



FINAL REPORT AA4 WORK PACKAGE

Foster Emerging Wiki-based Collaborative Work Practices in Research Teams.

Version 1.0, May 2007

Lead Investigator: Russell Sim,

Investigators: Dianna Hardy

Prepared by:

Russell Sim russell.sim@jcu.edu.au

Dianna Hardy dianna.hardy@jcu.edu.au

Franz Eilert frank.eilert@jcu.edu.au

School of Maths, Physics and Information Technology
Faculty of Science, Engineering and Information Technology
James Cook University

Executive Summary

This report presents the findings of research undertaken as a part of the DART (Dataset Acquisition, Accessibility and Annotation e-Research Technologies) project, specifically for Work Package AA4, foster emerging wiki-based collaborative work practices in research teams. The milestones for the work package were:

- Working software with rights management flowing from chosen directory services;
- Wiki content and metadata used to generate semantic web maps of content; and
- Series of discipline specific templates and wiki structures.

These milestones were achieved by utilising an appropriate open source set of tools that would enable researchers to easily access and annotate data within a wiki-like environment. The tool that was chosen as a framework was Plone. Plone is a Content Management System based on Zope. Due to Plone's use of python as a language it is very flexible and easily extendible.

An agile software development model was followed in the implementation of this work package. As the work package team developed the skills in Plone and the requirements of the DART project and other potential users became evident, several portals were established using Plone to test the effectiveness of the tools. Despite the challenges presented during the work package, this work package is has produced a number of operational Plone portals and these are located at: <http://plone.jcu.edu.au>

Several additional milestones were achieved as part of this work package including:

- Integration with Storage Research Broker (SRB);
As datasets used by researchers can be quite large, a document storage system called SRB was integrated with Plone to provide the capability of handling large datasets.
- Improved Security;
In order to encourage collaboration between members of different institutions it was necessary to implement inter-site authentication.
- Workflow;
Development of workflow within Plone to manage documents.
- Task lists;
Creation of To Do type lists for managing activities.
- WebBlogs.
Ability of researchers to create Blogs of their activity.

This work package has demonstrated that through agile project management of the project at JCU, the work package was able to deliver several significant outcomes in addition to the milestones in the work package.

Table of Contents

1	Introduction.....	4
1.1	Organisation of this report.....	4
2	Project Milestones.....	5
2.1	Identifying the most suitable form of Wiki.....	5
2.2	Rights Management Flowing from Chosen Directory Service.....	6
2.3	Wiki Content and Metadata used to Generate Semantic Web Maps of Content.....	6
2.4	Series of Discipline Specific Templates and Wiki Structures.....	6
2.5	Additional Milestone.....	6
2.6	Sites.....	7
3	Project Outcomes.....	8
3.1	Rights Management Flowing from Chosen Directory Service.....	8
3.2	Wiki Content and Metadata used to Generate Semantic Web Maps of Content.....	9
3.3	Series of Discipline Specific Templates and Wiki Structures.....	9
3.4	Additional Milestones.....	11
3.4.1	Integration with Storage Resource Broker (SRB).....	11
3.4.2	Improved Security.....	11
3.4.3	Research Workflows.....	12
3.4.4	Plone Task List.....	12
3.4.5	MetaWeblog API for Weblogs.....	13
3.4.6	Forums.....	13
3.5	Sites.....	14
4	Archival Storage of Project Deliverables.....	16
5	Recommendations.....	17
5.1	Future development.....	17
5.2	Graduate Research School.....	17
6	Publications.....	18
7	Terms of Reference.....	18
7.1	Glossary.....	18
8	Report Signoff.....	19

1 Introduction

The Work Package AA4 project goal was to foster emerging wiki-based collaborative work practices in research teams. The milestones for the work package were:

- Working software with rights management flowing from chosen directory services.
- Wiki content and metadata used to generate semantic web maps of content.
- Series of discipline specific templates and wiki structures.

For this Work Package it was decided to extend a Content Management System (CMS) called Plone. Plone was chosen because it has a large operational base and provided a solid platform with wiki type functionality and functioning rights management. Plone also has a large development community and extensible meta data support which promotes its flexibility and potential to support external systems.

1.1 Organisation of this report

The purpose of this report is to detail the work undertaken in relation to the development of the tools for encouraging collaboration between research groups and individuals.

This report covers:

- The achievement of initial milestones;
- Achievement of additional milestones;
- Extensions developed for Plone and
- Groups who tested Plone implementations.

