Translation of stranded IP to commercialisable software applications using Workspace

P. W. Cleary[a], L. Hetherton[a], D. Thomas[a], M. Bolger[a] and D. Watkins[a]

[a] CSIRO Data61, Australia
Email: Paul.Cleary@csiro.au

Abstract: For Australia and many other countries, which have large gaps between their research inputs and economic outcomes from their national innovation systems, understanding where and in what form the IP generated resides represents a significant opportunity. Aside from publication outputs (which are usually the primary measure of research excellence) and registerable IP (patents etc) which form only a small amount of the IP generated, there is significant development of enabling tools and data. However, such IP (termed "stranded IP") can often have value to third parties if these opportunities can be identified, the IP put in a suitable form and then provided to them for use.

Usage of research outcomes and research software can be highly valuable to the right customers if they can be provided in a usable form that is fit for their usage purpose in a cost effective and timely manner. Research IP is typically developed for specific research purposes and projects (most commonly with the aim of publication) with opportunities for translation to third parties either not considered or identified significantly after the form of the IP has been determined.. The form of the IP is usually decided by individual researchers or teams based on what suits their current needs rather than strategically or with the intention of translation or later commercialisation.

It is also common for enabling IP to be part of a larger process or workflow which usually then also involves multiple manual or semi-manual steps. In the modelling and simulation space, this is particularly common in relation to the preparation of model inputs and the analysis, preparation or presentation of outputs. These represent substantial barriers to translation, as potential users judge the cost of expert knowledge acquisition to be too high and the distortions to their internal workflow and decision making processes to be too costly and/or risky.

It is useful to consider research activities as workflow processes, being composed of a series or network of unit operations linked with input-output relationships and execution order dependencies. Typically some or many of these operations or steps will use software components (often including proprietary or open third party sources). The intervening manual steps present strong hurdles to the automation of these semi-manual workflows which inhibit reproducibility and transfer to third parties. It also allows errors in the use of the IP to be introduced. An effective solution is to convert all steps into workflow unit operations that can be manifested in a scientific workflow system. This can facilitate high levels of component re-use, improved collaboration, interoperability of the software components in an extensible and customisable way. This provides the ability to adapt to new and emerging requirements and to customise IP on a per customer basis at comparatively low cost.

This paper explores the nature of stranded IP and the obstacles that limit its exploitation. It then explores how the adoption of a software workflow platform, Workspace in this instance, can overcome a sufficient subset of these obstacles so to provide a low cost pathway from internal research to exploitable software applications and products. A methodology is proposed, based on the capabilities of Workspace, for taking legacy, enabling or stranded IP, particularly in the modelling and simulation space, and building translatable and commercialisable products.

Keywords: Workspace, stranded IP, commercialisation, IP translation pathway, workflow

1. INTRODUCTION

There are many obstacles to the translation of research activities and research software into distributable and commercialisable applications. Addressing such obstacles is important for improving the transfer of research into economically useful outcomes and an important path to impact that can justify funding for such research. Cleary et al. (2017) previously considered connections between the operation and efficiency of the national innovation system, the generation of significant volumes of stranded enabling IP and how translation of this IP to third party and commercial use can be facilitated using workflow approaches to software architecture.

This paper seeks to explore some aspects of the nature of stranded IP and obstacles to its exploitation and considers how the adoption of a software workflow platform such as Workspace can moderate a range of these obstacles by providing a low cost pathway from internal research outputs to exploitable software applications and products. A methodology is proposed, based on the capabilities of Workspace, for taking legacy, enabling or stranded IP, particularly in the modelling and simulation space, and building translatable and commercialisable products. This is based on the experience of several years of usage in CSIRO in around 50 research projects and areas.

