CSIRO MINERAL RESOURCES
www.csiro.au

Big Data improving the productivity of minerals processing

Final Report – Version B

Craig Mudge, Greg Roach, James Tickner

Report Number EP173774

1st Edition 26 July 2016 2nd Edition 30 November 2016 3rd Edition 5 May 2017

Mining and Petroleum Services Centre of Excellence
South Australian Department of State Development

Acknowledgements

The authors of this final report wish to acknowledge useful help and commitment from Jake Gallina, metallurgist at Olympic Dam. Jake helped us understand the processes involved, hosted a visit to Olympic Dam, and then as modelling progressed he was an important discussant. His assessment at project end was most useful. Alain Bouchard of BHP Billiton, acting as our main interface to BHP Billiton, provided guidance regarding his IT systems, sponsored meetings with the appropriate BHP Billiton staff, and provided leadership for follow-on work, which resulted in the Predictive Maintenance proposal. Our industrial trainee, Rob Wright provided great assistance.

Within CSIRO, our partnership with IM&T provided essential engineering and development, with David Benn and Brendan Speet being noteworthy contributors. In all ways, David Benn was a member of our team.

Overall sponsorship was provided by Steve Liddell at BHP Billiton and Nick Cutmore in CSIRO.

Contents

Figures

Executive Summary

Over the last two years, the Big Data project in CSIRO Minerals Resources has partnered with two global miners in a demonstration of the use of data analytics in improving performance parameters of portions of major minerals processing plants. These parameters, which include throughput, recovery, and energy use, contribute directly to plant profitability. We were entrusted with several years of sensor data from some 2000 variables, some every minute, some every hour, and some daily.

Three technology shifts support the development of more automated approaches to plant control. They are (a) increasingly sophisticated instruments for real-time measurements of chemistry and mineralogy, (b) information networks that can gather and transmit thousands of pieces of data second-by-second from every part of a plant, and (c) massive-scale data storage and processing, as cloud computing, and new machine-learning software.

Machine learning, currently the most significant part of artificial intelligence (AI), involves programming computers to learn from example data or past experience -- to allow prediction, recognising images, and discovering anomalies. It has given us speech recognition (Siri), excellent translation of languages, self-driving cars, and web search.

We began by formulating a data-sharing agreement with Anglo American to obtain plant data from their platinum plant at MGK in South Africa. The historical data came from a database system from OSISoft PI for the entire concentrator plant. We installed PI at CSIRO. However, we were unable to learn enough of the detailed processing steps to formulate a reasonable model of the entire plant. Moreover, a lot of missing data and, again, our lack of knowledge of the plant made it impractical.

We decided to take a simpler approach by modelling just one unit step of a plant and to select a data-sharing partner who could supply plant knowledge in a straightforward manner. BHP Billiton (BHPB) came forward and offered data from its Olympic Dam plant in South Australia.

The Svedala mill, the largest of three mills onsite, was chosen for analysis. Initial findings from analysing six months of data suggested there was a potential for mill throughput improvement.

To improve the chances of adoption of our results we ran a two-month pilot at Olympic Dam (OD). A scalable system for secure daily exchange of data between OD and CSIRO and cost-effective modelling was built around a public cloud service from Amazon Web Services, which complied with the security policies of both BHPB and CSIRO

Our modelling over a ten-month period showed that a small percentage improvement, but significant economically, in throughput could be achieved.

In addition to improving throughput, sensitivity analysis on key variables was carried out. Our modelling of the effect of twelve variables such as FeSiO2 ratio, mill fill, mill power, and final density provided new insights for plant metallurgists and confirmed others.

The three key results of this Big Data project are (1) the contribution of machine learning to improving OD mill throughput, (2) encouragement to extend our application of AI to other parts of process control and plant management, and (3) how cloud computing simplified data exchange between CSIRO and BHP, and gave straightforward scaling to other BHP mines.

1 Background

Across a wide range of base and precious metal commodities, typically 5-35% of the mined metal is lost at the initial concentration stage. With a large concentrator producing $1 billion or more of product annually, even small improvements in recovery can be worth tens of millions of dollars per year.

In the boom years of the early 2000's, producers focused on maximising throughput to increase their profitability. With stagnant or falling commodity prices, declining ore-body grades, increasing input costs and growing concerns over water and energy usage, attention is shifting back to processing efficiency.

Improving efficiency implies better control of the numerous components – mills, cyclones, flotation cells and thickeners – that make up a modern concentrator. However, it is becoming increasingly difficult for miners to find the skilled, highly experienced metallurgists, geologists and plant engineers needed to run their complex operations. Typically, local or fly-in-fly-out staff manage the day-to-day control of a plant, with occasional audits conducted by senior corporate or external groups to oversee long-term operational performance.

This creates a significant opportunity for new businesses looking to introduce novel technologies that can help plant operators to lift recovery, improve throughput, or reduce energy costs, particularly if these developments reduce the need for skilled and field-based staff.

This project, then, aimed to use modern Big Data and machine learning techniques to explore data-driven modelling, i.e., analysing historical plant data to produce a model of a processing plant which could be used to improve plant performance

'Big Data' and 'Machine Learning' have become buzz-words of the early 21st century, perhaps reflecting dissatisfaction with the growing gap between information and knowledge, engendered by our ability to collect and store data outstripping our ability to extract meaning.

Behind the hype, companies such as Google and Facebook have shown that data-driven modelling is a powerful approach for a range of complex problems, provided that sufficient historical data is available, and that it is matched with large computing and data-processing resources. Some applications where data-driven methods have been shown to be more effective than first-principles models include natural language translation, the self-driving car, voice recognition, and a University of California personalised medicine project, based on a massive repository of genetic sequences of thousands of cancer tumours.

To clarify the positioning of our project, we note that we are operating in step 5 of the minerals value chain (Table 1).

Table 1. Positioning of project on the minerals value chain

(Greenfields) Exploration	Resource assessment	Planning and mine devel.	Mining	Mineral processing	Smelting and refining	Closure and remediation
1	2	3	4	5	6	7

A mineral processing facility typically logs data in real time, uses the information immediately for plant control and then stores the data – often never to be used again! One of the underlying premises of Big Data is to utilise the large amount of existing data to extract value for future plant production. Figure 1 illustrates the various components of this 'Big Data' stream that are currently logged in a typical mineral process.