2 Project Milestones

The Work Package AA4 project goal was to foster emerging wiki-based collaborative work practices in research teams. The milestones for the work package were:

- Working software with rights management flowing from chosen directory services;
- Wiki content and metadata used to generate semantic web maps of content; and
- Series of discipline specific templates and wiki structures.

These milestones were achieved by utilising an appropriate open source set of tools to enable researchers easy access to data within a wiki-like environment. The tool that was chosen was the Zope based content management system, Plone.

2.1 Identifying the most suitable form of Wiki

Standard wikis were deemed unsuitable due to their lack of structure. It had been shown by research projects like pmd**CRC* (<http://www.pmdcrc.com.au/>) that after heavy use the sites Twiki (<http://twiki.org/>) became messy and difficult to navigate. Alternative wikis like Platypus's (<http://platypuswiki.sourceforge.net/>) Semantic Wiki were investigated. Platypus looked quite promising since it provided an RDF based storage system and is implemented in Java. But since the project was so new it wasn't going to ready in time for any real production implementations.

Zope had been used at JCU ITR for some time before being adopted by High Performance Computing (HPC) Unit as their de facto documentation system. Zope at ITR originally hosted a wiki implementation called Zwiki (<http://zwiki.org>) which suffered from inconsistencies after it had been in use for many years. After Zwiki became overgrown ITR switched to Plone because it provided a solid hierarchical structure. JCU's HPC unit adopted Plone after one of the staff members from ITR moved to HPC. Plone had been in use in HPC for a year before DART commenced.

Plone was chosen for use in DART because it was the most suitable web based collaboration system. Plone is built using the Zope framework and an extension to it called Content Management Framework (CMF). Plone has a large development community and has a hardened code base. Plone has excellent meta data support and focuses on users arranging documents in a hierarchical manner which supports growth.

One of the benefits of Plone is that it is open source software with a very active development community. This creates a dynamic development environment allowing for rapid and flexible deployment of additional functionality based on user requests. Several portals using Plone were established to gauge the effectiveness of the tools being produced and to gather further software requirements.

2.2 Rights Management Flowing from Chosen Directory Service

Rights management relies on delivering an integrated team management system that allows rights to be assigned to directories and files owned by the group and individuals. This needs to be integrated into a common directory service allowing the Researchers to organise themselves into groups. This is desirable because it reduces administrative overheads and allows simpler collaboration between research groups.

The emergence of the Shibboleth authentication and authorisation framework for providing authentication and authorisation services across institutions and for Virtual Organizations (VOs), provided another challenge for rights management but was outside the scope of this work package. A MAMS Grant from Maquarie University funded creation of a Shibbolised version of Plone.

2.3 Wiki Content and Metadata used to Generate Semantic Web Maps of Content

Generating semantic maps requires the development of dictionaries and thesauri of metadata that can be cross referenced for semantic searches. For example the use of the terms Bruker and Rigaku should be found by a search of X-Ray crystallography. Development of these types of semantic search engines is not currently advanced and requires strict adherence to metadata rules when creating data. Utilisation of smart search engines that can relate terms and understand semantics in a rudimentary fashion appears to provide a basic resolution to this type of issue.

2.4 Series of Discipline Specific Templates and Wiki Structures

The objective of this milestone was to develop specific templates and wiki structures for specific disciplines and this would be appropriate in a document process driven environment where document structures are predefined and utilised in workflows.

The appropriate solution for development of interactive content is to develop a set of tools that would allow researchers to develop their own portal, with appropriate interactive tools, and to create wikis appropriate to their own needs. Examples of such tools that would encourage communication include blogs and forums. Blogging is a popular and accepted method of communication where users post ideas and interesting discoveries for other users. These then can be connected to RSS feeds.

2.5 Additional Milestone

A number of additional milestones were achieved that were not initially identified. These included:

- Integration with Storage Research Broker (SRB);
- Improved Security;
- Workflow;
- Task lists; and
- WebBlogs.

2.6 Sites

A number of customer sites were implemented to trial tools and elicit feedback on the use of Plone in a live context.

3 Project Outcomes

The previous section detailed the process that was undertaken to determine the appropriate tools for collaborative work practices. This section details the components that were added and integrated with the Plone platform. It must be noted that throughout the DART project, team members used the Plone service to store information. The primary milestones were achieved as follows:

3.1 Rights Management Flowing from Chosen Directory Service

Rights management has two major requirements:

- To ensure that the correct member has been identified - authentication, and
- That the correctly identified member has the required privileges to access the resources – authorisation.

The rights management issues were address in the context of the Plone implementation, as it was chosen as the product that provided the ‘best fit’ of requirements to resolve the milestones of the AA4 work package.

