A Community of Practice for Data Professionals: Brief 2

Table of Contents



# What do we need?

What we need in the health system in South Australia are systems that promote the continuous development of data skills by data professionals and the continuous promotion of these skills throughout the health system and the broader health analytic profession. This amounts to a goal of creating a culture that values data skills and expertise and establishes clear incentives for those with less data skills to gain more skills and for those with more skills to go further to be world leading in their field of data expertise.

This goal is inconsistent with the maintenance of current systems for using data and managing data professionals. This goal is clearly a goal of change. Fast utilisation of data assets is inconsistent with analysing data in software that does not make its processes replicable. Bringing new technology to the point of use is inconsistent with software being written on individaul computers that do not maintain version control. Creating long standing data driven change is inconsistent with managing data projects via email in which the feature requests are uncoupled from the solutions provided.

This paper details the structures that are currently in place and the pportunities for layering novel systems that will bring this goal into reality.

# Background

The health system in South Australia employs some particularly high performing data professionals. However, there are some limitations on the current and future contribution that these professionals and their colleagues can make.

## A swiftly changing set of skills

The best "data skills" are constantly evolving. Not long ago, knowing how to create pivot tables in Excel would have constituented serious data skills. Likewise, there will come a day when the [R](https://www.r-project.org/about.html) language for statistical computing will not be the [de facto standard](https://bookdown.org/rdpeng/rprogdatascience/) programming language for data science. To this end, data professionals and their managers need to spend a period of their regular work time updating their skills. This updating means that data professionals will know the latest approaches to their current work and their managers will know what sort of work can even be achieved.

## Skills need to be utilised

As the economist Erich Zimmermann was wont to say: ["Resources are not; they become"](https://www.jstor.org/stable/4225924). In the context of data professionals in the health system in South Australia, it seems that the skills that are currently held in the system are not consistently utilised by management.Hence, these skills are not, in Zimmermann's phrasing, an asset. They are not being used and so are not given the opportunity to **become** assets. Evidence of this underutilisation of potential resources is seen in the amount of data driven work that is outsourced by the health system in South Australia. Below are a selection of contracts for data projects paid for by the health system in South Australia that may have been achievable by staff currently in the system.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Reference | Contract Title | Start Date | Expiry Date | Total Cost |
| SAH760 | Data refresh of My Home Hospital Business Case - Contract ID SAH2020-1198 | 03/08/2020 | 07/09/2020 | $ 31,720.00 |
| SAH701 | SAHMRI Consultant Engagement - Health and Aged Care Interface Data Project Contract ID SAH2020-533 | 08/05/2020 | 23/10/2020 | $ 87,625.00 |
| SAH746 | Think Research's Visit to Central Adelaide LHN | 02/12/2019 | 02/01/2020 | $ 60,000.00 |
| SAH680 | Evaluating the health transform, inspire, engage and redesign leadership and management training program | 01/05/2019 | 31/12/2020 | $ 56,184.00 |
| DH042264 | Data collection services for population health surveillance surveys Contract ID SAH2017-1069 | 01/05/2018 | 30/04/2022 | $ 3,404,573.00 |
| DH040717 | Health Activity Modelling and Forecasting Services Contract ID SAH2016-2026 | 20/02/2018 | 19/02/2022 | $ 686,400.00 |
| DH026484 | Data Management System. Contract ID: SAH2014-983 | 01/04/2015 | 30/03/2022 | $ 774,213.00 |

Greater visibility to management of the skills held by data professionals in the health system in South Australia would help turn these abilities into resources through their use.

## Siloed skills are less valuable

Data professionals in the health system in South Australia frequently work as the only data professional in a team that includes clinicians and administrators. There are several reasons why these isolated data professionals and the health system at large would benefit from regular interaction with their data professional peers.

* A data asset that is made available, software that is built or documentation that is written in the service of one use case should be made available to other data professionals in the health system in South Australia to reduce the number of times that same task is done in different places.
* Techniques and technologies that affect the health care system in South Australia are developed by people working in government at state and federal level, private industry and academia. It would help the system to fascilitate the trasfer of novel approaches across these institutional divides.

# Avenues for improvement and what already exists

There are a number of avenues available for meeting these goals. There already exists a range of regular events in Adelaide, around Australia and internationally that could serve as models to draw from or as opportunities to engage with.

## Colloquia

A sharing of work that focuses on the technical components in which informed peers could be invited to ask difficult and pertinent questions.

In this space already:

* The Adelaide University Faculty of Engineering, Computer and Mathematical Sciences hosts regular [colloquia.](https://ecms.adelaide.edu.au/events/list/2020/all)
* R Users Adelaide annual [meet-up](https://www.meetup.com/en-AU/Adelaide-R-Users/).
* The CEIH and SAAS have hosted a data visualisation [expo](https://www.eventbrite.com.au/e/data-visualisation-expo-tickets-104931753700).
* CALHN runs a monthly CoP that frequently invites internal presentations coordinated by Andrew McAlindon.
* The [HPC](https://www.hpcsa.com.au/) coordinates the DAGS group that meet monthly and invites internal and external presentations coordinated by Jane Austin.