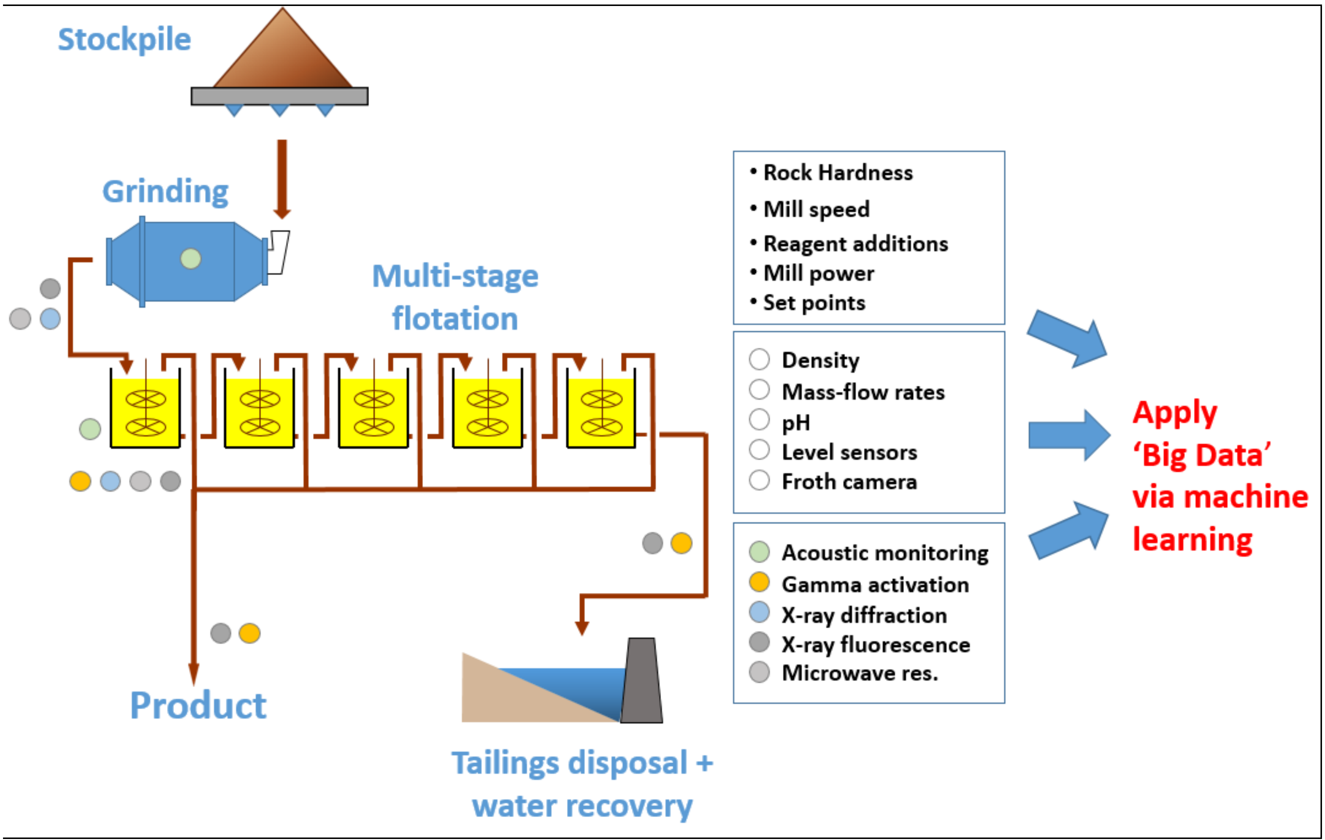


Figure 1 Process flow and sources of data in a typical minerals concentrator.

1.1 Three technology shifts

Three technology shifts are occurring to support the development of a more automated approach to holistic plant control:

- Increasingly sophisticated instruments capable of real-time measurements of chemistry and mineralogy throughout a plant. Two of the project partners, CSIRO and Thermo Fisher, are world-leaders in the development and supply of process instrumentation.
- Information networks that can gather and store thousands of pieces of data second-by-second from every part of a plant, and transmit this anywhere in the world.
- Massive-scale data storage and computing facilities, and associated 'Big Data' machine-learning software techniques for finding patterns and correlations in these data sets that can be exploited. Google is one of the pioneers in this area, showing that computers can excel at tasks, like language translation and driving vehicles, which were once considered to be exclusively human preserves.

2 Motivation and Objectives

The Centre of Excellence Project is an initial feasibility study, which aims to demonstrate the application of 'Big Data' and machine learning techniques to mineral process plant control and optimisation. Specific, original project targets included:

1. Acquisition of historical data sets covering real-time operating conditions and metal recovery from one or more large processing plants;
2. Demonstration of machine learning methods on these data sets, focussing on predicting recovery and plant upsets, and identifying features that correspond to good and bad periods of operation;
3. Identifying a target processing plant for a proof-of-concept, real-time demonstration of these machine learning methods;
4. Developing and installing a data integration platform at this plant to collect and transmit real-time information;
5. Developing a cloud-based service to analyse and report on the collected data in real-time.
6. Developing a commercial model for delivery of the service.

3 Initial steps and restart

We began by formulating a data-sharing agreement with Anglo American to obtain plant data from their platinum plant at MGK in South Africa. We were able to receive data from the entire concentrator plant as illustrated in Figure 2.

