

Nothing can ever be lost: Infrastructure for Archival Research and Dissemination.

Marco La Rosa, Gavan McCarthy, Ailie Smith

¹The University of Melbourne, Melbourne, Australia, m@lr.id.au

² The University of Melbourne, Melbourne, Australia, gavan.mccarthy@unimelb.edu.au

³ The University of Melbourne, Melbourne, Australia, ailie.smith@unimelb.edu.au

For close to thirty years the eScholarship Research Centre (ESRC) and its predecessors – the Australian Science Archives Project (1985-1999) and the Australian Science and Technology Heritage Centre (1999-2006) – have worked on the preservation and dissemination of knowledge and information resources for the benefit of the community. Since its establishment in 2007, the ESRC has continued and expanded on these aims by functioning as both an academic centre and a focus of infrastructure design, testing and deployment.

Central to this work has been the development of sustainable digital information resources, dating back to the launch of *Bright Sparcs* in 1994 (now incorporated into the *Encyclopedia of Australian Science* – www.eoas.info), which has been followed by the development of more than 25 online public knowledge spaces and over 50 online guides to archival collections, some of which also incorporate digitised collections. Over this time the Centre and its predecessors utilised a combination of its own technological infrastructure, infrastructure support from collaborators within the University of Melbourne, and infrastructure from external service providers.

During the first years of the ESRC's existence (2007-2008), as a consequence of ageing infrastructure, some of this significant digital legacy appeared under threat. Therefore, in late 2009 the Centre commenced significant development work on its own server infrastructure and related technology – a **data management platform** - achieved with support from the University Library and external expertise. Through deployment of this infrastructure in the University's data centre, the ESRC achieved much-needed confidence in the sustainability of its work and its ability to meet future challenges, in particular the growing number of Australian Research Council (ARC) funded projects and government funded research contracts.

The platform has been specifically geared to meet both research needs and the expectations of collaborators, in particular 'action research' methodologies combined with public domain information interventions that are the hallmark of the ESRC's research contributions. It enables the ESRC to respond quickly and has allowed the centre to apply for significant local, national international funding, including Australian National Data Service (ANDS) projects, ARC LIEF and Linkage grants, and government contracts.

To support the research needs of the ESRC, a **highly available data management platform** has been commissioned as a user file share encapsulated within a data management service which enables the research staff to work productively and efficiently. Within the platform ESRC researchers and collaborators curate and produce content which is then managed, published and monitored by **the** data management platform.

A schematic of the current platform is shown in Figure 1.

This talk will detail the highly available configuration of the platform as well as explaining the rationale behind the choice of services. Technologies that will be discussed include firewalls and loadbalancers, reverse proxies, application servers, message queues, databases, monitoring and automation.

A pervasive theme is the absolutely non-negotiable requirement for configuration automation and ubiquitous monitoring. Lessons learned during the evolution of the platform from a 'few VMs' to a vertically integrated yet horizontally scalable service will be detailed.