Initially the TeamSpace and MXMWorkgroups products were identified as possible solutions. Both allowed site members to manage groups much like content. However the requirement of both products weren't compatible with LDAP and added more complication to the authentication system migration which occurred between Plone 2.1.3 and 2.5.

Authentication Improvement

Plone already had some LDAP support through the LDAPUserFolder plugin but there were some key issues:

- All LDAP groups are printed when mapping LDAP groups to Plone. This can take a very long time and makes choosing the correct group difficult.
- Users cannot be individually mapped to Plone only groups can.
- LDAP users don't appear in Plone when searches are performed.

The LDAP plugin is essential to ensure security as it allowed the users within James Cook University, the host institution to access Plone. Other users must be created in the LDAP directory or added to Plone own authentication mechanism.

There were changes made to the Plone source, LDAPUserFolder, and until the changes made to the LDAP plugin are accepted back into the main repository we will maintain patch sets for application to the latest versions of LDAPUserFolder and components. The changes are especially relevant in assisting in deployment of Plone at institutions which have large LDAP servers and improves integration between LDAP and Plone.

Authorisation

Plone natively allows the owner of content to set permissions to the content at a user and group level, as long as the group is defined in LDAP. Content that is created can have a predefined set of permissions and generally these permissions are set to owner access only. When the authentication processes become more complex and include users and groups outside Plone's LDAP directory then authorisation becomes more complex. However, authorisation depends on policy and the users defining the security of the content are responsible for granting access to groups and users outside the standard LDAP directory.

3.2 Wiki Content and Metadata used to Generate Semantic Web Maps of Content

Whilst the concept of generating semantic web maps appears to be meritorious, it is complex and doesn't necessarily reflect the current capability of this type of technology or the complexity required to create the metadata to generate the maps. Two plugins were researched to identify if a solution was possible.

PloneOntology

PloneOntology replaces keywords with an ontology based solution. However, the team could not get version 0.9b3 to work and, unsurprisingly, it seemed overly complicated for new users.

PortalTaxonomy

PortalTaxonomy provided a tree like implementation of keywords. In reality it just provided a better keyword selection widget and again only served to unnecessarily increase the complexity of the site. Ideas from the ontology would be used to allow clustering of search result around similar keywords.

These solutions created work for the users and added complexity to the site without providing significant benefits. It was decided to utilise a search tool.

Plone Live Search

Plone Live search can be provided with a list of key terms that can be added by the users and will also allow searches within documents. It may not provide the semantic capabilities of a true semantic search tool but it satisfies 80% of the requirements of the milestone without significant overheads. Plone Live Search was implemented as the search tool.

3.3 Series of Discipline Specific Templates and Wiki Structures

Custom templates were explored, but the overhead of upgrading new products would grow quite substantially with each new template. The overhead grows because of API changes this is particularly relevant since Plone is currently in the process of changing from Zope2 to Zope3.

A better method would be to generate content types with GenericPloneContent product using the Work Package DA3 metadata schema repository, and to use the much simpler PloneTemplates product. This would push some of the template creation to the users and the rest would be generated and easier to update.

An important aspect of achieving this milestone was to remove the barriers to collaboration that was caused by the difficulty in sharing documents with co-workers. To resolve this issue, more advanced work flows were introduced. Work flows within Plone control the default permissions on content and can trigger events. These work flows contained extra states which allowed other researchers on the site to view selected documents in a read-only fashion, as well as provide comments and other input.

3.4 Additional Milestones

These milestones were not part of the original work package but provided significant extra value and capability to the Plone System and associated tools.

3.4.1 Integration with Storage Resource Broker (SRB)

A key element to the success of Plone is the ability to access data sets and content created and stored in SRB. Whilst not one of the initial requirements, it soon became evident that the use of SRB as a core technology required Plone to have access. The first task was for users to be able to access the SRB content type. The SRB content type allows objects from SRB to be added to Plone as content. This content metadata is then indexed and the SRB object can be linked to any other content. Currently only a site wide user can be used for authentication but there is room for individual user connections.

SRB integration in Plone was perhaps the most difficult task since SRB does not include a full API for Python. Therefore, the first task was to refactor the existing API built on the incomplete SRB Python API into a more object oriented and functional model. It was then extended to include more advanced metadata queries by default.

SRB objects actually sit quite nicely within Plone since Zope follows an object based model for storing all its data. The only issue is that Plone uses a catalogue to speed up queries within Zope's object database which causes an SRB object to become out of sync with Plone. Currently the ability to view SRB metadata and edit some of it has been implemented. Dublin core metadata, which is used through out Plone, is impossible to extract from SRB due to SRB's query design only using normal JOIN operation instead of OUTER JOINS which results in ~15 query's per object.