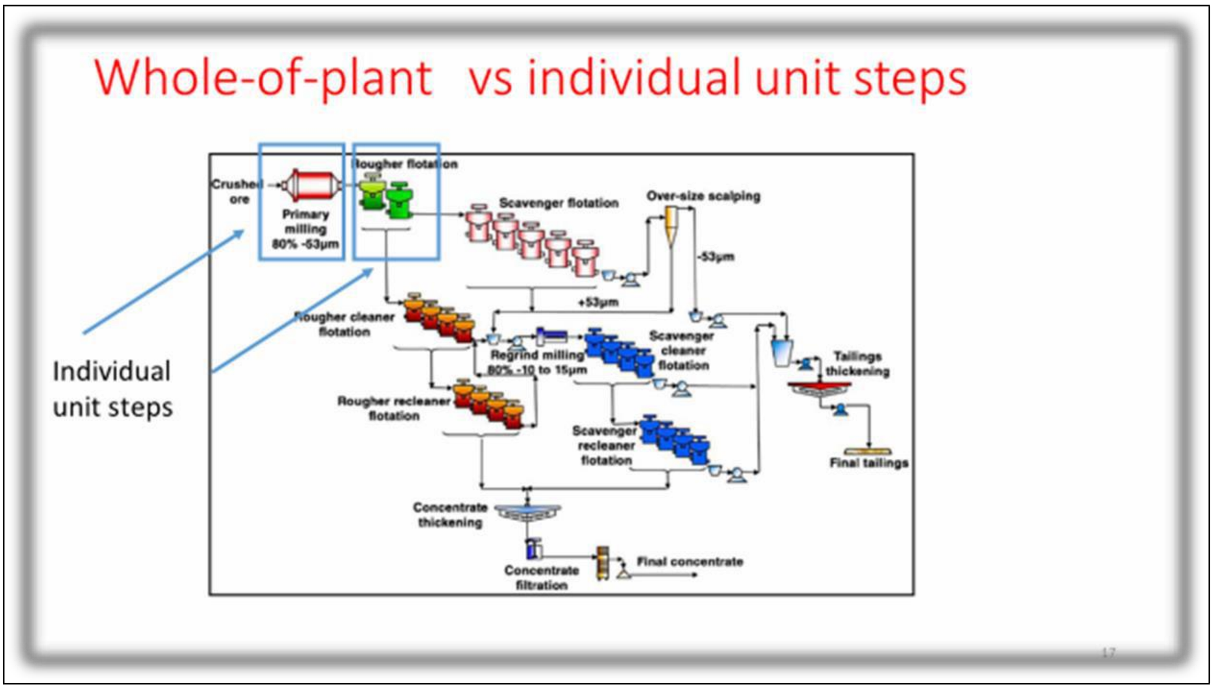


Figure 2 A complete processing plant made up of a sequence of unit steps

The historical data came from the plant's historian, a database system from OSISoft PI, which is common across the mining and other processing plants, such as water treatment. We installed PI at CSIRO to improve the access to historical data for our analysis programs.

However, we were unable to learn enough of the detailed processing steps to formulate a reasonable model of the entire plant. Moreover, there was a lot of missing data and again our lack of knowledge of the plant made dealing with that impractical.

We thus decided to apply the Big Data approach to just one unit step in a plant and to select a data-sharing partner, preferably Australian-based, who could supply plant knowledge in a straightforward manner.

3.1 Project restart with BHP Billiton

BHP Billiton came forward and offered data from its Olympic Dam plant in South Australia. In joint discussions with BHP Billiton metallurgists and plant operations management, we considered four different units steps in the Olympic Dam plant and selected the main mill, as the best candidate for data-driven modelling.

This is shown as the primary milling step in Figure 2, where the mill takes crushed ore some ten cm in size and grinds it to fine powder form which is input to the first stage of flotation.

Figure 3 The Svedala mill at Olympic Dam

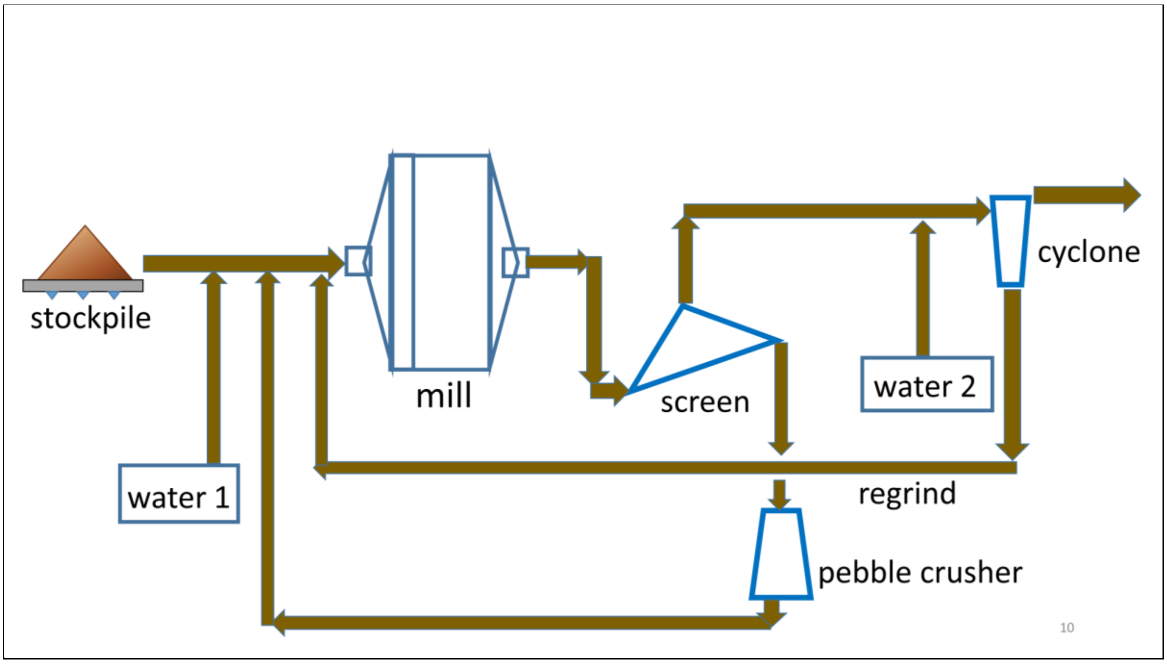


Figure 4 Flow within the Svedala mill at Olympic Dam

Several factors contributed to the success of BHP Billiton as a data-sharing partner. First, the data was provided to us in a simple text format, which we requested to be closely aligned with our data cleaning tools. Second a lot of supplementary material was provided by metallurgists at Olympic Dam, including training manuals, plant daily status reports, and tabular descriptions of the attributes of hundreds of variables in the sensor data. Finally, and most importantly, plant personnel were available to us for our questions on the data received, and later for discussion as we developed our models.

We first began modelling with six months of data. Some of the data was sampled every minute, some every hour some daily, and some every few days. We up sampled and down sampled to bring everything to a common time base of hourly data. Since initial findings suggested there was a potential for mill throughput improvement, we developed a proposal for an on-site trial at Olympic Dam, where we would provide advice on set points to improve mill throughput.

