



ADMIRE D3.2– ADMIRE Platform Release 1

| | |
|-----------------------------|---|
| Project Title | ADMIRE |
| Document Title | ADMIRE Platform Release 1 |
| Deliverable Number | D3.2 |
| Authorship | Radek Ostrowski, Rob Baxter |
| Document Filename | ADMIRE-D3.2-description.tex |
| Document Version | 1.0 |
| Distribution Classification | Project Internal |
| Distribution List | ADMIRE Project Team |
| Approval List | Malcolm Illingworth, Project Manager, Executive Board |

| <i>Personnel</i> | <i>Date</i> | <i>Comment</i> | <i>Version</i> |
|------------------|-------------|--------------------------------------|----------------|
| RHO | 04/03/2009 | Document stub | 0.1 |
| RHO | 12/03/2009 | First version | 0.2 |
| RHO | 17/03/2009 | Second version | 0.3 |
| RMB | 20/03/2009 | Added figures, enhanced descriptions | 0.4 |
| RMB | 23/03/2009 | Final draft | 0.5 |
| RMB | 27/03/2009 | Incorporated review comments | 0.6 |
| RMB | 27/03/2009 | Final edit and signoff | 1.0 |

Contents

| | | |
|----------|--|-----------|
| 1 | Overview | 3 |
| 1.1 | Current status of Workpackage 3 | 3 |
| 1.2 | Related documents | 3 |
| 2 | Definition of the ADMIRE Platform | 5 |
| 3 | Definition of the ADMIRE Testbed | 6 |
| 4 | Current State of the ADMIRE Platform: Release 1 | 7 |
| 4.1 | ADMIRE Platform Release 1 | 7 |
| 4.1.1 | OGSA-DAI (v3.1) with ADMIRE DQP extensions | 8 |
| 4.1.2 | USMT (v1.6.1) | 8 |
| 4.1.3 | ADMIRE prototype DMI-L Gateway (v0.0.1) | 8 |
| 4.1.4 | SKSA (v0.1) | 9 |
| 5 | Current State of the ADMIRE Testbed | 10 |
| 5.1 | EPCC nodes | 10 |
| 5.2 | IISAS nodes | 11 |
| 5.3 | Comarch node | 12 |
| 5.4 | Deployed Platforms | 12 |
| 5.4.1 | Production Platform | 12 |
| 5.4.2 | Development Platform | 12 |
| 5.4.3 | Integrated Test Platform | 12 |
| 5.5 | Testbed use | 13 |
| 6 | Future Plans | 15 |
| 6.1 | Project months 13-18 | 15 |

1 Overview

This report is a description of Deliverable D3.2, the ADMIRE Platform Release 1, as defined in the ADMIRE Description of Work [1].

In this report we describe the ADMIRE Platform and Testbed, and in particular the first ADMIRE Platform Release, the current state of the ADMIRE Testbed on which the Platform Releases are deployed, and summarise plans for near future.

1.1 Current status of Workpackage 3

The overall aim of Workpackage 3 is to create in the ADMIRE Platform a simple, integrated environment deployed at multiple sites to enable the testing of and experimentation with advanced DMI tools.

The concept of the ADMIRE Gateway forms the core of the ADMIRE Platform (see [2, 3]), and the foundations of the Gateway are laid by the USMT infrastructure services bus and the OGSA-DAI dataflow engine. To bring the ADMIRE Platform together, WP3 has thus worked closely with WP4, to assess the progress of the USMT infrastructure service bus, and with WP5, to determine current stable functionality within OGSA-DAI. In both cases it has been WP3's role to ensure that these key foundation components are properly tested and integrated and thus suitable for release as part of the Platform.

WP3 also works with WP5 to help identify tools and services on the “Workbench side” [2] which have reached a level of maturity sufficient to allow them to be included in the Platform Release. These components will be tested against the other components in the platform, to ensure compatibility and interoperability. The integration of these components with the Workbench or Gateway and/or other components takes place in WP4 and WP5, but with input from WP3 on testing and quality assurance.

These Releases will be made at six monthly intervals throughout the project. At each cycle, a release process is followed which assesses the status of the infrastructure and tools, assesses integration status, runs system tests, updates documentation, packages distributions and tests releases.