2. MEASURING INNOVATION

The Global Innovation Index (GII) is an annual assessment of the relative strengths of the innovation systems of most countries in the world. It provides high level measures of the relative inputs and outputs of the national innovation systems and tracks their relative change from year to year. Australian innovation performance was ranked 23, 20 and 22nd (all rankings out of around 127 countries) by the Global Innovation Index (GII) in 2017-2019 (Cornell University, INSEAD, and WIPO, 2017, 2019). The Innovation Efficiency Ratio (IER) is given by the ratio of the Innovation Output to Input sub-indices. In 2017, Australian innovation efficiency in 2017 was 0.6 which ranked 76 (with Innovation Input ranked highly at 12th but output ranked only 30th) (Cornell University, INSEAD, and WIPO, 2017). In 2019, Australia's IER was lower at 0.57 (with Input rank declining to 15th while output rank declined to 31st). These rankings fluctuate year by year but do so within consistent ranges (Office of Chief Economist, 2017) and so give a consistent long term picture of relatively high levels of investment in research but with relatively low levels of translation into economic outputs. In contrast, Australia ranked 9th in terms of publication outputs and 10th for Citable Document H-index in 2017 which was little changes in 2019 with a ranking of 10th for each measure (Cornell University, INSEAD, and WIPO, 2017, 2019) suggesting a stronger focus on measurable publication outputs than on commercial translation of the underpinning IP (Intellectual Property). Output components most relevant to scientific IP translation are IP receipts and ICT services exports as a % of trade for which Australia ranked 29th and 89th respectively. Countries that perform best from an innovation efficiency perspective are Switzerland, Luxembourg and China whose IER were 0.96 to 0.92 in 2018. The USA and UK, whilst having larger innovation footprints have only intermediate levels of IER (0.76-0.77) with rankings of 21 and 22 respectively. In terms of publication outputs their normalised publication rates are somewhat lower than for Australia but they are equal first for Citable Documents H-index. For the most innovation efficient countries, publication outputs in terms of H-factor vary with Switzerland being stronger than Australia's (2nd and 9th for the two measures), China being 42 and 14th while Luxembourg was ranked 49 and 73th respectively. This could support a conclusion that a strong focus on publication could at least mildly correlate with weaker innovation efficiency.

ATSE (Australian Academy of Technology and Engineering) (2016) provided analysis of Australian innovation performance is primarily focused on policy structures, funding mechanisms, the business-research interface and the resultant outcomes in comparison to those of other advanced economies. A discussion paper (Australian Government, 2014) identified general factors that influence success of translation of public research into commercial outcomes:

1. Research excellence (as measured by publication based metrics)
2. Targeted research effort (meaning research investment directed into specific areas, typically relating to current or emerging perceived strengths)
3. Cooperation between researchers and industry (as measured by the proportion of businesses collaborating with research enterprises and the relatively high proportion of researchers working outside businesses)
4. Entrepreneurship (where "Countries that excel in innovation tend to exhibit a high degree of entrepreneurship in both the research community and industry"). This is related to the predisposition of researchers to be prepared to take their research outcomes into commercial ventures and to try to scale-up uptake of the technologies via appropriately structured commercial vehicles or arrangements.

This top down view provides evidence of relative under-performance in Australia for the translation of research into commercial outcomes but it does not provide understanding for the reasons that cause this to occur. As discussed by Cleary et al. (2017), they do not typically consider questions around the nature of the outcomes of large scale research investment or their form (that is, "What did all the investment end up creating?") or about the obstacles to translation of such IP into commercial outcomes. They also do not consider mechanistic questions at the researcher or research team level (the levels where most operational research decisions are made which control the form and nature of the IP generated) as to why they make the types of decisions that they make. These decisions made by individual researchers or teams frequently result in significant IP being created within research projects that is either not translated or is not in a form that is suitable for translation, which we term "stranded IP". Where do the research inputs go?