4 Data analytics

4.1 Analytics pipeline

Common to most data analytics projects, the sequence of steps begins with data acquisition, data curation, and data cleaning, as shown in Figure 5

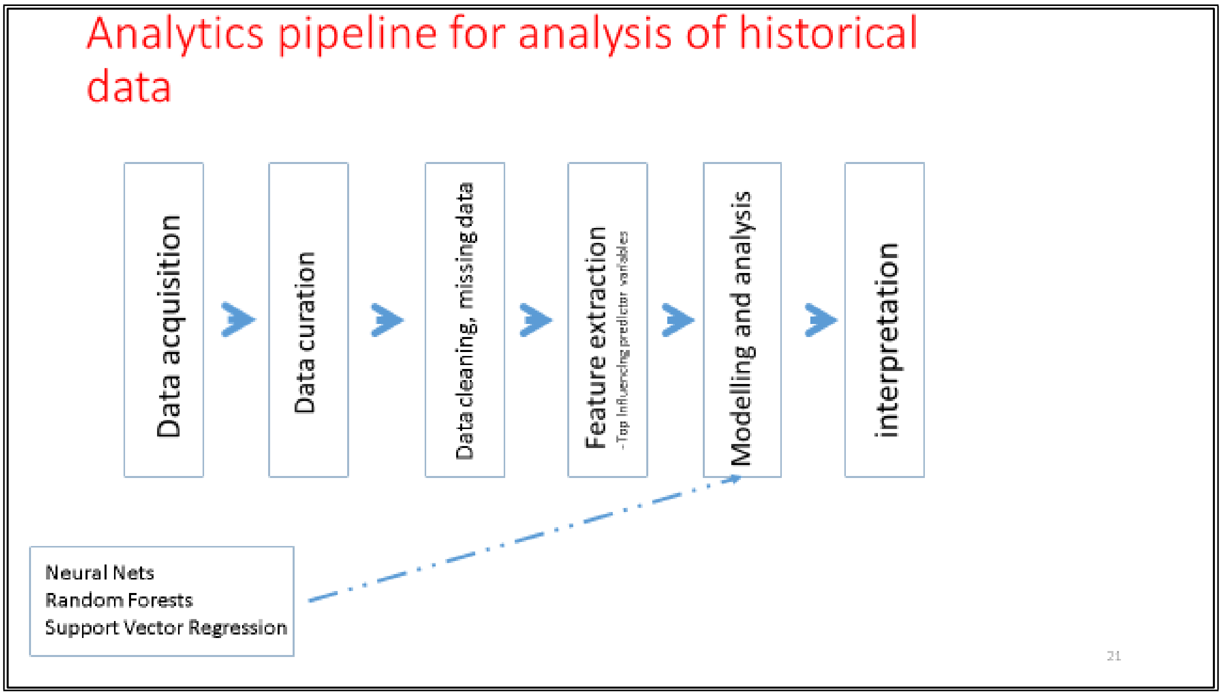


Figure 5 The usual steps in a data analytics project

The feature extraction step determines which variables are the important ones, i.e., have the most effect on the prediction. These so-called predictor variables contain sufficient relevant information from the input data, so that the desired modelling can be performed by using this reduced representation instead of the complete initial data. Some of these variables are shown in Figure 8.

We determined by hand which variables to use in our modelling. We initially drew up a list of 50 or so candidate variables and then systematically dropped them one by one to determine their contribution to the model error, eventually reaching 12 variables.

Although this was a manual process, variable importance was informed (during our initial data exploration) by a random forests model. For a future modelling project, we recommend that decision tree methods and other statistical methods be used more extensively to determine the most important predictor variables.

4.2 An alternate view of the analytics pipeline

An alternate pictorial representation to Figure 5 of the data flow we followed, is shown in Figure 6, which also shows the different types of modelling methods. The two other approaches used were, decision trees and neural networks. However, the tractability of the linear/quadratic regression methods above led to their use in most of the work.

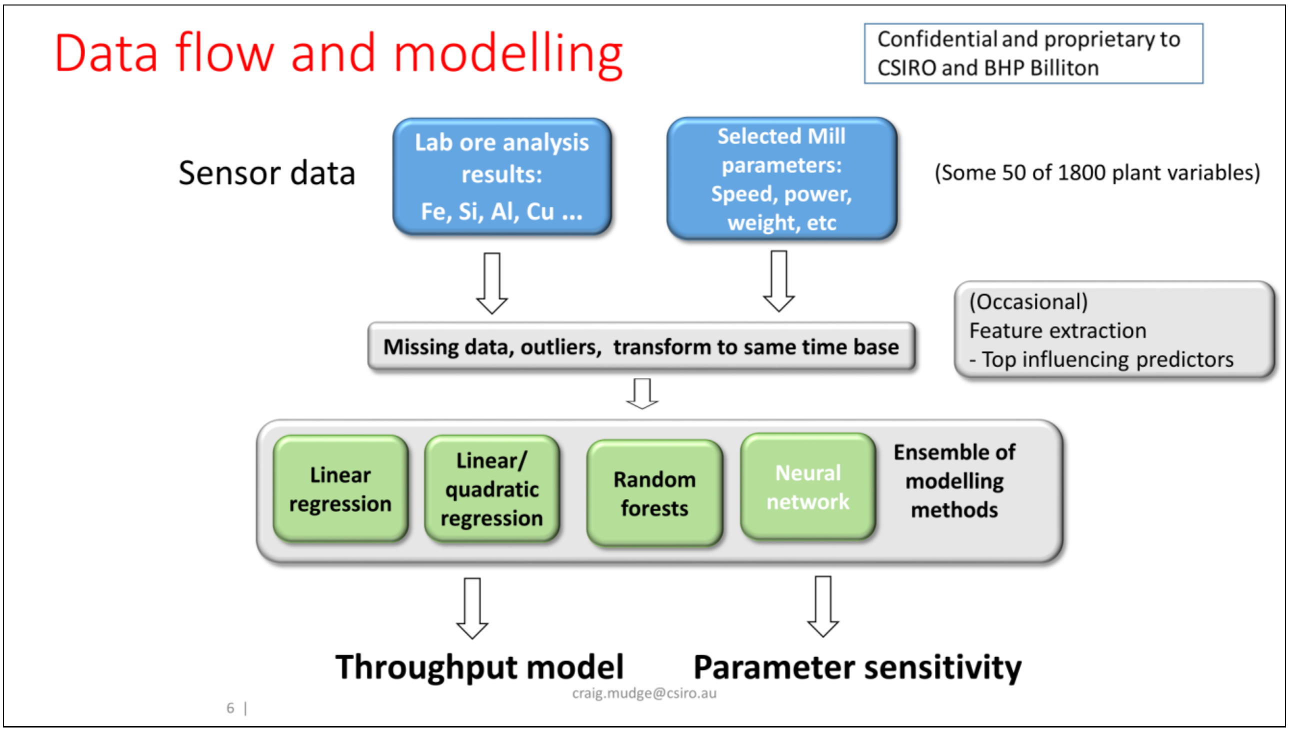


Figure 6 An alternate view of data flow in our modelling of mill throughput

5 A two-month pilot at Olympic Dam

5.1 Reasons for a pilot

To improve the chances of adoption of our results we ran a two-month pilot at Olympic Dam. We built a scalable system providing secure daily exchange of data between Olympic Dam and CSIRO and cost-effective modelling. A public cloud service from Amazon Web Services, which complied with the security policies of both BHPB and CSIRO, was the centrepiece of our IT infrastructure, and allows straightforward migration, in the future, to another BHPB mine, such as Escondida,

