section, specifically:

1. The base version of ARR will be free of charge for all users;
2. The concept of ownership has changed in that copyright will no longer be vested with Engineers Australia but rather the Federal Government;
3. Distribution will be 100% electronic and will leverage existing infrastructure to consistently apply updates, and;
4. Non-commercial enabling software and data sets will be centralized.

The rationale behind this framework is to reduce, or ideally, eliminate, the isolated data island problem discussed in Section 3. This will be done by utilizing the increased-speed-at-lower-cost internet access available within Australia.

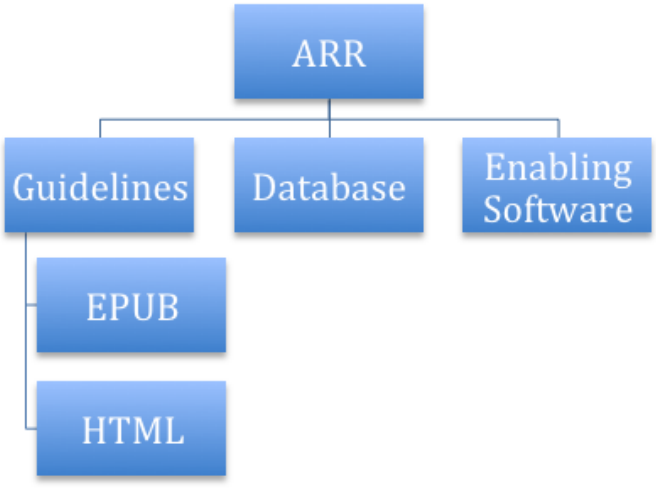


**Figure 4 – Schematic of the new ARR.**

The key concept here is not control but rather **consistency**. This framework is designed to create a **living document** subject to more frequent but smaller updates than a 30-year cycle. To borrow from software engineering this is almost a **continuous revision** system.

With respect to item two in the framework list, the Government owns the copyright to ARR works. However, any future work will be a collaboration, as the technical expertise rests with the community of hydraulics and hydrologic engineers through EA's National Committee on Water Engineering. The copyright is more of a technicality than a practicality, as any updates not managed through the NCWE would lack the authenticity required for acceptance by the community.

In short, the days of downloading a data set once and continually reusing it are over. A centralized system, such as this, combined with download records and revision control provides for a system –

different from ARR87 – with reproducible outcomes.

Early feedback from stakeholders raised some concerns that such a centralized system would lead to a control in the flow of information. While this is a valid concern the use of web-based systems allows for exactly the opposite, in that it facilitates the rapid dissemination of updates as needed. For example, we can use the existing Amazon and iBooks web shops to distribute an eBook edition of ARR – just because it is in these stores does not mean that there will be a cost associated with a download of the book. In the future when (note the use of when and not if) typos are found the team can update the book within the online store and every user is automatically notified of the update and in most cases the update is applied automatically. This leads to a constancy in the information that all practitioners are working from.