Cleary et al. (2017) observed that the accumulated difference between the input and output sub-indices of the GII (Cornell University, INSEAD, and WIPO, 2017, 2019), could be interpreted as an IP output gap. Integrated over the life of the IP, this represents a potentially very large pool of sunk investment that is not usually considered commercially exploitable. Some of this IP may be re-used in current or future projects by some researchers or teams. There is no currently available data to suggest what fraction of this stranded IP is currently being internally re-used within research entities or how much of this may be able to be exploitable if sufficiently low cost translation pathways were available. Since the size of the accumulated under-exploited IP pool is extremely large, even minor improvements in the amount of IP exploited could generate substantial improvements in measured translation outcomes.
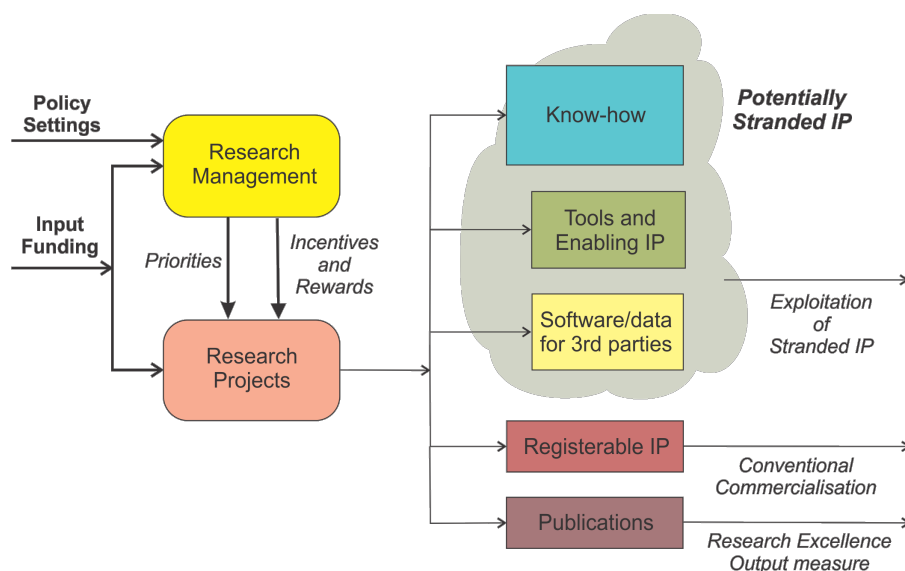
Figure 1. Flow chart showing idealized process of IP generation within research projects.

Research projects are the operational level where research activity is performed and where IP is generated. Figure 1 shows a conceptual model of IP generating research activities with inputs of funding and policy setting influencing the behaviours of research management and research projects in how they spend their research funding. Research projects are often planned with a view that publications are the dominant IP output (as indicated by the comparatively high ranking that Australia has for Citable Document H-factor at 10th in 2019). This is heavily influenced by research quality metrics at both organisational and individual levels that are predominantly focused on publication outputs. Some registerable IP (e.g. patents) are also produced but typically these are a small fraction of output IP. Much of the effort expended goes into developing "know-how" and intermediate enabling IP (such as numerical or experimental methods, data, computer software for simulation, pre- or post-processing of data or visualisation). This is sometimes regarded as Trade Secret, but more commonly is not explicitly thought of as output IP, is not tracked and is not subject to meaningful effort to translate this into any form of commercial outcome. It would not be uncommon for a significant fraction of research costs to be expended in creating these types of enabling or internally held IP. This is typically not created with the intention of either future re-use or commercialisation, and is almost always in a form that is not suitable for commercialisation. Consequently, a large proportion of the know-how and enabling technology is stranded in the sense that it has no obvious viable path to external exploitation. Cleary et al. (2017) argued that a reasonable fraction of the national research output gap manifests in the creation of such stranded IP and showed how use of a workflow methodology could ameliorate these problems.

3.   BARRIERS TO TRANSLATION OF STRANDED IP

Most consideration of barriers to IP translation or commercialisation focuses on issues relating to IP protection (IP registration such a patents, copyright and trademarks) and impediments to crossing the research-business interface. However, critical barriers often relate to the form of the IP and to the interest (or lack thereof) of the IP generators (researchers) for attempting such translation. So the research translation process often fails well before the more broadly considered system level barriers are able to influence the outcome.