The rest of this report describes the first ADMIRE Platform Release.

1.2 Related documents

- [1] The ADMIRE Consortium. ADMIRE: Description of Work , Feb 2008.
- [2] Amy Krause, Carlos Buil, Rafał Gąsiorowski, Branislav Simo, Michal Laclavik, Ivan Janciak, and Rob Baxter. ADMIRE – Tools Development Report and Requirements Analysis. Deliverable report D5.2, the ADMIRE Project, Feb 2009.
- [3] A. Hume, Liangxiu Han, J.I. van Hemert, and M.P. Atkinson. ADMIRE – Architecture. Public report D2.1, the ADMIRE Project, Feb 2009.

- [4] Vivian Lee and work package partners. ADMIRE – Development and Deployment Report for USMT V2: capabilities of USMT V2. Deliverable report D4.2, the ADMIRE Project, Feb 2009.
- [5] Rob Baxter and Radek Ostrowski. ADMIRE – Platform Delivery and Support Progress Report. Deliverable report D3.1, the ADMIRE Project, Aug 2008.
- [6] Ondrej Habala, Rafał Gąsiorowski, Karol Strzelecki, Maciej Gańczyk, and Viet Tran. ADMIRE – Report on Validation of Initial ADMIRE Model and Architecture. Deliverable report D6.2, the ADMIRE Project, Feb 2009.

2 Definition of the ADMIRE Platform

The ADMIRE Platform should provide ADMIRE collaborators with tools and software components sufficient to carry out experiments and prototyping in advanced data mining and integration (DMI). The results of such work produced at one site should be easily transferrable to another. The Platform should consist of one or more nodes located at different sites in order to provide a distributed environment. There should also be an easy way of extending the Platform by third-party software or new components developed by the ADMIRE team.

The Platform should be divided into three parts: *production*, *development* and *test*.

The *Production Platform* is defined as a stack of software developed by the ADMIRE team and named as an *ADMIRE Release*. The latest stable version of an ADMIRE Release should be installed on a set of machines at distributed locations.

The *Development Platform* should be a newer, ‘living’ version of the above. It should be based on incremental updates to the previous Platform plus experimental services deployed as candidates for the next ADMIRE Release.

The integrated *Test Platform* should be a test framework which verifies that all components of an ADMIRE Release work and provide the expected functionality. The framework should be easily extendible and the test coverage should be wide and include ADMIRE specific tests. The most important system tests should be based on use cases for applications presented in WP6.

All three parts of the ADMIRE Platform should be installed on the ADMIRE Testbed.

3 Definition of the ADMIRE Testbed

The Testbed comprises a number of dedicated servers drawn from across the project consortium across which it is intended to deploy Production, Development and integrated Test Platforms. Selected sites host one or more servers. Our aim with the Testbed is to provide a mix of capabilities and operating systems to exercise the ADMIRE Platform tools and services across a range of hosting environments. The Testbed has not only ADMIRE Releases installed but is also augmented by third-party software necessary to conduct experiments and prototyping. These include file, RDBMS and XML storage systems populated with test data as well as other pieces of software found to be required at the later stage of the project.

The current state of the Testbed is described in Section 5.

4 Current State of the ADMIRE Platform: Release 1

4.1 ADMIRE Platform Release 1

ADMIRE Platform Release 1 consists of the following components:

- OGSA-DAI (v3.1) with ADMIRE DQP extensions;
- USMT (v1.6.1);
- ADMIRE prototype DMI-L Gateway (v0.0.1);
- Semantic Knowledge Sharing Assistant (SKSA) (v0.1).

Together, the first three form the core of the services expected of the full ADMIRE Gateway — conceptually, Figure 1 sketches the logical structure of the Release 1 Gateway services.