### Specifics

#### Venue

A suitable venue has the following characteristics:

* Central location to lower the friction associated with attending.
* No venue hire costs.

One venue that would fulfill these characteristics is the Minister's Board Room at Hindmarsh Square.

#### Frequency

Too often and it would feel like a burden but too rare and it would be forgotten. To this end, meeting every month would be preferred.

#### Cost

To some extent it would be desirable for the event to stand on its merits as a place to hear and debate the latest ideas in health bioinformatics rather than a place to nibble food and drinks. Hence, the event could prosper with a smaller budget.

#### Invitation list

Lukah Dykes and Owen Churches have begun drafting a list of potential speakers and people to invite. This will be updated here.

## Expert reading group

A reading of serious texts on an aspect of data analysis, software design, change management or ethics then an informed discussion about the applicability of the author's work to our organisation.

In this space already:

* Owen Churches convenes a monthly book club on [AI ethics](https://github.com/ofchurches/AI_ethics_bookclub/wiki).
* The [NHS](https://nhsrcommunity.com/blog/nhs-r-book-club/) runs a data book club.

### Specifics

#### Venue

Both the NHS group and the AI ethics group meet online and this works particularly well. It would help to use a wiki style collaboration medium for adding additional resources and a threaded communication platform such as Slack to maintain communication between meetings.

#### Texts

What we read would be a reflection of the aims and the capacity of the group. Professor Mark Mackay has suggested that this be a vehicle to immerse people who are not currently experts in a more quantitative way of thinking. This would be in keeping with the NHM book club. Hence, books that could be read could include:

|  |  |
| --- | --- |
| Book | Author |
| Calling Bullshit: The Art of Skepticism in a Data-Driven World | Carl T. Bergstrom Jevin D. West |
| How To Make the World Add Up | Tim Hartford |
| The Ostrich Paradox Why We Underprepare for Disasters | Robert Meyer and Howard Kunreuther |
| The Art of Statistics: Learning from Data | David Spiegelhalter |
| The Art of Logic in an Illogical World | Eugenia Cheng |
| Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming | Naomi Oreskes, Erik M. Conway |
| The Signal and the Noise: The Art and Science of Prediction | Nate Silver |
| Standards: Recipes for Reality | Lawrence Busch |
| The Sum of the People: How the Census Has Shaped Nations, from the Ancient World to the Modern Age | Andrew Whitby |
| The Design Thinking Playbook: Mindful Digital Transformation of Teams, Products, Services, Businesses and Ecosystems | Michael Lewrick |
| Factfulness: Ten Reasons We're Wrong About the World and Why Things Are Better Than You Think | Hans Rosling, Ola Rosling, Anna Rosling Rönnlund |

#### Cost

We should budget on $50 per book but after this, this is a function of the number of people in the group.

## Technical skills practice

Testing and developing analytical and programming skills individually or as a group on public data sets that promote the quick exploration of public sharing or outcomes.

In this space already:

* Owen Churches coordinates the [PublicSectR](http://publicsectradelaide.rbind.io/) monthly meetup which includes a sprint through that week’s [TidyTuesday](https://github.com/rfordatascience/tidytuesday) dataset.
* Adelaide University's statistics specialisation runs a GentleR group each week for people to bring in their R related problems.
* The [CAIN](https://www.cain.science/) at UniSA runs a weekly R group focused around reading the [R4DS](https://r4ds.had.co.nz/) book.

### Specifics

#### Venue

There are pros and cons to meeting online and in person. Online meetings reduce the friction of attending but in person meetings promote a stronger group identity and commitment to the process. To start with, the suggestion is that we meet in person at SAHMRI. If more people want to take part at different venues, then we could have multiple locations.

#### Technology

It would be fitting if the technology we used to work on this was itself a learning opportunity for those taking part and an example to others of what we can achieve. Hence the proposal is that we maintain a portfolio of our work in a [blogdown](https://bookdown.org/yihui/blogdown/) site with an associated GitHub repository.

#### Focus

Two potential sub-avenues exist:

* A methodological learning together group that could move sequentially through:

1. [R4DS](https://r4ds.had.co.nz/)
2. [Advanced R](https://adv-r.hadley.nz/)
3. [Mastering SHiny](https://mastering-shiny.org/)
4. [Hands-On Machine Learning with R](https://bradleyboehmke.github.io/HOML/)

* Greer Humphrey and Owen Churches have begun work on a [TidyTuesday](https://github.com/rfordatascience/tidytuesday) group to design together and create a visulaisation and analytic product based on a novel data set each week.

#### Frequency

This group would benefit from meeting weekly.

## Access to resources and tools

Access to the latest journal articles to learn from and to use modern software design tools.

## Formal training

To some extent, formal training is helpful.

In this space already:

* The CEIH paid for a large number of data professionals in the South Australia health system to undergo the [CHIA](https://www.healthinformaticscertification.com/) exam.
* The HARC pays for the Data Fellows to attend some training that is chosen by each fellow individually.
* There are numerous paid and free data skills courses available (e.g. [Coursera](https://www.coursera.org/specializations/data-science-foundations-r)).