Increasingly, much of the enabling technology-related stranded IP is in the form of software or could be packaged and distributed in software form (for example AlteTreat, 2015). Components of this stranded IP are often useful to others (either other researchers or commercial entities that would benefit from the use of more leading edge IP than is available in existing commercial solutions). Cleary et al. (2017) identified five key obstacles to researchers or their organisations providing this IP in some form of usable product, including:

1. Lack of interest from the researchers in translation of research outputs into economic value.
2. Lack of expertise in the creation and support of software applications.
3. The form of the IP being unsuitable for use by third parties, often because of complex and/or unintuitive interfaces and the existence of manual steps in the underlying workflows.
4. Functionality being insufficiently close to what is needed therefore requiring adaption or extension.
5. Use of third party libraries and tools where licensing conditions are either unsuitable or not understood.

Obstacle 1 is a policy and management related issue in which incentives, rewards and operating structure are important. Solutions to this are not related to the provision of improved tools for translation. Obstacle 3 results from the IP having been developed to the minimum level required to be able to deliver on the specific project goals, typically not considering later re-use or use by anyone other than the project team. Project goals are often narrowly conceived in order to constrain time and cost which means that there are usually gaps when compared to third party requirements for re-using this IP.

The key consequence of obstacles 2-5 is that the cost to develop a commercial or externally usable version of any such software tools is usually prohibitive and/or the timescale for such creation is too long. Reliable estimation of the potential revenue from commercialisation can also be very difficult to make. Market evaluation also usually requires that the software already exists so that it can be usable and tested. Unsurprisingly, only limited opportunities with very high perceived value are typically commercialised.

## 4.   WORKSPACE – A WORKFLOW PLATFORM FOR COMMERCIALISATION

Workspace (Workspace, 2014; Cleary et al. 2019) is a workflow platform that has been specifically developed to support commercialisation. It does this in several ways. Firstly, it is built with a very narrow range of third party dependencies which all have LGPL (Lesser General Public License) or more permissive licensing conditions so there are no inherited licensing restrictions on commercial exploitation of anything built on Workspace giving broad freedom to operate (obstacle 5). Workspace itself is licensed freely for non-commercial application (which includes internal usage by corporate entities and any free provision to third parties of applications built using Workspace). Secondly, it supports open and closed source deployment with developers able to decide whether they supply workflows, workflows with GUI components or full applications that are compiled as closed standalone executables (thereby enabling protection of trade secret IP embedded within the application). Thirdly, it supports the inclusion of license management into applications. Fourthly, the ability to easily attach custom GUI elements to workflows provides a low cost path to tailoring user experiences to customers' domain-specific requirements.

The construction of workflows (networks of operations) and applications from collections of operations with a clear set of inter-connection requirements inherently forces modular design and development on users. This is particularly important for providing both intrinsically better structured workflow code for non-software engineering qualified users and reducing the software skill level required to produce quite complex near-commercial-level software applications (obstacle 2). A common variant of obstacle 3 is that the software does not have a usable or friendly interface, instead being controlled by complicated and opaque text input files for example. Significant improvements can be made to usability of software by attaching easy to use but powerful GUI's which is very well supported by Workspace. The easy adaptability of workflows means that the scope of applications can be changed at relatively low cost to meet new or alternative requirements for third parties who would value usage of this IP (obstacle 4).

