Opportunities for workflow tools to improve translation of research into impact

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

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

Abstract: There are many obstacles to the translation of research activities and research software into distributable and commercialisable applications. This 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. Usage of research outcomes and research software can be highly valuable to the right customers if it is deployed to them fit for purpose and easy to use in a timely and cost effective way. Typically research IP is initially developed to target specific research outcomes (most commonly with the aim of publication) and wider opportunities for translation are only identified well after the form of the IP has been set, for example after a project has been completed. It is also common for the use of the IP to involve manual and semi-manual steps, particularly relating to the preparation of inputs and the analysis and/or preparation of outputs, often requiring intimate knowledge of the underlying technology. These represent substantial barriers to entry, as customers generally deem this knowledge overly costly to obtain and its loss too high a risk to their business processes. Large amounts of capability developed within research and many commercial organisations, which we term Stranded IP, have varying combinations of such translation obstacles. This Stranded IP cannot, therefore, be easily exploited for opportunities beyond the original research focus.

For Australia, which has a large gap between research inputs and economic outcomes, 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. Some may be provided in the form of open source software or curated data, but often this is not considered to be externally valuable. However, such stranded IP can often have value to third parties if these opportunities can be identified, the IP put in a usable form and then provided to them for use.

It is useful to consider research IP usage as a workflow process, composed of a series of unit operations linked in a network manner 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 allows errors in the IPs use to be introduced. Some basic integration of such software components can be achieved by using scripting languages such as Python to automate them. A more complete solution is to convert all steps into workflow unit operations with a dedicated workflow platform to facilitate the interoperability of the software components and best expose the required inputs and outputs to the less technically-minded end user.

In this paper, we will describe one specific workflow engine, Workspace, which has been developed to provide relatively low cost pathways for the translation of Stranded IP into useful applications for use by third parties. This can provide commercialisation opportunities for valuable IP whose market sizes range from large to ones that are sufficiently small so as to prevent cost-effective exploitation by more traditional software development processes. Key attributes of the workflow engine that support this include the ability to

- easily add a Graphical User Interface (GUI) layer over a workflow, including complex data visualisation options,
- create closed source compiled products with installers,
- easily convert pre-existing software into workflow compatible operations,
- arbitrarily inter-connect unit operations and to be able to adapt, customise and evolve both the workflow and GUI layers, and
- extensively re-use IP once it is within the platform framework.

These attributes provide a comparatively low cost pathway for commercialisation of both existing and new IP and software.

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

1. INTRODUCTION

Australian innovation performance was ranked 23 (out of 127 countries) by the Global Innovation Index (GII) in 2017 (Cornell University, INSEAD, and WIPO, 2017). Innovation input was ranked much higher at 12 but output was ranked 30. It is instructive to consider the "Innovation Efficiency Ratio" which is given by the ratio of the Innovation Output to Input sub-indices. Australian innovation efficiency (again from Cornell University, INSEAD, and WIPO, 2017) in 2017 was 0.6 which ranked 76 (out of 127 countries surveyed). This suggests, as reported by AFR (2016) that "Australia was considered particularly poor when it came to foreign direct investment net outflows, ICT exports and knowledge diffusion" and by Business Insider (2017) who reported that "the supporting environment for innovation is good but the results aren't matching the investments". These rankings fluctuate year by year but do so within consistent ranges (Office of Chief Economist, 2017) and so present a consistent long term picture of relatively high levels of investment in research but with relatively low levels of translation into economic outputs. It is useful to observe that Australia ranked 9th in terms of publication outputs and 10th for Citable Document H-index, but in contrast its ranking for "ICT services exports" was only 89 (Cornell University, INSEAD, and WIPO, 2017).

This top down view provides evidence of relative under-performance in the translation of research into commercial outcomes but it does not provide mechanistic understanding for the reasons that lead to this. Analysis of the overall situation tends to focus on policy structures, funding mechanisms, the business-research interface and the resultant outcomes in comparison to those of other advanced economies (ATSE, 2016). The "Boosting the Commercial Returns from Research" discussion paper (Australian Government, 2014) identifies general factors that influence translation of success of public research into commercial outcomes:

1. Research excellence (as measured predominantly by publication based metrics)
2. Targeted research effort (meaning focusing research investment 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").

The resulting "agenda for action" (DIIS, 2015) focuses on research area prioritisation, public funding mechanisms, research training, business tax incentives, research infrastructure, frameworks for reducing IP impediments to research-industry collaboration and ensuring that ARC and NHMRC "competitive grants appropriately recognise industry-relevant expertise or research".

