Shaping Ramps for Data-Intensive Research

Malcolm Atkinson* David De Roure‡ Jano van Hemert* Danius Michaelides†

Abstract

An 'Intellectual Ramp' enables researchers to move incrementally from their current practice into the adoption of new methods. An investigation of Ramps is an important step towards "crossing the chasm" so that researchers can benefit from new tools, technologies and approaches. This paper will define and explain the concept of Ramps, discuss requirements and the anatomy of Ramps, and proposes a measurement framework illustrated by examples to improve the understanding of why ramps work.

1 Introduction

Three events triggered the production of this paper: *a*) a 2009 tour of the USA to develop our understanding of how researchers use data [1], *b*) *the RCUK review of e-Science* [2], which identified 'crossing the chasm' as an immediate and pressing challenge and *c*) *the Data-Intensive Research (DIR) Workshop in March 2010, Edinburgh* [3], which identified Ramps as the method to scaling interactions by reaching deeper into communities and reaching out to new communities

During the tour we saw many examples of effective aids that enabled researchers to progressively adopt new methods; we identified these *Intellectual Ramps* as a key concept behind many of these aids. They will prove invaluable in accelerating the passage across the 'chasm' and facilitate the broad adoption of data, software and computationally intensive methods. This paper will define and explain the concept of Ramps, describe requirements and the anatomy of Ramps, provide a series of illustrative examples, and suggest a strategy for measuring and improving the design and provision of Ramps.

2 Defining Ramps

Researchers are typically focused on challenges in their own particular domain, which may be absorbing all of their time and intellectual effort. Thus they are reluctant to allocate any time and thought to selecting and trying new methods, particularly if there is uncertainty whether that investment will pay off. Consequently, they persevere with their existing research practices possibly missing opportunities for better research results and often undertaking immense amounts of 'manual' data handling, organisation and computation.

An 'ideal' Intellectual Ramp enables researchers to move incrementally from their current practice as far as they wish into the adoption of new methods, i.e. there is no requirement to abandon current tools and practices, there is immediate advantage from new methods that grows linearly as a researcher develops knowledge of the new methods and tools, 'ascends the Ramp', and there is *no obligation* to proceed further up the Ramp than a researcher wishes. Achieving these properties perfectly may be infeasible, e.g. some steps occur or the Ramp may not yield the full potential of the method that purpose-built tools would. However, such Ramps are so effective in empowering large numbers of researchers with new capabilities, that R&D into achieving good Ramps is worthwhile.

Technological Ramps enable the research technologists who support researchers to incrementally move an existing research environment so that advances in technology may become available to their supported researchers. A particular interest is technological Ramps that empower the technologists with the ability to easily supply Intellectual Ramps to their community.

*School of Informatics, University of Edinburgh, UK
†School of Electronics and Computer Science, University of Southampton, UK
‡Oxford e-Research Centre, University of Oxford, UK

3 Characteristics of Ramps

The Ramp typically starts from a computing context already familiar to researchers: a web-browser, a spreadsheet, a file system, a tool set such as the statistical package R, etc. This indicates that its design and provision depends on a good understanding of the research community it is intended to help.

The Ramp should present the capabilities and concepts of the new methods incrementally in as small a steps as makes logical sense. In some cases, multiple capabilities may be reachable via the Ramp, and researchers may be able to choose the order of adoption. Organising capabilities and concepts so that they may be adopted incrementally is a major challenge for the creators of Intellectual Ramps. It is essential to eliminate all extraneous technical detail, which requires careful engineering and automation. It is essential to maintain the look and feel of the original environment, and to provide continuity in the semantics of the operations it provides. However, generally the new facilities will have new information of concern to the research users, e.g. on the progress of work or the status of information in a distributed system—it is both a technical design and social challenge to introduce this incrementally and non-disruptively.

Ramps grow in different parts of the ecosystem. Resource providers may provide documentation, software and training to facilitate use. But the most effective Ramps perhaps start in the researcher's work environment instead of starting at the resource. Web 2.0 services often use Ramps to facilitate users moving onto them, e.g. Dropbox www.getdropbox.com uses the file interfaces from the local file systems familiar to users to introduce version-controlled file sharing over the Internet with fine-grained access control. The user can move on to sharing files and to accessing earlier versions. Minor changes appear in the interface, e.g. to indicate that a file has yet to be synchronised. Dropbox itself is used as the starting point and mechanism for a *Drop and Compute* Ramp that facilitates biologists submitting jobs to a computational service. The service receives jobs and returns results using the Dropbox service blog.openwetware.org/eresearch. It is a good example of providing a Ramp where users can incrementally advance their use to gain more from the system.