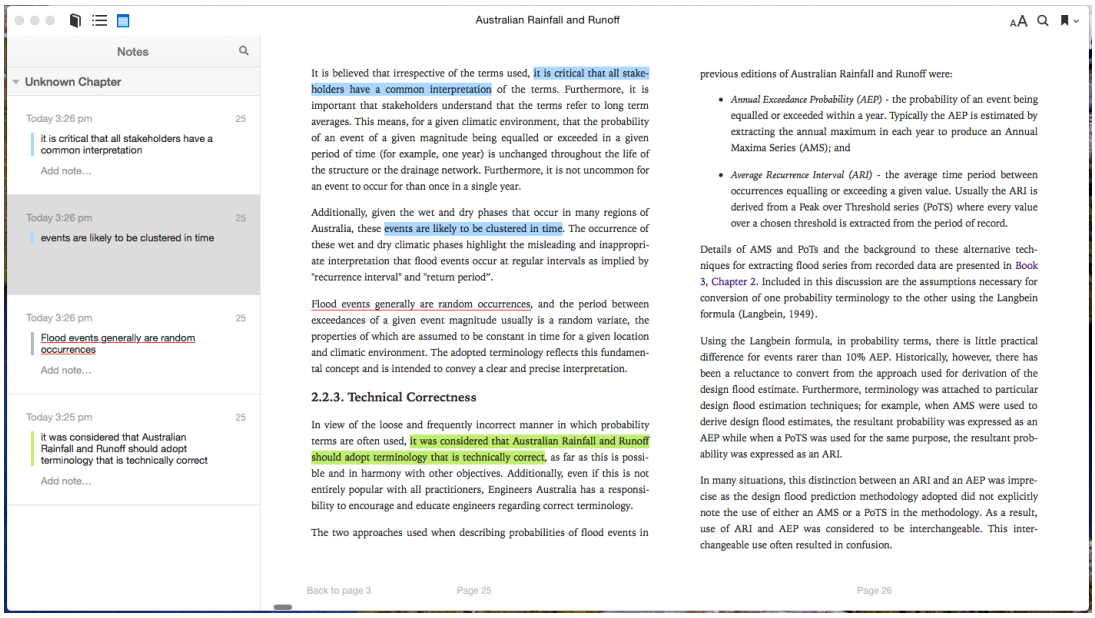


**Figure 5 – An annotated copy of the new electronic EPUB.**

A similar philosophy applies to centralized enabling software. This includes packages such as FLIKE, web-based Regional Flood Frequency Estimation[3] (Project 5 - RFFE) and Joint Probability[4] (Project 18), which the ARR team, and associated entities such as the Bureau of Meteorology, have developed as part of the revision process. The benefit of a single centralized software platform is consistency: all users access exactly the same program and data set. When bugs are found they only need to be pushed to one site and then, as before, all users will be automatically working off the same platform.

It is worth emphasizing: these are not software packages in the traditional commercial sense. They need to be thought of in an alternate framework that they are enabling software in the manner of a public utility. They are unlikely to ever be profitable, in fact more likely to be a burden to the managers, but are necessary for consistency across the user base.

User feedback concerning the centralized software has indicated a level of caution within the community that the third party organisations that host the software may be able to "spy" on the commercial in confidence information that is passing through the servers. This is a legitimate concern given that confidentiality clauses are mandatory on most projects. Moving beyond the word of the host there are a number of methods to disguise information in this system. Consider the example of the ARR enabling software. Given that in most usage scenarios an engineer would need a number of combinations computed for their work, the simplest approach is to batch the data and send it through immediately. However, where this is not suitable for commercial in confidence work the batch processed can be modified by arbitrarily increasing the sample count with random, or semi-random, samples to disguise the samples of interest. In certain cases the raw data can also be normalized. These two methods

[3] http://rffe.arr.org.au/

[4] http://p18.arr.org.au/

should be enough to assuage the nerves of most users.

While the free delivery of ARR eliminates capital and ongoing costs to an organization with respect to license purchases there are two potentially significant issues to be addressed with this centralized framework. First: this generates a reliance on a third party to maintain the underlying systems. This introduces a small risk that is not present with a physical book: the practitioner own the book for life. However, with this delivery system the practitioner will be reliant on the ARR team to maintain the server/s and push updates. Second: with centralized server usage most businesses will be moving more data across their external network connection. Given the discussion in Section 2 this should be a minor issue that diminishes over time.

It is clear that while the delivery of ARR in this manner will challenge the established community traditions, the new system is not free from implementation issues.

## 5. CHALLENGES FOR THE ARR TEAM AND THE COMMUNITY

There are two broad categories of ongoing issues that the ARR team will have to deal with:

1. So called "stay in business" costs, and;
2. Managing the continuous revision process.

The "stay in business" costs are those associated purely with maintaining the current system. It must be reinforced at this point that the new ARR is more than just a book but rather it is a cluster of books with integrated web-server based enabling software packages and supporting data stores. At the present time, and over the duration of development, the ARR servers are hosted on various VMware clouds run by organisations involved in the development of ARR. As of December 2015 the ARR team are planning to continue this arrangement over an evaluation period after which formal decisions for future planning can be made. Viable options that are being considered include hosting at one or more locations that donate colocation facilities or even hosting on cloud provides such as the Amazon Web Service Elastic Cloud Compute (AWS EC2) facility. However, given conservative costs of $215USD/month[5] ($295AUD/month[6]) for low spec machines the EC2 option would need significant ongoing funding, not to mention the significant ongoing risk of currency fluctuations given EC2 is priced in USD. Though EC2 is not cost effective for long term deployments, it is much more efficient to host onsite where efficiencies of scale can be made to slot ARR services in with other corporate workloads.