These high level analyses do not however consider questions around the nature of the outcomes of large scale research investment, their form and what obstacles exist to the 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 frequently result in significant IP that is either not translated or is not translatable, which we term "stranded IP". This paper seeks to explore some aspects of the nature of stranded IP and obstacles to its exploitation and considers how the adoption of software workflow platforms can moderate a range of these obstacles by providing a low cost pathway from internal research outputs to exploitable software applications and products. In this context, Workspace, which is a specific workflow platform developed by CSIRO Data61 with an explicit aim of trying to reduce translation barriers, will be discussed with some examples of such IP translation.

2. BARRIERS TO COMMERCIALISATION OF RESEARCH

The accumulated difference between the input and output sub-indices of the GII (Cornell University, INSEAD, and WIPO, 2017), which is the IP output gap integrated over the life of the IP, represents an enormous pool of sunk investment cost that is not usually commercially exploitable. Some of this will be re-used in current or future projects by researchers and their teams. There is no currently available data to suggest what fraction of this stranded IP may be able to become exploitable with differing levels of barrier reduction resulting from adoption of new tool sets and models that better support commercialisation. However, the size of the accumulated under-exploited IP (which we term stranded IP) pool is extremely large, so even minor improvements in the proportion of IP exploited could lead to very substantial improvements in measured translation outcomes.

In the broadest terms, government provides policy settings and funding input (either directly through block grants or indirectly through competitive grants) to the public research enterprises (universities, government research organisations, CRCs and other specialised centres). Research management in these entities provides priorities and sets incentive and reward structures to try to influence staff with specific research projects being funded (on a large range of scales). These research projects generate IP in the course of their activities. These relationships are shown in the flow chart in Figure 1. Many research projects have publications as the sole planned IP outputs, because the existence and quality of these is used as a primary measure of research excellence. Some IP is suitable for registration (e.g. patent) but this usually occurs for only a small fraction of project IP. However there are a number of other classes of IP generated during such projects. Some can be classed as "know-how" and represents general increases in knowledge by project participants. In order to create the publication or registerable outcomes it is very common to create intermediate enabling IP (which could be numerical or experimental methods, data, computer software etc). These are sometimes regarded as Trade Secrets, but more commonly are not explicitly thought of as output IP. It would not be uncommon for a significant fraction of project costs to be expended in creating these types of enabling or internally held IP. This IP is typically not created with the intention of either future re-use or commercialisation, and almost always is in a form that is not suitable for commercialisation and so can therefore be regarded as being stranded. It could be argued that a reasonable fraction of the national research output gap is used in the creation of these types of stranded IP. Another class of output IP which is increasingly common is source software and curated data which is intended to be usable by third parties, which can be in the form of open source software. This trend is to some degree a response to policy changes that look to generate broader outcomes from projects than just publication and patents. Depending on the form of this IP and the degree to which the creators have been successful in making this usable by others, these outputs can either be useful outcomes or a different part of the accumulated stranded IP.
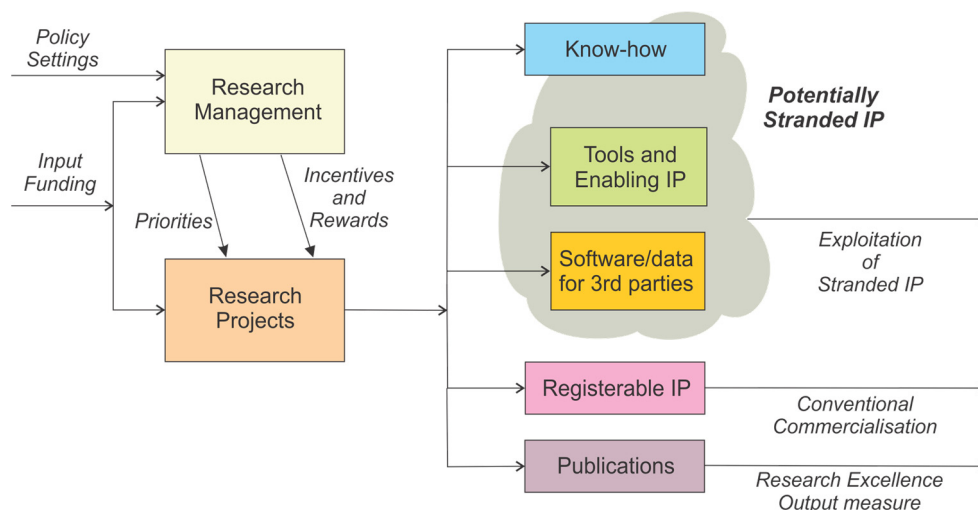

Figure 1. Flow chart representation of IP generation by research projects.