The myExperiment www.myexperiment.org social web site is a Ramp for workflow users because it makes it easy it find scientific workflows, either through a familiar Web browsing experience or by bringing the site's functionality through into the work environment [4]. Workflows themselves are a Ramp for users to access advanced functionality, and workflow authoring systems are a Ramp for research technologists to develop workflows for the users they support.

Another example of a *Technical Ramp*, Rapid [5] provides an easy means of building and refining portals. The chemists then used Rapid to build the portal, an Intellectual Ramp, that supported their teaching and research [6]. Here again, we see the requirement to build from a starting point that is already familiar, in this case the use of XML to describe things, and the Ramp is provided by the incremental addition of features. Rapid facilitates the automatic production of all of the detailed components that are needed to deliver a web portal for scientific computing.

There is a beneficial interplay between Intellectual Ramps and education. The e-Stat project focuses on statistical software development for complex statistical models that are used in the social sciences and other disciplines www.cmm.bristol.ac.uk/research/NCESS-EStat. It allows advanced users to create domain-specific model templates to make particular instances of complex models more accessible to novice practitioners. The system is also created in such a way that it will allow mathematically orientated users who have low levels of computing expertise to reason about models and algorithms in mathematical terms, letting the system do the 'grubby' business of creating efficient computer code to execute the models.

A larger set of representative examples, including those in Table 1, will be classified to characterise the principal features of Ramps.

4 Measuring and improving Ramps

To improve our understanding of why ramps work we propose measurement of properties such as: *a*) the amount of perturbation of existing practices required for adoption, *b*) the extent to which the power of methods can be accessed incrementally, and *c*) the scope of the ramp, i.e. the amount of potential new power it offers researchers. Comparison of these measures with analyses of the working practices and expectations of the research communities will reveal principles to guide ramp design, implementation and introduction.

Ramp name	Description	References
Dropbox to manage file archiving, sharing and FTP publication	Starting from the standard file-management interface the user places files in Dropbox folders, where they can be shared, change tracked, synchronised and published	www.getdropbox.com
Drop and Compute to submit jobs	This ramp enables biologists familiar with Dropbox to submit jobs, monitor their progress and retrieve results	blog.openwetware.org/eresearch
MyExperiment a social environment for sharing scientific workflows	Using familiar web browsing and social website metaphors it enables researchers to collect, share, re-use and improve workflows	www.myexperiment.org [4]
Rapid is a technical ramp for constructing portals	Building on familiarity with XML and style specifications it facilitates portals for job submission and data management	[5]
eCAT shared on-line lab-book system	Builds on familiarity with word processors and mail attachments to enable the construction and controlled sharing of lab notebooks	www.axiope.com
eSAT enables the collaborative development of complex statistical models	It builds on statistical and mathematical reasoning skills and automates transformation into model enactment and analyses	www.cmm.bristol.ac.uk/research/NCESS-EStat
Excel data-exploration ribbon	Builds on familiarity with spreadsheets to enable researchers to explore, understand and share collections of data held in Excel	[7]

Table 1: Examples of Intellectual Ramps

References

[1] Malcolm Atkinson and David De Roure. Realising the power of data-intensive research. Technical report, National e-Science Centre, 2010.

[2] Daniel Atkins, Christine Borgman, Nathan Bindoff, Mark Ellisman, Stuart Feldman, Ian Foster, Albert Heck, Dieter Heermann, Julia Lane, Luciano Milanesi, Jayanth Paraki, Alexander Szalay, Paul Tackley, Wolfgang von Rüden, Han Wensink, and Andres Ynnerman. Building a UK Foundation for the Transformative Enhancement of Research and Innovation: Report of the International Panel for the 2009 Review of the UK Research Councils e-Science Programme. Technical report, Research Councils UK, March 2010.

[3] Malcolm Atkinson, David De Roure, Jano van Hemert, Shantenu Jha, Bob Mann, Stratis Viglas, and Chris Williams. Data-Intensive Research Workshop Report. Technical report, e-Science Institute, 2010.

[4] D. De Roure, C. Goble, and R. Stevens. The Design and Realisation of the myExperiment Virtual Research Environment for Social Sharing of Workflows. *Future Generation Computer Systems*, 25:561–567, 2009.

[5] J.I. van Hemert and J. Koetsier. Giving computational science a friendly face. *Zero-In*, 1(3):12–13, 2009.

[6] J. Koetsier, A. Turner, P. Richardson, and J.I. van Hemert. Rapid chemistry portals through engaging researchers. In A Trefethen and D De Roure, editors, *Fifth IEEE International Conference on e-Science*, pages 284–291, 2009.

[7] Roger Barga, Dennis Gannon, Nelson Araujo, Jared Jackson, Wei Lu, and Jaliya Ekanayake. Excel DataScope for Data Scientists. In *e-Science All Hands Meeting*, 2010 submitted.