A basic summary of Workspace and a comparison to other Scientific Workflow Systems is given by Watkins et al. (2017). Thomas et al. (2019) provides an update on recent developments that support the creation of more complex workflow solutions. The broader space of tools, reusable workflows and applications built in Workspace are termed the Workspace ecosystem. The use of Workspace in several use cases relating to modelling and simulation, particularly in the commercial deployment of research derived software applications is described in details in Cleary et al. (2019). Examples of externally deployed applications built using Workspace include ArcWeld (a welding simulation application, Murphy and Thomas (2017, 2019)), AlteTreat (for managing heat treatment of castings, AlteTreat (2015)), Spotsizer (Bischof, 2016), Grainscan (Whan, 2014), Amicus (bush fire prediction, Sullivan et al. 2013)), Flood prediction (Hilton et al. 2015), SPARK (bushfire simulation, Miller et al. 2015), Oventus (for customizing asleep apnea device, Oventus, 2016) and Dive Mechanic (Cohen et al. 2017). Worked examples are given in Thomas et al. (2019).

5. PROCESS FOR CREATING SOFTWARE APPLICATIONS FROM STRANDED IP

The following process is divided into a sequence of specific steps, several of which have a number of sub-steps. It is shown as a flow sheet in Figure 2. This process was developed over several years as a range of different sets of legacy code were developed into exploitable forms using the Workspace platform (for example). It is worth noting that much of the legacy code was stranded (by which we mean it had no identifiable viable commercialisation or translation path to customers or users and/or was too difficult and unfriendly to be used by their potential user base).
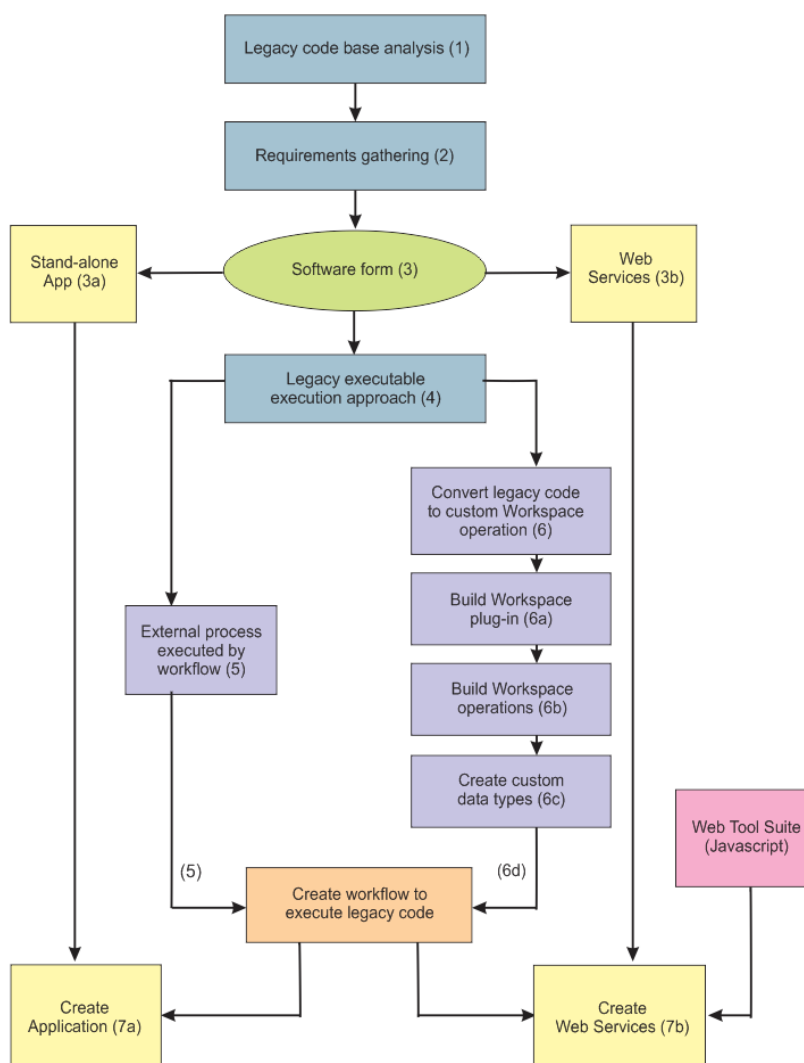
Figure 2. Flow sheet shows key steps and their relationships for building user-friendly software applications from legacy code using Workspace.