Most consideration of barriers to IP commercialisation appear to focus on issues relating to IP protection (IP registration such a patents, copyright and trademarks) and impediments to crossing the research-business interface. However, there can be significant barriers that relate to the form of the IP or to the interest of the IP generators that often prevent serious consideration of translation. So the translation process often fails well before these more broadly considered barriers are able to influence the outcome. Many tools and elements of the stranded IP pool are in the form of software or could be distributed within a software package. It is not uncommon for components of this IP to be of value to external third parties. The obstacles to providing these in some form of commercial product include:

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. Their form being unsuitable for use by third parties, where common reasons include complex and unintuitive interfaces and the existence of manual components in the supporting workflow processes.
4. Functionality is close to what is needed but not a precise match, therefore requiring some adaption.
5. Use of third party libraries and tools where licensing conditions are either unsuitable or not understood.

The first reason is a policy and management related issue in which incentives, rewards and operating structure are important. This in general is not addressable by provision of improved tools. Obstacle 3 results from the IP being 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 also often narrowly conceived in order to constrain time and cost which means that there are usually gaps when compared to third party requirements for using this IP. The consequence of obstacles 2-5 is that the cost to develop a commercial version of any software tool is often prohibitive for the commonly smaller markets that high end research based tools address and/or the timescale for creation is too long. Reliable estimation of the potential revenue from commercialisation can also be very difficult to make. Proper evaluation also often requires the software to be built so that it can be usable and testable by these stakeholders. Therefore usually only opportunities with very high perceived value are typically commercialised.

A software workflow platform is a visual programming environment that allows applications to be constructed as networks of discrete operations (Wikipedia, 2017). This structure intrinsically provides a framework for high levels of re-use of operations within workflows and for automating manual processing steps. It creates flexibility since software design and structure can be easily modified, which provides the ability to customise and adapt software operation to meet needs that are only understood after the majority of development has already been done. These are essential requirement for the later commercialisation of IP that is found to be stranded, since its form and specific capabilities are typically set by the original needs of the research project(s) that developed it and not by any understanding of third party requirements for using some variant of it later (and explicitly address obstacles 3 and 4).

3. WORKSPACE – A COMMERCIALISATION FOCUSSED SOFTWARE WORKFLOW PLATFORM

Workspace (Workspace, 2014; Cleary et al. 2014, 2015; Bolger et al. 2015) 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 (Qt, OpenSSH, Python, Boost, ZeroMQ) which all have LGPL (Lesser General Public License) or better (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. Fourth, the ability to easily attach custom UI elements to workflows provides a low cost path to tailoring user experiences to customers' domain-specific requirements.

The construction of workflows 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).

Commercial engineering applications that are examples of initially stranded IP whose cost effective translation to usable products was enabled by Workspace include (but not a complete listing):

1. ArcWeld (a welding simulation application, Murphy and Thomas (2014).
2. AlteTreat (for managing heat treatment of castings, AlteTreat (2015).
3. Spotsizer (Bischof, 2016).
4. Grainscan (Whan, 2014).
5. Amicus (bush fire prediction, see Sullivan et al. 2013).

Workspace has also been used as a platform for building new applications in a way that maximises commercialisation potential and minimises the chances of this IP becoming stranded by poor execution:

1. Flood prediction (see Hilton et al. 2015).
2. SPARK (see Miller et al. 2015).
3. Oventus (for customizing the Oventus sleep apnea device, see Oventus, 2016).
4. Dive Mechanic (Cohen et al. 2017).

CSIRO is actively, if slowly, building up a pool of higher level, re-usable Workspace based components for performing key functions (e.g. meshing, visualisation, data clean-up etc), which we term a "software product platform" (SPP) and whose availability can provide a second level of cost savings and acceleration. Any entity can in principle also use Workspace (or perhaps also other workflow platforms) to build up similarly intentioned layers of pre-built configurable components to assist with rapid translation to exploitable forms.