We decided to run an onsite pilot for two reasons

a. To improve the chances of adoption by plant personnel of our results

b. We wished to experience the stringent infrastructure requirements of an eventual real time automatic control system, which would involve the cooperative running of sections of the plant, BHPB Information Systems, and CSIRO.

Before we committed to the pilot, we added another six months (for Q4 2015 and Q1 2016) of historical data to the modelling. This improved the accuracy of our model and gave us the confidence to proceed to a pilot.

5.2 Information flow for the pilot

The information flow occurred as in Figure 7, enabled by a secure data exchange daily between BHP Billiton Olympic Dam and CSIRO.

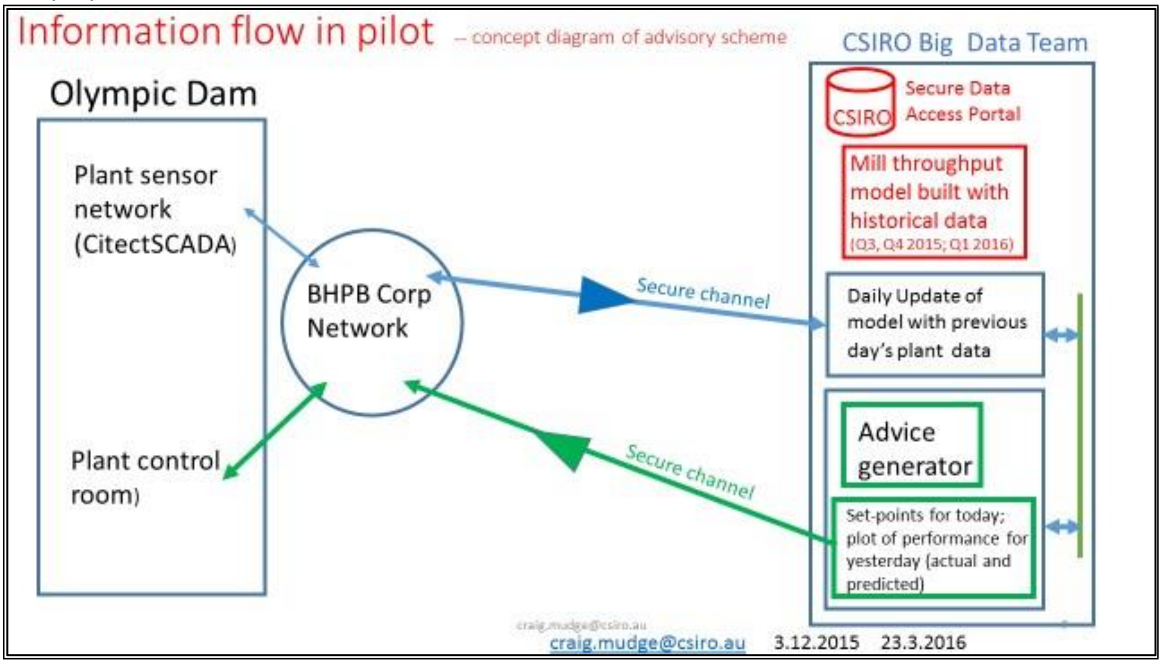


Figure 7 Information flow during daily data exchange in the on-site pilot

Some of the 50-odd variables in the daily data exchange

39	4221FI0049	flow rate of water to the mill (addition #1) -- Feed
40	4221FI0050	flow rate of water to the cyclone (addition #2) - discharge
41	4221FIC0049VS	WaterFl#1SP
42	4221G.FE_SIO2_RT	Feed Prep Pond2 Fe:SiO2 Ratio
43	4221II0178	Cyclone Feed Pump VFD Load Current
44	4221II0415_M	Pebble crusher speed
45	4221IIC0415_MVS	4221IIC0415 Operator Set Point
46	4221IIC0415VS	... Setpoint 4221IIC0415 Operator Set Point
47	4221NY0903_V	Svedala Mill Feed Dry Mass (mill feed (feed rate of solids to the mill))
48	4221NYQ0750.D	Svedala Mill Total Power Daily
49	4221NYQ0751.D	Svedala Unit Power Consumption (power draw)
50	4221PI0023	cyclone feed pressure
51	4221PI0023KS	... Setpoint 4221PI0023 Cyclone Pressure Cube Setpoint
52	4221SI0751	MillRPM - mill speed
53	4221UE0001A	Svedala Cyclone UF Density
54	4221UE0066A	Svedala Cyclone OF Density
55	4221UE0415A	Svedala Pebble Crusher Gap
56	4221WI0011	Pebble crusher conveyer rate
57	4221WI0285	Svedala Mill Weight (Load)

Figure 8 Some of the variables selected for modelling mill throughput

6 IT Infrastructure, security, and software engineering practice

This section discusses the infrastructure we designed and built as a scalable system providing secure data exchange on a daily basis between CSIRO and Olympic Dam as well as cost-effective modelling. The components of the main cloud-based system are given in Figure 9.