1. Identify the types of interfaces that exist to the legacy codebase (there may be more than one type of interface). Generally speaking, most substantial legacy research applications take the form of a monolithic executable (1b) that consumes input files/data and produces output files, so this is the type we will cover in more detail here. Specific tutorials covering the component aspects of (1a) and (1c) exist in the Workspace core tutorial set, (Workspace, 2014).
2. Once the form of the legacy code is identified, it is essential that a requirements gathering process is undertaken in order to ascertain:
    a. Detailed input interface requirements, such as file formats, data types etc, as well as the form and format of the output data
    b. Data visualisation and reporting requirements
    c. If a custom GUI is required, then a preliminary design process is undertaken in which wireframes are drafted for each significant screen or page of the application
3. It is also critical at this juncture to determine the specific form that the software will take. Generally, there are two approaches available for creating software suitable for commercial delivery and/or use by a typical engineering user (as opposed to a highly experienced domain expert such as those who originally develop the IP) who may not have specific detailed expertise in the application. Either:
    a. The software is delivered as a standalone, native software application (e.g. an .exe and a number of .dlls on Windows). Workspace provides built-in capabilities to expedite the creation of such an executable, or
    b. The software is delivered as a web-based application. Although this is often the more immediately popular choice, customers often become hesitant once challenges associated with hosting and long-term maintenance are discussed.
4. Decide on the approach desired for execution of the monolithic executable from a workflow. There are generally two possible approaches:
    a. Use OOTB (Out Of The Box - meaning that they come bundled with the Workspace platform) operations in a workflow.
    b. Wrap the external process (and/or input/output generation processes) within custom Workspace Operation(s).
5. If proceeding with (4a), at this point, a workflow (or set of workflows) is created.
6. If proceeding with (4b), then Workspace plugin is created along with a suite of data types and operations to wrap the codebase
7. Once (5) or (6) are suitably advanced, if a GUI front-end was selected (2c) then developer can start to put this together. Its form necessarily depends on the choices made at (4):

This process for using the Workspace ecosystem tools to create applications suitable for third party use from stranded IP is elaborated on in more details in Hetherton et al. (2019). This process now exists and has been shown repeatedly to be able to deliver the required IP translation, but this is not intended to imply that other processes or variations of the process could be followed successfully for different scenarios.

6. CONCLUSIONS

The Workspace software ecosystem provides low cost pathways for the translation of existing (stranded) IP and software developed for internal usage into products for use by third party customers. The process and drivers to produce commercial-grade third party usable applications deployed as either standalone products or as web services, from existing or stranded IP using the Workspace workflow platform has been explored. It describes a now well tested process for taking valuable in-house IP that cannot be externally used (that is, stranded) and creating commercialisable software. It achieves this by:

1. Reducing development cost (by enabling significant operation re-use).
2. Increasing agility, flexibility and ability to customise (as software design and capabilities can be changed by adding/removing parts of the operation networks).
3. Increasing user friendliness with easy attachment of user interfaces to the workflows.
4. Reducing re-development time thereby decreasing time to market.
5. Supporting improved programming and architecting of such software, particularly by enforcing modular software design and development which reduces the software engineering skill level required to create commercialisable software applications.

The difficulties of translation of research IP into commercial benefit were analysed in the context of the lower than desirable levels of translation efficiency by the national innovation system. An IP taxonomy for research IP was used to understand the scale and reasons for stranding of so much research output.

REFERENCES

AlteTreat, (2015), https://research.csiro.au/workspace/altetreat/

Australian Government, (2014), Discussion paper "Boosting the commercial returns from research", https://submissions.education.gov.au/Forms/higher-education-research/Documents/Boosting%20Commercial%20Returns%20from%20Research%20%20-%2024102014.pdfATSE, (2016), "Research Engagement for Australia: Measuring research engagement between universities and end users", https://www.atse.org.au/content/publications/reports/industry-innovation/research-engagement-for-australia.aspx

Bischof, L., Prevorovský, M., Rallis, C., Jeffares, D.C., Arzhaeva, Y., and Bähler, J., (2016), Spotsizer: High-throughput quantitative analysis of microbial growth BioTechniques, 61, 191-201.