4.   AN EXAMPLE OF A WORKFLOW BASED SOFTWARE APPLICATION

One particular type of IP that has a strong prevalence for becoming stranded IP is physics-based simulation software (Cleary et al. 2015). Complex simulation solvers (which are able to solve valuable problems) developed in-house over long periods are often hard to use with; for example with command line execution using collections of text input files for control and problem specification. One such commercial example of a fully-fledged application based upon originally stranded solver IP exhibiting this is ArcWeld (Murphy and Thomas, 2014).

Another example that we will consider in some detail here is GF-Mill, an application used to simulate the operation of grinding mills (see Cleary, 2004 for details of the underlying Discrete Element Method (DEM) solver and its use for mills). This software can enable significant financial and time savings to be made in the design and optimisation of grinding mill liners. This DEM solver is typical of much scientific software in that it is non-trivial to use due to the inherently complex nature of the problem being tackled and that it was not developed with third party use taken into consideration. Many input parameters (mill geometry, material, physical and numerical parameters etc.) need to be specified in a range of input files. Pre- and post-processing operations and visualisation (when used in-house) were performed using separate (sometimes commercial) software applications with manual processes in between. Such semi-manual workflows are a significant obstacle for commercialisation which requires the usage to be automated, made simple to use and to not require any separate third party (particularly commercial) software tools to be required. Workspace provides significant capability to easily streamline and automate such a workflow and to make usage significantly easier by enabling attachment of intuitive GUI components. This allows non-experts easy access to the exposed capabilities of the underlying software without having to master all of its complexity.
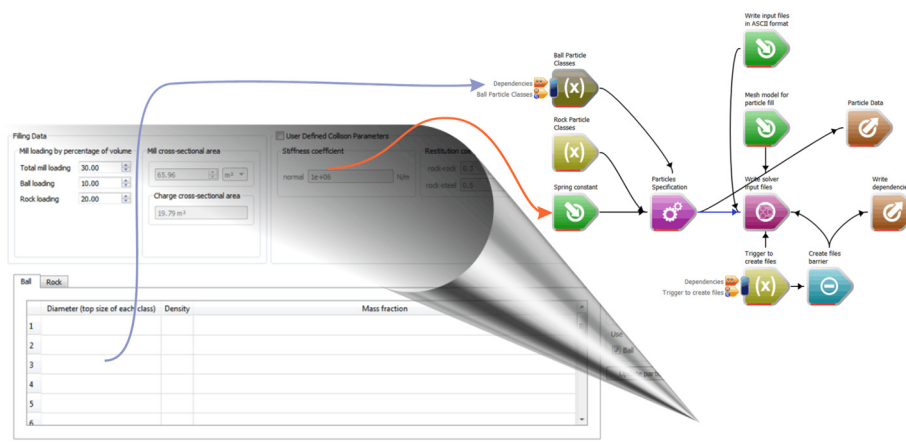

Figure 2. Connection of GUI widgets for setting particle attributes which connects to the underlying particle setup workflow and then passes data for preparation for simulation execution.

Figure 2 shows an example of the part of the application that allows a user to set collision properties and to specify the details of the rocks and balls to be used in the simulation. On the left is the section of UI where these input fields are provided in a form that is familiar to mill users and designers. The underlying workflow is shown on the right with the UI layer being "peeled back" so as to convey the manner in which the workflow underlies the GUI. This workflow sets the simulation parameters and enables the writing of the input files that the solver requires. The ball specification fields (in the GUI) are connected to the Workspace operation for setting the particle properties (in the workflow shown) with the connection between layers shown by the blue arrow. The spring stiffness, which is also set in this GUI panel, is connected (as shown by the orange arrow) to the input of an operation that controls the spring stiffness in the workflow.