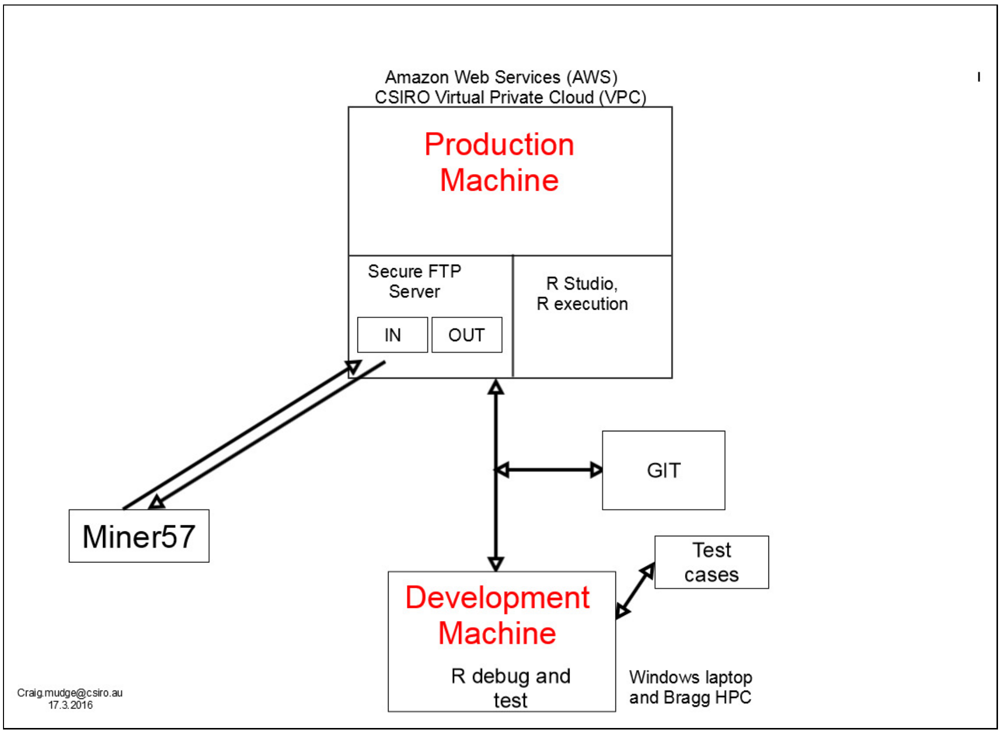


Figure 9 Our arrangement of production and development machines and cloud service provider in the Big Data project. Miner57 is the code name given to BHP Billiton

6.1 Compliance with security policies of both BHP Billiton and CSIRO

The two industries which traditionally are the most security conscious are pharmaceutical (drug discovery) and minerals and energy (exploration data). There are now examples of firms from both industries using Amazon Web Services (AWS).

Our off-site processing and storage uses CSIRO's Virtual Private Cloud (VPC), located in the Asia Pacific (Sydney) region. A VPC is formed by AWS dedicating a secure network sub-net to a single Amazon customer. The CSIRO VPC is accessible only to users with CSIRO nexus credentials over a secure link.

The Australian Signals Directorate maintains an ASD Certified Cloud Services List, which shows cloud service providers that have been through their Information Security Registered Assessors Program (IRAP); the list is at http://www.asd.gov.au/infosec/irap/certified_clouds.htm

6.2 Advantages of cloud computing

The advantages of using a public could computing service, such as Amazon Web Services (AWS) are

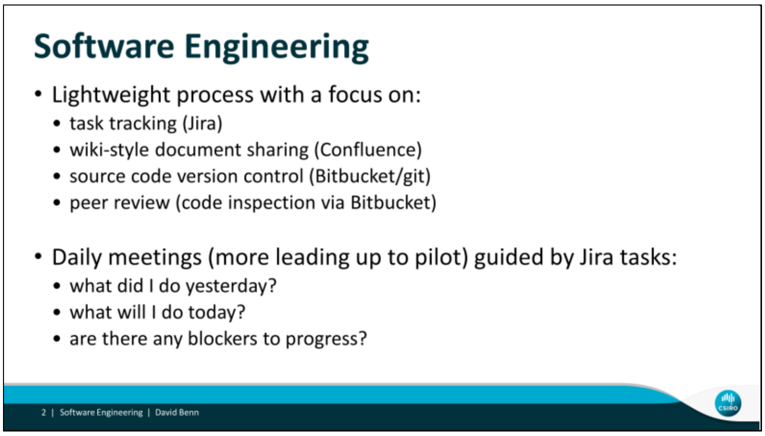
a. It provides a high-availability compute and storage infrastructure, which is essential for servicing a daily advisory service during the April-June pilot with Olympic Dam. Availability applies to software licenses as well as hardware.
b. We can quickly move to parallel execution when our modelling requires it, because of the map- reduce framework supported by AWS in Spark, Hadoop, and other distributed computing frameworks.
c. Amazon Machine Images (AMIs) pre-package a system ready for one to boot on one's own custom virtual server. Ours have a base operating system, such as Ubuntu, with pre-configured extra software, built by CSIRO. We use one targeted at R and RStudio Server, which required only a few minutes to get going on CSIRO's VPC.
d. Given that mineral processing analytics as a service is a potential commercialisation path for CSIRO, cloud-based infrastructure provides a low-entry cost for a CSIRO spinoff. Moreover, it is a readily scalable delivery model that could be sold around the world.
e. It is a straightforward matter to replicate the modelling, daily receipt of sensor data, and advice generation at another BHPB mine, such as the Escondida mine in Chile.

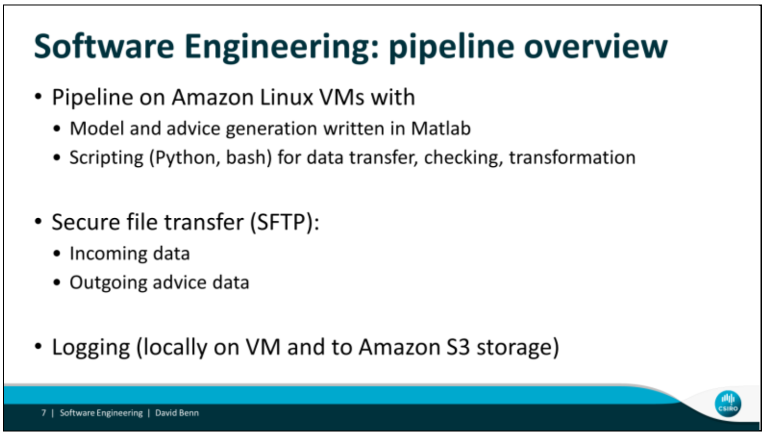
6.3 A separate repository for document collaboration