Cleary, P. W., Watkins, D., Hetherton, L., Bolger, M. and Thomas, D., (2017), Opportunities for workflow tools to improve translation of research into impact, 22nd International Congress on Modelling and Simulation (MODSIM 2017), Hobart, Tasmania, Australia, 3-8th December.

Cleary, P. W., Thomas, D., Hetherton, L., Bolger, M., Hilton, J. E., and Watkins, D., "Workspace: a workflow platform for supporting development and deployment of modelling and simulation", revised for: Mathematics and Computers in Simulation, (2019).

Cohen, R. Harrison, S. M., and Cleary P. W., (2017), Dive Mechanic: Bringing 3D virtual experimentation to elite level diving using the Workspace workflow engine, MODSIM 2017, Hobart, Australia, December.

Cornell University, INSEAD, and WIPO, (2017), The Global Innovation Index 2017: Innovation Feeding the World, Ithaca, Fontainebleau, and Geneva. ISSN 2263-3693. ISBN 979-10-95870-04-3.

Cornell University, INSEAD, and WIPO (2019),The Global Innovation Index 2019: Creating Healthy Lives—The Future of Medical Innovation, Ithaca, Fontainebleau, and Geneva. ISSN 2263-369. ISBN 979-10-95870-14-2.Hetherton, L., Thomas, D., Bolger, M., Watkins, D., and Cleary, P. W., (2019), How to use Workspace to convert legacy modelling and simulation codes into productive, customised, user-friendly, commercial quality software applications, submitted to: MethodsX.

Hilton, J., Miller, C., Bolger, M., Hetherton, L., and Prakash, M., (2015). An Integrated Workflow Architecture for Natural Hazards, Analytics and Decision Support, in: Environmental Software Systems. Infrastructures, Services and Applications, IFIP Advances in Information and Comm. Tech., 448, 333-342.

Miller, C., Hilton J., Sullivan A. and Prakash M., (2015). SPARK–A Bushfire Spread Prediction Tool, R. Denzer et al. (Eds.): ISESS 2015, IFIP AICT 448, pp. 262–271.

Murphy, A. B. and Thomas, D. G. (2017), A computational model of arc welding – from a research tool to a software product. Paper presented at the 22nd Int. Congress on Modelling and Simulation, Hobart, Australia,

Office of Chief Economist, (2017),

https://industry.gov.au/Office-of-the-Chief-Economist/Publications/IndustryMonitor2015/section2.html

Oventus, (2016), https://www.csiro.au/en/Research/MF/Areas/Metals/Lab22/Mouthguard

Sullivan, A., Gould, J., Cruz, M., Rucinski, C., and Prakash, M., (2013). Amicus: A national fire behaviour knowledge base for enhanced information management and better decision making, 20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013.

Thomas, D., Hetherton, L., Bolger, M., Watkins, D., and Cleary, P.W., (2019), Example workflows for use of Workspace in modelling and simulation applications, submitted to: Data In Brief.

Thomas, D. et al. (2019), Workspace – a scientific Workflow System with Commercial Impact, MODSIM 2019, Canberra, Australia, December 2019.

Watkins, D., Thomas, D., Hetherton, L., Bolger, M. and Cleary, P.W., (2017), Workspace - a Scientific Workflow System for enabling Research Impact, MODSIM 2017, Hobart, Australia, December 2017.

Whan, A. P., Smith, A. B., Cavanagh, C. R., Ral, J-P. F., Shaw, L. M., Howitt, C. A., and Bischof, L., (2014), GrainScan: a low cost, fast method for grain size and colour measurements, Plant Methods, 10:23

Workspace, (2014), DOI: http://dx.doi.org/10.4225/08/54D03170101B7