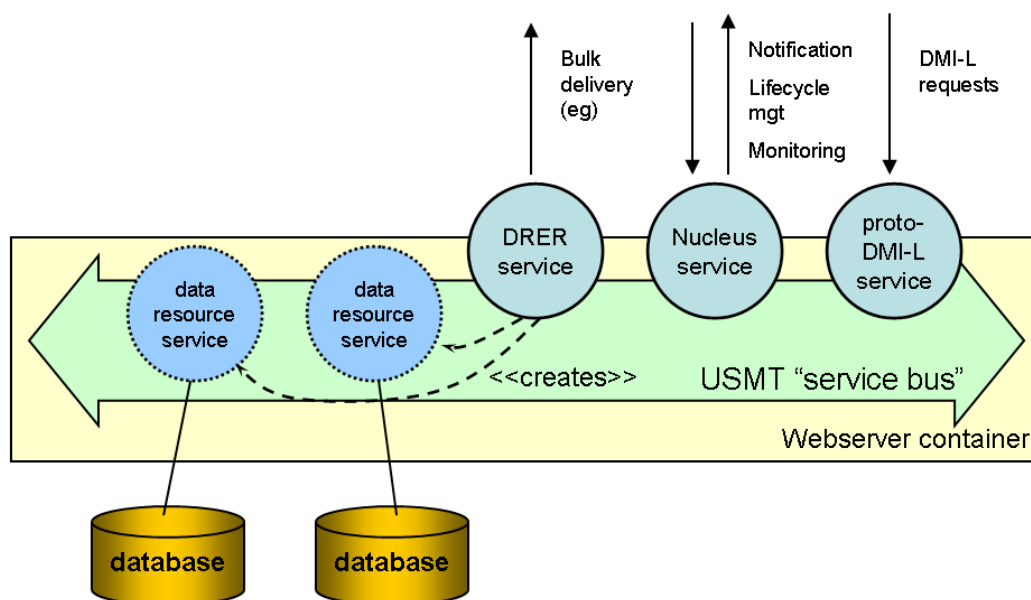


Figure 1: Conceptual sketch of the ADMIRE Gateway services in Platform Release 1. The merger of USMT and OGSA-DAI has created a common “service bus” which allows the Gateway services to use standard USMT mechanisms to orchestrate OGSA-DAI data resources through the agency of the Data Resource Execution service (DRER). This provides the ADMIRE Gateway with its core foundation. The prototype DMI-L Gateway service provides an example of the main interface to be presented to the ADMIRE Workbench tools. The Nucleus service provides the Workbench tools with the interfaces needed to benefit from the notification, monitoring and lifecycle management functions of USMT. The DRER, while primarily an internal service, can be exposed to the Workbench tools as a mechanism for bulk data delivery or other “sideband” communication mechanisms of this nature.

4.1.1 OGSA-DAI (v3.1) with ADMIRE DQP extensions

Full OGSA-DAI user documentation can be found at

<http://www.ogsadai.org.uk/documentation/ogsadai3.1/>

The DQP extensions for ADMIRE are described in §4 of Deliverable D5.2 [2].

4.1.2 USMT (v1.6.1)

USMT is documented online at

<http://www.admire-project.eu/trac/wiki/USMT>

and its current status is covered in Deliverable D4.2 [4].

4.1.3 ADMIRE prototype DMI-L Gateway (v0.0.1)

The ADMIRE prototype DMI-L Gateway is currently running on USMT v1.6 at

<https://admire4.epcc.ed.ac.uk:8443/>.

Code for the Gateway is located in the ADMIRE SVN repository at

`trunk/testbed/DummyGateway_USMT`.

The idea behind this first prototype release (v0.0.1) of an ADMIRE DMI-L Gateway is to build something that begins to look like the planned main ADMIRE Gateway interface and can be used by tool developers when developing their tools.

The first aim is to support an example where a client submits a request to the Gateway and is returned the end-point reference (EPR) of a newly created request resource. This request resource allows the client to observe (and, in later phases, manage) the request.

So we have two services:

- Gateway
 - Operations
 - * EPR execute(String requestInDML)
 - Resource Properties:
 - * None beyond the standard lifetime ones you get for free
- Request
 - Operations

- * None — all interaction will be through resource properties or lifetime management operations.
- Resource properties
 - * Status: RUNNING | COMPLETED
 - * Progress: 0..100

Client calls `execute` on the Gateway. The Gateway creates a new Request service instance and returns the EPR of this request to the client. The request service instance starts a thread that does the work (in this trivial demo case, counts from 0 to 100) updating the Progress and Status resource properties as it goes. The Client registers an interest in these resource properties and receives notification events when they change.