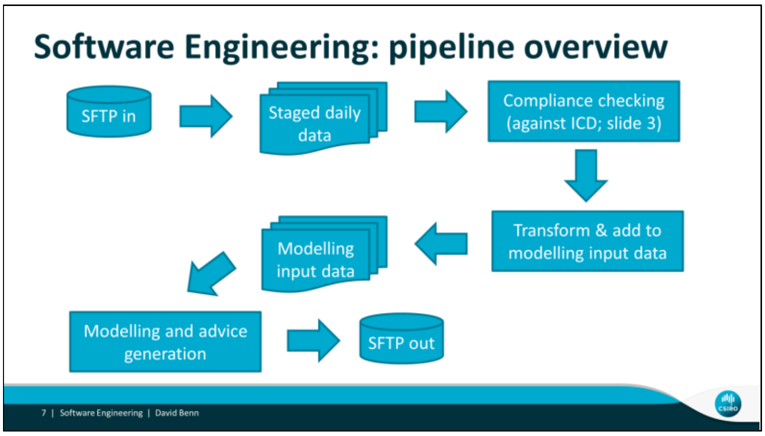
Data exchange of sensor data and advice generated from our Big Data modelling is achieved through the Secure FTP mechanism shown on Figure 9. However, a quite separate facility, complying with the security policies of both CSIRO and BHPB, was needed to facilitate sharing of technical papers, videos and co- authoring of reports. This has been achieved by BHPB providing CSIRO user accounts on their SharePoint system.

6.4 Software engineering practice

The following two slides extracted from David Benn's presentation at the June 21 final meeting, are a good summary of the practices we followed.



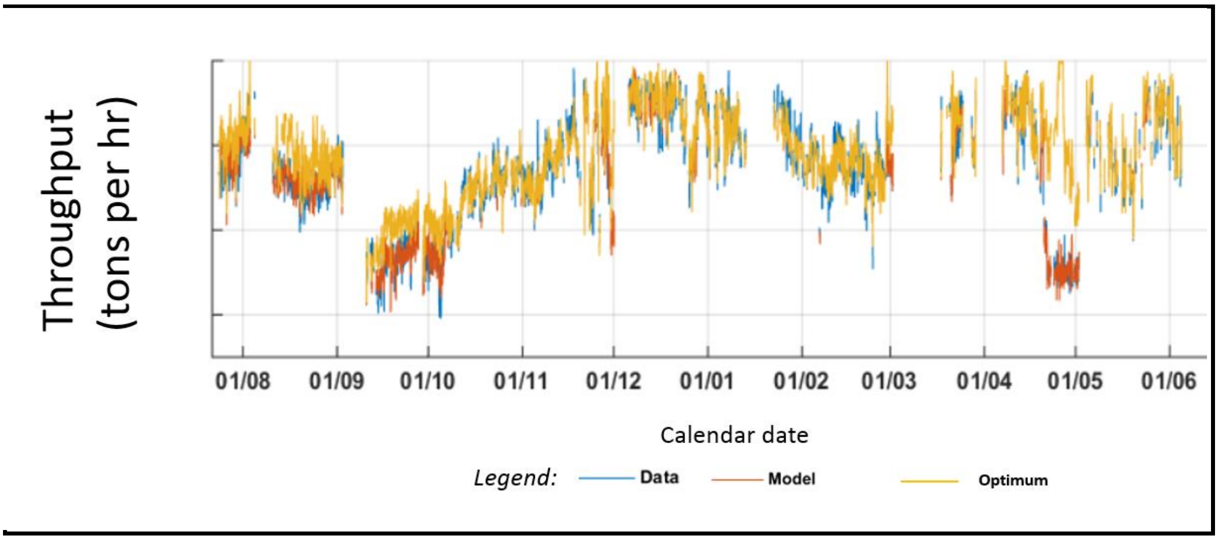


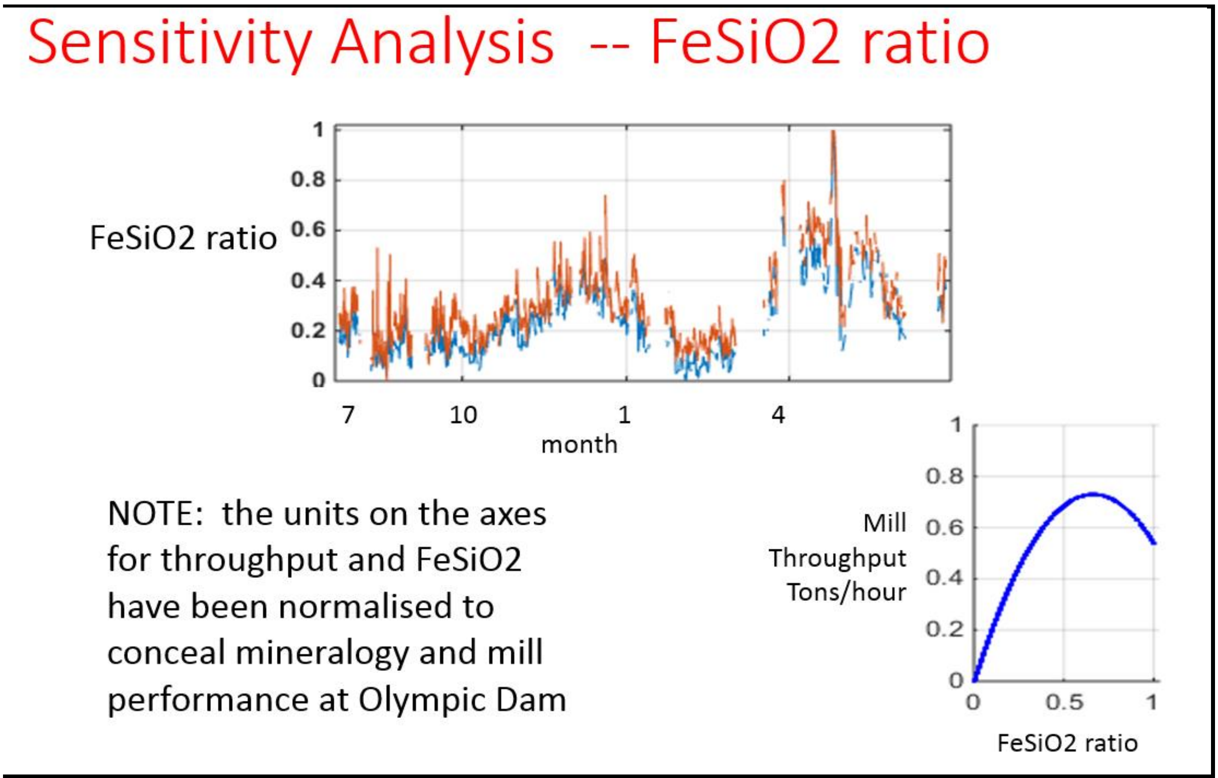


Note that context for these slides is **Figure 9 "Our arrangement of production and development machines and cloud service provider in the Big Data project."**

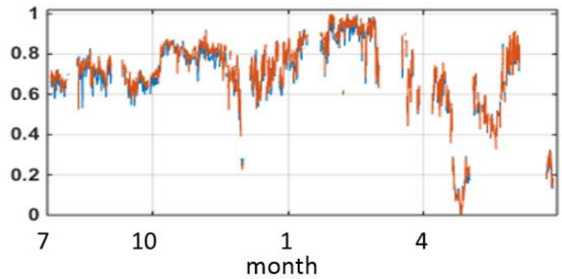
7 Results

The following short section, extracted from the Confidential Chapter 7 of the internal BHPB-CSIRO Report, summarises our findings.