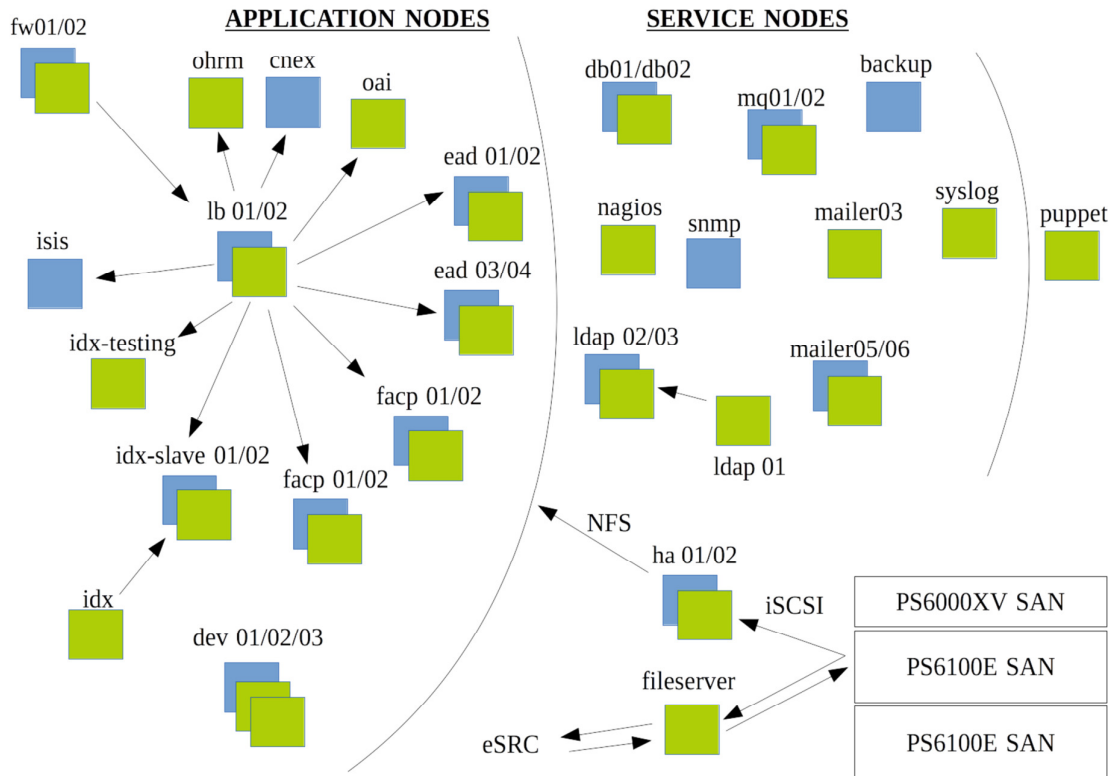


Figure 1: Schematic of the current platform

ABOUT THE AUTHORS

Dr Marco La Rosa.

After gaining a degree in Computational Chemistry Marco moved into the research support space building cluster and grid computing facilities for Chemistry and High Energy Physics; specifically, the initial Australian ATLAS Tier 2 facility for the Large Hadron Collider (LHC) Computing Grid. Taking his infrastructure expertise to VerSI Marco commissioned a Storage Resource Broker (SRB) installation federated across The University of Melbourne, Monash and La Trobe Universities; at the time the only one of its size (> 400TB across the systems of RAW capacity) in the country.

Following VerSI Marco worked as a contractor for the eScholarship Research Centre commencing what would become the basis of this presentation and the Australian Synchrotron building the initial prototype proof of concept for MASSIVE before finally taking a role with Strategic Data (a privately held software company) as Infrastructure Architect and Dev/Ops Tech Lead. During this time Marco prepared the company for an infrastructure certification by the Australian Signals Directorate gaining significant expertise in the development of the policy required to operate infrastructure for handling data at the PROTECTED LEVEL. Prior to rejoining the ESRC Marco was contracted to the Humanities Networked Infrastructure (HUNI) Project where he architected and built a cross dataset harvesting / processing solution as well as a proof of concept application and virtual lab environment. The current work Marco is tasked with at the ESRC combines his infrastructure service expertise as well as development work to produce archival tools and software for the group.

Associate Professor Gavan McCarthy

Associate Professor Gavan McCarthy is Director of the University of Melbourne eScholarship Research Centre in the University Library founded in 2007. His research is in the discipline of social and cultural informatics with expertise in archival science and a long-standing interest in the history of Australian science. He contributes to research in information infrastructure development within the University and his projects highlight strong engagement with community. His distinctive cross-disciplinary research reaches into other fields such as education, social work, linguistics, anthropology, population health and history. He re-examines theoretical foundations and tests new theories through practical interventions with a focus on public knowledge domains, contextual information frameworks and knowledge archives.

His scholarly output includes: 1 monograph; 1 edited special edition; 6 book chapters; 18 journal articles; 47 conference papers; 11 substantive and influential industry reports; over 20 large scale public knowledge web resources; and over 30 scholarly guides to archival collections of national significance. In addition he has led the development of three major software systems that express in a very practical way the results of this research. Two of these systems are widely used in both Australia and the United Kingdom and one of them has been central to recent successes with the Australian Research Council (ARC) as a Chief Investigator. In addition he devised and implemented a sustainable web publication methodology to meet scholarly and research needs. His innovations and conceptual advances have been essential to the creation, in collaboration with other researchers, of high impact public research outputs in the form of web-based public knowledge spaces. In 2011-2012 he co-edited a themed edition of Australia's leading journal in archival science, *Archives and Manuscripts*. He has been actively engaged at an international level in archival standards development and in various projects and committees focussed on the archives and history of science. Most recently he has been active in helping develop the World History of Science Online for the IUHPS Commission on Bibliography and Documentation. In 2012 he was invited to join the International Council on Archives Experts Group on Archival Description to help remodel the conceptual foundations of archival practice to enable effective intergenerational knowledge transfer in the digitally networked 21st Century.

Ms Ailie Smith

Ailie Smith is a Research Archivist at the eScholarship Research Centre at the University of Melbourne. Her work includes managing projects, the publication and management of a range of web resources, working with a range of organisations to enable them to manage their own archival collections and resources, as well as collaborating with researchers in projects with a social and cultural informatics focus. Ailie also makes significant contributions to the development of tools for capturing information for the documentation of archival collections and the documentation of contextual information, with a particular focus on the use of Extensible Markup Language (XML). She started her career in archives at the Australian Science and Technology Heritage Centre and has also worked at Bethlem Royal Hospital Archives and Museum, and the Corporate Records Unit at Imperial College London. In 2013 Ailie graduated Master of Business Information Systems from Monash University, specialising in corporate information and knowledge management, and archival and recordkeeping systems.