These projections of server costs have been significantly reduced because, under current agreements, Geoscience Australia will host the spatial data sets. These large sets create unique challenges for hosting in that they consume large volumes of disk space to store and even higher bandwidth costs as they are downloaded. GA sponsorship has removed a source of concern for the ARR team in this area.

In addition to hosting charges there is the cost of providing ongoing maintenance of the server systems. For example, the application of security patches and recompiling/testing programs as bug fixes and patches are deployed. The provision of these services has not been fully scoped as of writing this paper.

The second major component of running ARR is the continuous revision process. The logical vehicle to provide ongoing management would be the Engineers Australia National Committee on Water Engineering (NCWE). As the cost of running revision projects is significantly higher the most likely source of funding would be government or industry grants and as such it would make sense for the NCWE to maintain, through effective collaboration with stakeholders, a "wish list" of projects that can be easily accessed if and when grant funding is available.

[5] Based on Amazon Web Services Elastic Cloud Compute prices as at 2015-12-02 for four (4) linux m3.medium instances in the Sydney, Australia data centre on a three-year all up front reserved instance contract.

[6] Exchange rate as of 2015-12-02

## 6. CONCLUSION

This paper has presented the results of an examination of the drivers of industry disruption and how this can be leveraged to provide a significant benefit to the hydraulic and hydrologic community. Specifically we have presented the change to the established system from a user-owns system to a centralized system. The centralized system provides a number of benefits, which we believe, more that outweigh the negative benefits. Namely, the twin benefits of consistency and ease of maintenance of a centralized system exceed the disadvantages of possible bandwidth increases and a reliance on third parties to maintain systems. In the interests of full transparency we briefly presented the known ongoing challenges that the ARR team must work through to ensure that the current edition remains relevant to its user base.

The new ARR will challenge the work methods of the engineering profession. It is an exciting time as the new edition is deployed.

## 7. ACKNOWLEDGMENTS

James Ball is the Editor of Australian Rainfall and Runoff. Mark Babister is the Chair of the ARR Revision Technical Committee. Monique Retallick is an ARR Project Engineer. Peter Brady is a Senior Engineer at WMAwater.

## 8. REFERENCES

Akamai 2015a, 'Q2 2015 Report', *Akamai's State of the Internet*, vol. 8, no. 2.

Akamai 2015b, *'Time-based trends in Internet connection speeds and adoption rates'*, Available at: https://http://www.stateoftheinternet.com/trends-visualizations-connectivity-global-heat-map-internet-speeds-broadband-adoption.html (2015-12-02)

Bingemann, M. 2015, *''Netflix effect' is slowing the nation's broadband internet speed'*, Available at: http://www.theaustralian.com.au/business/technology/netflix-effect-is-slowing-the-nations-broadband-internet-speed/story-e6frgakx-1227520116416 (2015-12-02)

Mcilroy, T. and Belot, H. 2015, *'Uber to launch in Canberra; ACT government to cut taxi licence fees and regulate ride-sharing'*, Available at: http://www.smh.com.au/act-news/uber-to-launch-in-canberra-act-government-to-cut-taxi-licence-fees-and-regulate-ridesharing-20150929-gjx7a7.html (2015-12-02)

Pilgrim, D. H. (ed.) 1987, *Australian Rainfall & Runoff – A Guide to Flood Estimation*, Institution of Engineers Australia, Barton, ACT, Australia.

Prince, M. 2014, *'The Relative Cost of Bandwidth Around the World'*, Available at: https://blog.cloudflare.com/the-relative-cost-of-bandwidth-around-the-world/

Reynolds, C. 2015, *'Social, Streaming & On-Demand: Key Trends In The Music Industry Today'*, Available at: http://socialmediaweek.org/blog/2015/06/social-streaming-demand-key-trends-music-industry-today/ (2015-12-02)

Spicer, D. 2015, *'Taxi drivers rally against Uber in Sydney, calling drivers for ride-sharing service 'animals''*, Available at: http://www.abc.net.au/news/2015-09-10/taxi-drivers-protest-outside-nsw-parliament-over-uber/6764398 (2015-11-02)

Turnbull, M., The Honorable Minister for Communications and Cormann, M., Senator the Honorable Minister for Finance 2015, *Government Expectations*.