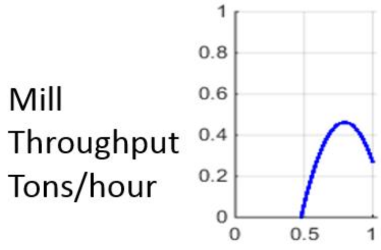




Sensitivity Analysis -- Mill fill



NOTE: the units on the axes for throughput and FeSiO2 have been normalised to conceal mineralogy and mill performance at Olympic Dam

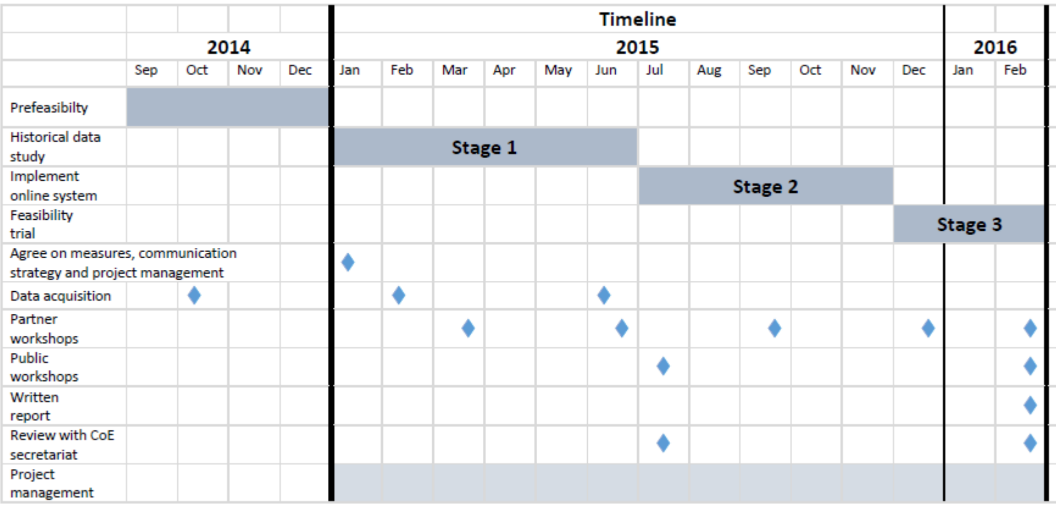


As BHP Billiton's Steve Liddell noted at our June 21 workshop, "the sensitivity analysis revealed the key drivers. It confirmed some of our intuitions and highlighted others that were more important than we had thought."

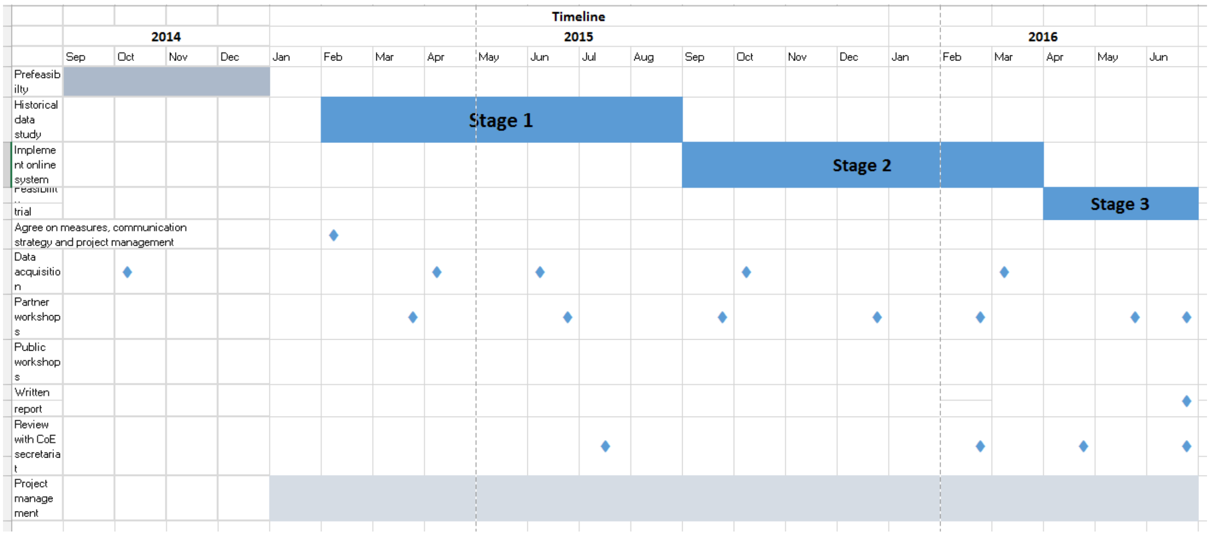
8 Appendices

Appendix 1 Project Implementation Timeline

The original timetable proposed in our Application to the CoE was as follows.



The actual timetable, incorporating a four-month project slip, is as follows.



CONTACT US
t 1300 363 400
+61 3 9545 2176
e csiroenquiries@csiro.au
w www.csiro.au

AT CSIRO, WE DO THE
EXTRAORDINARY EVERY DAY

We innovate for tomorrow and help improve today – for our customers, all Australians and the world.

Our innovations contribute billions of dollars to the Australian economy every year. As the largest patent holder in the nation, our vast wealth of intellectual property has led to more than 150 spin-off companies.

With more than 5,000 experts and a burning desire to get things done, we are Australia's catalyst for innovation.

CSIRO. WE IMAGINE. WE COLLABORATE. WE INNOVATE.

FOR FURTHER INFORMATION
CSIRO Mineral Resources
Dr Craig Mudge AO
t +61 8 8303 8621 | Mob: +61 417 679 266
e craig.mudge@csiro.au
w www.csiro.au/cmr