4.1.4 SKSA (v0.1)

The SKSA is described in §3.5 of Deliverable D5.2 [2].

5 Current State of the ADMIRE Testbed

Figure 2 offers a rough sketch of the current state of the ADMIRE Testbed.

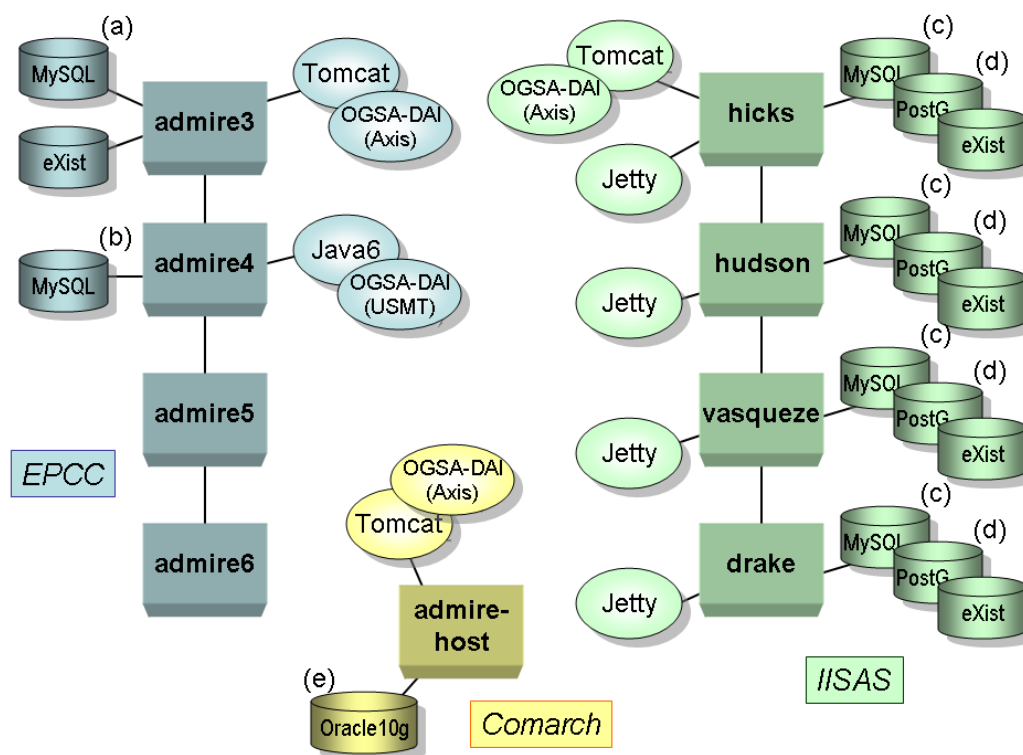


Figure 2: Current ADMIRE Testbed (at project month 13), spanning three sites: EPCC, IISAS and Comarch. The data deployed in each of the lettered databases is as follows:

- a example synthetic CRM data (cf. [5], §3.3.2);
- b UK census and postcode data (cf. [5], §3.3.2);
- c ORAVA river scenario data (cf. [6], §4.1);
- d Slovak meteorological data (cf. [6], §4.1);
- e Comarch ACRM test data (cf. [6], §5.1).

5.1 EPCC nodes

1. Machines

- admire3.epcc.ed.ac.uk, 129.215.63.145 – x86; Scientific Linux 5
- admire4.epcc.ed.ac.uk, 129.215.63.237 – x86; Scientific Linux5
- admire5.epcc.ed.ac.uk, 129.215.0.197 – x86; CentOS 5
- admire6.epcc.ed.ac.uk, 129.215.0.198 - x86; Scientific Linux 5

2. Software

- Tomcat/Axis/OGSA-DAI: <http://admire3.epcc.ed.ac.uk:7070/dai>
- MySQL (5.0):
 - admire3 – access allowed from machines admire3-5.
 - admire4 – access allowed from machines admire3-4.
- eXist: <http://admire3.epcc.ed.ac.uk:9999/exist>
- OGSA-DAI/Prototype Gateway/USMT v1.6: <https://admire4.epcc.ed.ac.uk:8443>

3. Deployed Data

- Artificially generated commercial data for datamining (‘clean’ and ‘polluted’ customer and contract data) – MySQL on admire3
- Census information and postcode mappings data – MySQL on admire4