3.4.2 Improved Security

Decentralised Authentication

In order to encourage collaboration, Plone required that an authentication scheme be implemented to allow inter-institutional authentication. The original goal was focused on the use of certificates, similar in function to that used by groups such as APAC to handle authentication. However due to Zope's deprecation of CGI support as an outdated technology, the Apache webserver was going to allow Zope complete access to the details within each user certificate, thus exposing the system to a potential security breach.

As this was clearly untenable, focus then moved to Shibboleth which provides an interface to LDAP or other authentication services via the web. Shibboleth also has support for bridging to the Grid through shibgrid (Shibbolised MyProxy) as an added benefit. Shibboleth support in Plone was possible without using CGI because the Shibboleth Service Provider can be configured to pass details through as headers to the client application. Shibboleth also provides more information about the user since any attributes from the institutions LDAP can be passed through to the Shibboleth Services.

A Shibboleth authentication extension was built for Plone/Zope. Shibboleth authentication was added as a PAS plug-in which operated in conjunction with Apache to achieve the correct authentication/authorisation. Apache actually performs the authentication and gathers the information about the users which is then passed to Zope through the HTTP request headers. Zope then grants the users the appropriate level of access depending on what details were passed through from shibboleth.

Secure Login

By default Plone provides insecure cookie based authentication and no SSL support. This product focused on replacing the default Plone cookies with improved session based cookies, it also addressed security problems by forcing the users to authenticate to the service through HTTPS which was provided via Apache.

CMF Advanced Membership Tool

This tool provides an improved method of assigning users default content on account creation (initial login). It was created to improve the flexibility of the default Plone membership tool which created the users folders on login. This tool helps enforce the authorisation level the users receives and is required when used with authentication mechanisms like LDAP/Shibboleth.

3.4.3 Research Workflows

These work flows are available to provide researchers with an easy method of sharing documents. Currently Plone provides advanced fine grained authorisation on every content item but this can be quite complicated for new users. AA4 developed work flows with extra states that allow users to share content with peers without needing an understanding of the complicated Plone sharing system.

3.4.4 Plone Task List

A task list product was developed on request from the Graduate Research School (GRS) at JCU. It provides a simple task list product for use by supervisors and research students. This 'to do' list was designed to be light weight and suit projects where a large task tracking does not suit. PloneTasks product was developed after GRS identified the basic facilities that were expected from a web collaboration system. Plone already provided some task tracking solutions but all of these were targeted at software and contained hard coded software terminology and complicated work flows. AA4 developed a solution which provided researchers with a simple but effective task tracking system which would be generic and allow crossover into other fields.

3.4.5 MetaWeblog API for Weblogs

The ability to create Weblogs where researchers could jot their daily work was desired by some of the customers and within the JCU DART team. Blogging is a popular and accepted method of communication where users post ideas and interesting discoveries for other users. Modifications were made to the MetaBlogger API to make it more compatible with tools such as ScribeFire. ScribeFire (<http://scribefire.com/>), a FireFox add on which allows posting to blogs through an XML RPC interface namely MetaWeblog API (<http://www.xmlrpc.com/metaWeblogApi>). The weblogs supported BloggerAPI by default and contained a broken implementation of MetaWeblogAPI. Some minor changes to the XML-RPC interface fixed all the problems and extensions were made to allow exporting of keywords to allow classification of weblog entries.

The Weblogs could be aggregated into an RSS feed by anyone wishing to keep track of what was occurring in the project on a daily basis. Research was conducted on a tool called Planet which aggregates RSS feeds together into a combined feed with the latest news first. This turned out to be extremely successful but access control was a big problem because it had to be done entirely by the web server, and there was not fine grained control. Administration was also an issue because it was time consuming to add/remove feeds in a dynamic environment.

3.4.6 Forums

Forums are a much older technology where users can have threaded discussions in a shared space. Forums were implemented in a basic form although connection to the mail server at the host institution is required for the forums to be truly effective. Further work investigation a solution such as Google Mil is required.

3.5 Sites

There were a number of sites that were created as part of this work package. Some sites were simple static content whilst others attempted to utilise the capabilities that were available and requested additional features. The groups School of Earth and Environmental Science, Graduate Research School and The Student Research Club all used Plone throughout its development with varying success.