### Specifics

This will be perceived by some as an indulgence for some to avoid doing work. However, structures should be put in place to make a clear return on investment for the health system in South Australia. These could include:

* If the exams are not taken, the learner could be asked to return the money.
* The choice of training could be decided by managers with experience of each learner's skill base and the needs of the health system.

## Secondments

Hosting experts from industry and academia within the government system for a period of time and sending government employees to work in industry and academia on the condition of their return would help foster relationships and build skills.

### Specifics

This would require strong structures to ensure that measurable outcomes were attained. These could include:

* A process for disseminating the results of the project at its conclusion.
* A competitive entry program based on merit of person and project

## Hiring and propotion practices

Within the contemporary software and data analytics industry it is common to recruit through the based on an applicant's engagement in the [open source programming community](https://www.socialtalent.com/blog/recruitment/source-amazing-developers-nobody-else-can-find) including their GitHub profile or history of stack overflow answers. Following the recruitment stage it is also common to have a selection process that includes a [code intervew](https://www.tutorialspoint.com/r/r_interview_questions.htm). The health system in South Australia could provide similar practices.

The deficiencies of quantitative skills at the appropriate levels in the South Australian health system may be similar to those found in the Australian Public Service in the 2019 [review](https://pmc.gov.au/sites/default/files/publications/independent-review-aps.pdf) of the APS. Importantly, this review proposed the following:

**Recommendation 14**

Building data and digital skills. Establish a new APS digital and data profession or professions to ensure that the APS has the skills required to develop, build and maintain ICT systems that enable the Government to be easy to deal with.

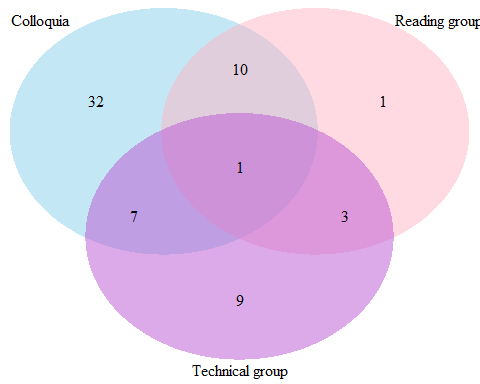
If implemented in the South Australia health system, this recommendation would have the result of data professionals being able to be promoted based on their data skills rather than their managerial skills. It would make the system more likely to retain and attract data professionals by making the system competitive with the private sector. And, it would incentivise data professionals to become better data professionals. Importantly, it would make data professionals reasonably expect a return on their investment in gaining more technical skills rather than pursuing a managerial path.

# Common risks

There are several reasons why this community of practice could not thrive or could fail to have an impact on the health system in South Australia. These can be mitigated against but can not be removed. It is worth saying at the start that this project could be so valuable that even if it does fail, it will still have been worth attempting.

## Overlap

Ideally, the different avenues described above will feed each other and expand the total pool of people engaged rather than cannibalizing the same pool. In particular, the first three avenues which all entail regular events could overlap in the following manner:



## The right people

Many of the people who could most benefit from these avenues may not readily realise the benefit to them as individuals that would come from their participation. It may help to fold these events into professional development.

## Managers need to help

Participation would also have to be supported by managers who would have to ensure time was quarantined for participation in these activities and who led by example in attending every event. There would have to be a perception amongst data professionals that being part of these activities would realistically lead to better individual outcomes such as promotion.

## Code of conduct

It is imperative for communities of practice such as these to have well defined and agreed to codes of conduct. Tina Hardin in the CEIH has begun work on a code of conduct for this purpose.

# The Proposal

## Phase 1

The proposal is that three avenues be started in February 2021. These will include:

* A colloquia
* A technical skills group
* An expert reading group

These will be funded to run for 11 months to the end of 2021. Experimentation with the format, frequency and people will be encouraged during this year. An evaluation in November 2021 will decide on the longer term implementation of these avenues.

The starting configuration will be:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Avenue | Frequency | Attendees | Cost per meeting (not including staff time) | Cost per year |
| Colloquia | Monthly | 50 | $100 | $1100 |
| Technical skills | Weekly | 20 | $0 | $0 |
| Reading group | Monthly | 15 | $750 | $8250 |

### Action to take

Each of the three avenues has specific detailed formats that are accessed here ( [colloquia](https://github.com/ofchurches/HARC_Development/blob/main/seminar.md), [technical](https://github.com/ofchurches/HARC_Development/blob/main/technical.md), [reading](https://github.com/ofchurches/HARC_Development/blob/main/reading.md)). Actions to be achieved are logged as [issues](https://github.com/ofchurches/HARC_Development/issues) in this repository.

## Phase 2

It is further proposed that the other four avenues be implemented in July 2021. These will include:

* Access to resources and tools
* Formal training
* Secondments
* Hiring practices

### Action to take

Each of these will require liaising with human resources and information technology management to facilitate. This process will begin in February 2021 with actions to be achieved logged as [issues](https://github.com/ofchurches/HARC_Development/issues) on this repository.