5.2 IISAS nodes

1. Machines – virtual machines all running Ubuntu 8.04 (512Mb RAM; 300Gb space)

- hicks.ui.sav.sk, 147.213.65.94
- hudson.ui.sav.sk, 147.213.65.95
- vasqueze.ui.sav.sk, 147.213.65.96
- drake.ui.sav.sk, 147.213.65.97

2. Software

- Tomcat(5.0.28)/Axis/OGSA-DAI(3.1): <http://hicks.ui.sav.sk:8080/dai/>
- MySQL(5.0) on all testbed machines (port 3306)
- PostgreSQL(8.3) on all testbed machines (port 5432)
- eXist(1.2.1) on all testbed machines (7070/exist)
- Jetty(6.1.9) on all machines (port: 7070)
Access currently is allowed from IISAS IP range (147.213.65.0/24) and UEDIN IP range (129.215.0.0/16).

3. Deployed Data – GRIB files: binary files containing meteorological data from Slovak Hydrometeorological Institute:

- MySQL:
 - OravaWaterStations – data from water station on Orava river
 - SK_CGMS_SAV1 – biomass production of selected crops (for are of Slovakia) (simulation data)
 - SK_CGMS_SAV2 – biomass production of selected crops (for are of Slovakia) (simulation with influence of underground water)
 - grib_meta – metadata on GRIB files content
 - svp – data on waterworks in Slovakia
- PostgreSQL:
 - MarsStat – interpolated meteorological data from year 1975

5.3 Comarch node

1. Machines

- admire-host, 192.168.117.1 – x86; Windows Server 2003 R2 EE

2. Software

- Tomcat(5.0.32)/Axis/OGSA-DAI(3.1)
- Oracle 10g EE

3. Deployed Data

- Example data in the analytical CRM (ACRM) system schema.

5.4 Deployed Platforms

5.4.1 Production Platform

At this stage in the project only the Development Platform is up and running.

5.4.2 Development Platform

The Platform Release 1 described in Section 4 is currently deployed as the development platform.

5.4.3 Integrated Test Platform

The strategy of testing in the ADMIRE project aims to minimise the number of defects in the delivered Platform by applying several different levels of testing to each Release:

- *unit testing*. This targets single units of the platform code, and concentrate on assessing whether each unit performs within its specification. WP4 and WP5 are responsible for implementing the tests which WP3 will review for coverage;
- *integration testing*. This tests integration of ADMIRE modules in the Platform, and will be performed by WP3, with assistance from WP4 and WP5;
- *system testing*. This ultimately will test system performance, stability and functionality according to the application requirements, and will be performed by WP3 and supplemented by WP6 for application driven testing.

Currently, the ADMIRE Test Platform has integration and system tests exercising OGSA-DAI and the prototype Gateway. The installed tests for OGSA-DAI include standard system tests and additional tests developed for the extended distributed query processing (DQP) functionality [2]. The tests for the Gateway are currently quite simple; these will be extended when the functionality of the Gateway increases.

Running the system tests. In order to run the system tests one has to log in to `admire3.epcc.ed.ac.uk` as username ‘test’, go to `/home/test/new-tests/` and execute ant target ‘ant install-all; ant run-all’:

```
$ ssh admire3.epcc.ed.ac.uk
$ cd /home/test/new-tests/
$ ant install-all; ant run-all
```

This will install the software and run all the tests and generate a report available at

<http://admire3.epcc.ed.ac.uk/ci2>

(see Figure 3). The tests are configured to be run nightly. The build script provides more functionality which can be viewed by running ‘ant -p’ command.