[BlueNet](#) (Australian Marine Data project)

[Coastal Zone Network](#) (Coastal Zone Management)

[PRAGMA 10](#) (International Grid Initiative)

[ReefGrid](#) (Tropical Sensor Networks)

[High Performance Computing](#) (for JCU Students & Staff)

[Information Technology and Resources](#) (JCU access only)

[RV James Kirby](#) (Marine Research Vehicle)

[DART](#) (Data Acquisition ...Annotation Technologies)

[ARCHER](#) (Australian ResearCH ... enviRonment)

[VeRG lab](#)(JCU e-Research group)

[Graduate Research School](#)

[School of Earth and Environmental Sciences](#)

[School of Maths, Physics and Information Technology](#)

Graduate Research School

The GRS portal began after JCU discovered GRS was going to out source to a collaboration tool called BaseCamp. The GRS site was then created to demonstrate that Plone and DART tools could be used to manage small groups of researchers and research students. It was fortunate, as GRS became a stakeholder in AA4. GRS staff could provide feedback from a live research environment. Another benefit of being part of the GRS project was that all the stakeholders had an interest in moving resources into a virtual environment.

School of Math, Physics and Information Technology (MPIT)

The MPIT site provided AA4 with a good test of a high traffic site, since not only was the site being used for sharing research it is being used for administration purposes and communication within the school. The MPIT site has grown to include almost all the staff from the school. Very few change requests have been reported by the users, which is a testimony to Plone suitability as a CMS but unfortunately doesn't reflect all the requirements of researchers.

Dataset Acquisition Accessibility & Annotation e-Research Technologies (DART)

DART and Plone sent mixed messages, mostly due to extreme speed problems experienced at sites external to JCU. Because of these issues adoption of Plone never truly occurred, as external sites sort alternative solutions. During the project site content mostly included presentations and DART project news. Towards the end of the DART project site traffic had improved with JCU submitting all meeting notes and all Plone custom documentation.

4 Archival Storage of Project Deliverables

The software produced by this work package, the initial version of Plone will be created on a Virtual Machine (VM), zipped up and shipped to Monash University for archival storage with the other software artefacts from the DART Project.

5 Recommendations

5.1 Future development

The products developed during DART were all developed around Plone 2.1.2 which focused on Zope 2 technologies, unfortunately Plone 2.5 which was released late 2006 focused on Zope 3 technologies. It took several months before Plone could be updated on the production server. In light of this all further development should focus on Zope 3 technologies since they are more flexible and follow the MVC patterns better.

5.2 Graduate Research School

All efforts should be made to continue supporting GRS because the users are volunteering for this solution. A recent presentation to the Australian Deans and Directors of Graduate Studies (DDOGS) was very well received with significant interest from several universities in deploying this Plone based model. Until projects like JAINIS are depositing real research data that can be annotated and shared directly from Plone many researchers will view Plone (and any other web based documentation systems) as an unnecessary overhead. However, there are other groups of researcher who are fascinated by the prospects and capabilities offered by collaborative workspace, and this group needs to guide our thinking and future development.

6 Publications

Nil publications were created from this work.

7 Terms of Reference

7.1 Glossary

Acronym	Definition
APAC	Australian Partnership for Advanced Computing
API	Application Programming Interface
CGI	Common Gateway Interface
CMF	Content Management Framework
CMS	Content Management System
HPC	High Performance Computing
GRS	Graduate Research School
JAINIS	JCU and Indiana Instrument Service
LDAP	Lightweight Directory Access Protocol
MVC	Model View Controller
PAS	Pluggable Authentication Service
RDF	Resource Description Framework
RPC	Remote Procedure Call
RSS	Really Simple Syndication
SRB	Storage Resource Broker
XML	Extensible Markup Language
Zope	Z Object Publishing Environment

8 Report Signoff

It is agreed between

Franz Eilert, Russell Sim and Dianna Hardy

and

Associate Professor Ian Atkinson

and

Dr Andrew Treloar

That the **Final Report Document** for the DART AA4 – ‘Foster emerging wiki-based collaborative work practices in research teams’, gives a full account of the work undertaken for the DART Project.

Franz Eilert	0401 710 272
Russell Sim	07 4781 6729
Dianna Hardy	07 4781 6921

- has been read and reviewed by all parties,
- shows that the DART AA4 – ‘Foster emerging wiki-based collaborative work practices in research teams’ has been completed satisfactorily,
- clearly outlines the deliverables stated in the AA4 requirements documentation have been met.

Dated this 4th day of June 2007

Signed
Chief Investigator
Associate Professor
Ian Atkinson

Signed
For and on behalf of DART
Project Director
Andrew Treloar