A second example of how the workflow and GUI interact is given by the 3D visualisation engine used to display the flow of the particles within the mill which is represented in Figure 3. A section of the visualisation workflow is shown in the center panel. Attributes for controlling visual appearance are set in the panel on the left with form part of the visualisation GUI. These can be set or altered by the user. Such changes are passed to the workflow with the connection represented by the blue arrow between the left and center panels and lead to updates of data in the centrally shown workflow. The resulting updated visual representation of the mill is created by this workflow and is passed to the visualisation window with the 3D graphics widget from the GUI shown on the right. So the left and right panels of Figure 3 form parts of the GUI with which the user interacts while the workflow underneath drives the updating of control parameters, rendering and display of the image shown in the GUI. Workspace's high performance C++/OpenGL 3D rendering allows workflows like this to drive complex, real-time interactive 3D visualisations which can form part of the final software product provided to end users.
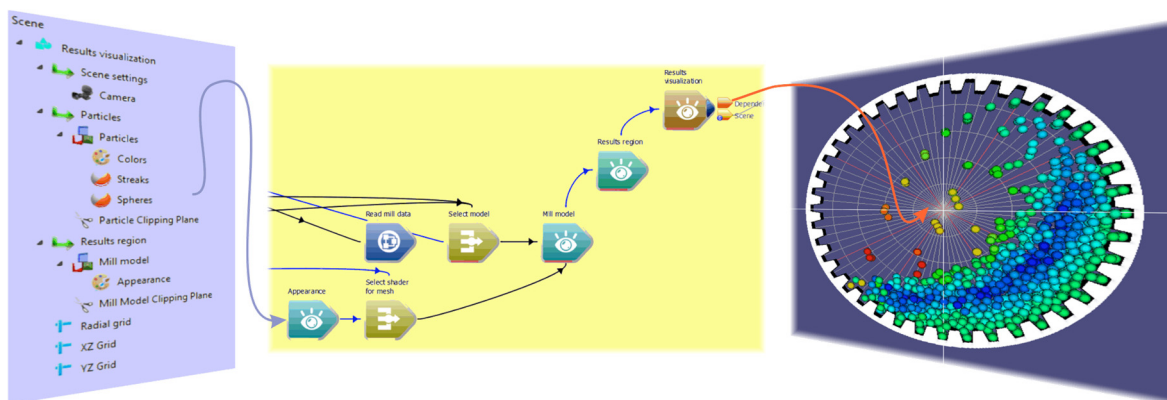

Figure 3. Connection of GUI widget for setting visualisation attributes which connects to the underlying visualisation workflow and then passes data to the render window in the application GUI.

The GF-mill application consists of several primary GUI screens connected together in the application by a simple linear workflow which makes the application simple to use by non-experts in both modelling and software engineering. The underlying workflows are nested with layers and consist of several thousand operations including ones for executing the simulation solver and for communication of both input and output data to it. Many of these operations are generic and apply to many modelling areas and so can be re-used in other applications with only minor adaption which provides significant cost reductions in the building of applications from such stranded IP.

5.   CONCLUSIONS

A relatively low level of research efficiency for Australia means that there is a large gap between the level of input funding and the commercial value of research outcomes. The nature of the IP generated by research projects was discussed and classes of IP consisting of enabling software was proposed as often having no direct path for commercialisation due to one or more common obstacles applying. Five such obstacles to the translation of Stranded IP were identified.

Use of Workspace was presented as a way to create a low cost, low skill pathway to allow research teams to translate enabling IP and software developed for internal usage to products for use by third party customers. 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.

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.pdf

AFR (2016). Australian Financial Review, http://www.afr.com/technology/australia-drops-in-global-innovation-index-20160815-gqsiou

ATSE (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.

Bolger, B., Cleary, P., Hetherton, L., Rucinski, C., Thomas, D., and Watkins, D. (2015). Workspace: Scientific Workflows and Applications for multiple Environments, *Proc. eResearch Australasia Conference*, Brisbane, Australia, 19-23th October.

Business Insider (2017). https://www.businessinsider.com.au/australia-keeps-slipping-down-the-global-innovation-ladder-2017-6

Cleary, P. W. (2014). Large scale industrial DEM modelling, *Eng. Computations*., 21, 169-204.

Cleary, P., Bolger, B., Hetherton, L., Rucinski, C., Thomas, D., Watkins, D. (2014). Workspace: A Platform for Delivering Scientific Applications", *Proc. eResearch* 2014, Melbourne, Australia, 27-31 October.

Cleary, P.W., Thomas, D., Bolger, M., Hetherton, L., Rucinski, C., and Watkins, D. (2015). Using Workspace to automate workflow processes for modelling and simulation in engineering, MODSIM 2015, Gold Coast, Australia, December 2015.

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.

DIIS (2015). https://industry.gov.au/industry/IndustryInitiatives/Pages/Boosting-the-Commercial-Returns-from-Research.aspx

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, T., Thomas, D. (2014). A user-friendly predictive model of arc welding of aluminium, Proc. 4th IIW Welding Research & Collaboration Colloquium, Wollongong, Australia, 5-6 November 2014, pp. 47.

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.

Wikipedia (2017). https://en.wikipedia.org/wiki/Scientific_workflow_system

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