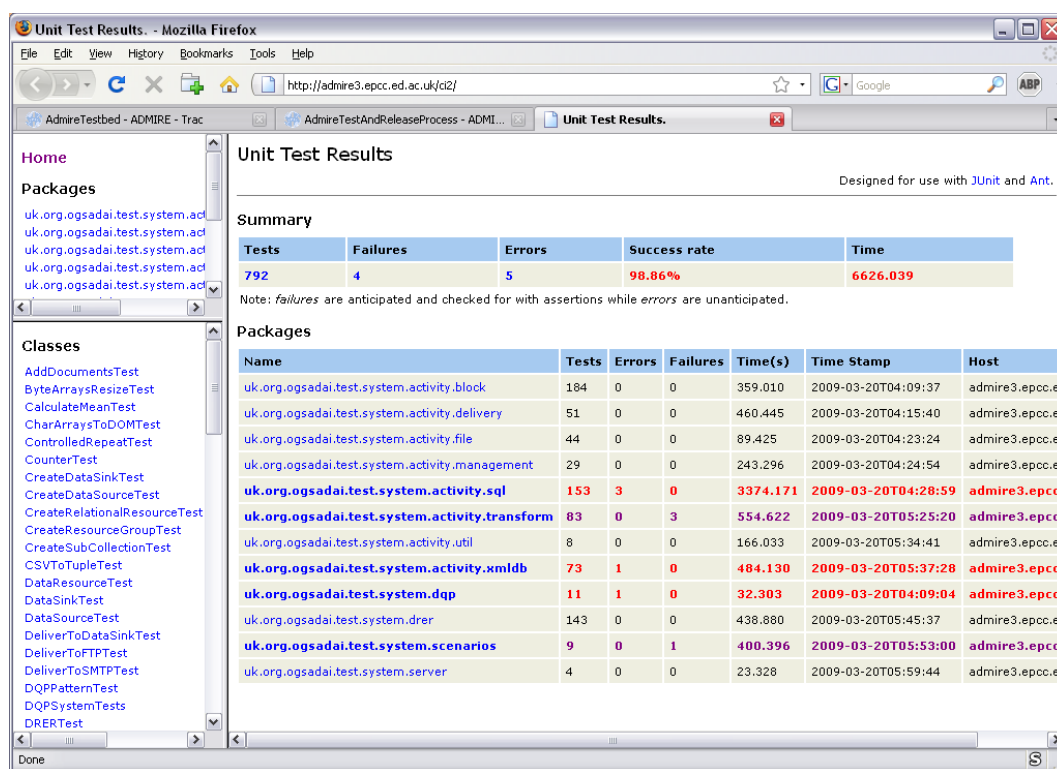


Figure 3: An example snapshot of the ADMIRE Integrated Test Platform dashboard at <http://admire3.epcc.ed.ac.uk/ci2>.

5.5 Testbed use

In its current basic form the Development and Test Platforms on the Testbed can be used within the ADMIRE consortium. Effective use is currently only supported from within tools under active development — connections to Testbed services from developers’ desktops, for

instance. As the Production platform is deployed in the next phase of the project the Testbed will become more user friendly.

ADMIRE developers can currently:

1. use installed ADMIRE Platform Release 1 components to:
 - see them at work and run demos;
 - work on interoperability and interface definitinos, particularly with the prototype Gateway;
 - develop new components;
 - extend existing components;
2. utilise third-party software installed on the Testbed (for instance the WEKA data mining component library);
3. employ data stored on the testbed for data mining through OGSA-DAI or JDBC (MySQL, PostgreSQL, Oracle 10g EE):
 - ORAVA scenario flood data at IIAS;
 - commercial ACRM data at Comarch;
 - synthetic commercial data at EPCC;
4. observe automated test results to verify if particular components behave as expected.

6 Future Plans

ADMIRE Platform Release 1 is the first of a series of six-monthly Platform releases. As ideas are developed within the research workpackages, and software components matured in the development workpackages, WP3 will expand the supported Platform and Testbed infrastructure accordingly.

6.1 Project months 13-18

Plans for the next release of ADMIRE software are likely to encompass the following:

1. ADMIRE Platform Release 2
 - First integrated ADMIRE Workbench, including Process Designer, SKSA, Data Preparation Tool and Model Visualiser.
 - First functioning ADMIRE Gateway, including basic DMI-L interpreter and enhanced USMT and OGSA-DAI.
2. ADMIRE Testbed hardware
 - Roll out ADMIRE Platform Release 1 onto EPCC and IAS nodes.
 - Extend Testbed with additional nodes at Comarch and FLE.
3. ADMIRE Testbed infrastructure
 - Create system tests based on use cases from WP6 which will exercise the whole platform.
 - Improve mechanisms for collecting and presenting software metrics and test results.