



ADMIRE D3.4 — ADMIRE Platform Release 3

Project Title	ADMIRE
Document Title	ADMIRE Platform Release 3
Deliverable Number	D3.4
Authorship	Radek Ostrowski, Rob Baxter
Document Filename	ADMIRE-D3.4-description.tex
Document Version	1.0
Distribution Classification	Project Internal
Distribution List	ADMIRE Project Team
Approval List	Maciek Jarka, Malcolm Illingworth, Project Manager Executive Board

<i>Personnel</i>	<i>Date</i>	<i>Comment</i>	<i>Version</i>
RHO	18/02/2010	Document stub	0.1
RHO	22/02/2010	First draft	0.2
RHO	23/02/2010	Incorporated reviewer's comments	0.3
RMB	24/02/2010	Minor edit and signoff	1.0

Contents

Executive Summary	2
1 Overview	4
2 ADMIRE Platform Release 3	5
2.1 ADMIRE Platform R3: Gateway Services	5
2.1.1 ADMIRE Gateway Service	5
2.1.2 OGSA-DAI Services	5
2.1.3 ADMIRE Registry Service	6
2.1.4 Database Monitoring Service – DSMON	6
2.2 ADMIRE Platform R3: Workbench Components	6
2.2.1 ADMIRE Update Site	7
2.2.2 Sharing Assistant (aka SKSA) Plugin	7
2.2.3 Process Designer Plugin	7
2.2.4 DMI Model Visualiser Plugin	7
2.2.5 Chart Visualiser (aka Data Preparation Tool) Plugin	7
2.2.6 Gateway Client Plugin	7
3 ADMIRE Testbed	8
3.1 EPCC	8
3.2 IISAS	9
3.3 Comarch	9
3.4 NeSC	10
3.5 U. Vienna	10
3.6 UPM	10
3.7 ADMIRE Libraries Dependency Management	10
3.8 Integrated Test and Monitoring Platform	11
3.9 Testbed Usage	11
4 Future Plans	13
4.1 Project months 19-24	13

Executive Summary

ADMIRE Platform Release 3 provides a stable, end-to-end DMI software platform for the ADMIRE project. In this release, the emphasis has been placed on upgrades to the Workbench tools. The release also includes updates to the Gateway and the Registry.

The ADMIRE Testbed has been augmented with new versions of the ADMIRE Platform software and now boasts five Gateway services across Europe.

The ADMIRE Platform and Testbed progress up to PM18 is described in D3.3 [1]. The key goals for PM24 were the following:

Deploy Platform Release 2 across Testbed: *achieved* — Platform running at EPCC, IISAS, Comarch, NeSC and Vienna.

First integrated ADMIRE Workbench: *achieved* — all plugins are scheduled to be published into the ADMIRE Update Site from where can be easily installed into Eclipse.

Improvements to the ADMIRE Gateway: *achieved* — Gateways are now able to process DISPEL requests and redistribute them across remote gateways.

Improve collecting and presenting software metrics and test results: *partially achieved* — the focus of this goal changed slightly to create a test system which presents the results of different tests run on a regular basis against all the Gateways installed on the Testbed. The result page is called the Dashboard and is further described in Section 3.8.

Create system tests which will exercise the whole platform: *not achieved* — ongoing, those tests will get incorporated into the Dashboard.

1 Overview

This report is a description of Deliverable D3.4, the ADMIRE Platform Release 3, as defined in the ADMIRE Description of Work [2].

In this report we describe the third release of the ADMIRE Platform and the current state of the ADMIRE Testbed.

The ADMIRE Platform is a collection of tools, services and software engines designed to demonstrate the current state of DMI research and development within ADMIRE. Every release of the Platform gathers the most recent stable components into a logical package, each time building on the previous release;

The ADMIRE Testbed is an instantiation of the ADMIRE Platform across a number of compute and data resources shared across the ADMIRE consortium.

For more detailed conceptual definitions of the ADMIRE Platform and Testbed, and the full description of activities within Workpackage 3, please see ADMIRE D3.2 [3].

We conclude this report by drawing plans for the near future.

2 ADMIRE Platform Release 3

The third release of the platform, in comparison with the second release, hands to the user a more complete set of client-side tools. These facilitate easier creation of DISPEL documents as well as allow deeper analysis of the data before any execution on the Gateway side is performed. Also featuring in this release is an upgrade to a new DISPEL parser in the Gateway which is able to delegate parts of DISPEL documents for processing to other distributed Gateways. The Gateway also integrates better with Registry which has also matured since the previous release.

A one page overview of platform, tools and testbed is now available on the ADMIRE website:

<http://www.admire-project.eu/platform.html>.

2.1 ADMIRE Platform R3: Gateway Services

On the Gateway side, Release 3 includes the following services:

- USMT WSRF Container Services v1.7.0 (described more fully in Deliverable D4.4 [4]);
- ADMIRE Gateway Service v1.0.2;
- OGSA-DAI Services v1.0.0;
- Registry Service v0.0.2;
- Database Monitoring Service (DSMON) v0.7.0.

2.1.1 ADMIRE Gateway Service

All services running within the ADMIRE Gateway are hosted within the USMT web services framework.

The ADMIRE *Gateway Service* is now equipped with a new DISPEL parser which is able to process DISPEL requests that can be distributed across remote Gateways. It also manages all required data transfers between the processing instances. It has also been augmented with a number of new extension points to allow the modification and optimisation of DISPEL graphs in the future. The ADMIRE Gateway Service has been extended to support Registry updates. This change will allow new Processing Elements to be registered in the ADMIRE Registry. Both these components are described in ADMIRE D4.4 [4].

2.1.2 OGSA-DAI Services

The ADMIRE Platform currently uses the OGSA-DAI data workflow engine as an enactment platform to provide all the functionality of ADMIRE Processing Elements (PEs). The

core OGSA-DAI services deployed within USMT in the ADMIRE Platform are discussed in ADMIRE D4.2, D4.3 and D4.4 [5, 6, 4]. Full OGSA-DAI user documentation can be found at:

<http://sourceforge.net/apps/trac/ogsa-dai/wiki/UserDocumentation>

2.1.3 ADMIRE Registry Service

The ADMIRE Registry provides access to the canonical descriptions of the Processing Elements designed by the ADMIRE end users. It is a semantic registry based on the WS-DAI-RDF specification which is queried using SPARQL. ADMIRE Platform R3 includes an updated version of the Registry described in D4.4 [4].

2.1.4 Database Monitoring Service – DSMON

DSMON is a service-oriented monitoring tool providing rich metadata about relational data sources via publish/subscribe mechanisms in a uniform manner. For the current release of the platform DSMON has been upgraded to the latest version of USMT and is described in ADMIRE D4.4 [4].

2.2 ADMIRE Platform R3: Workbench Components

This release of the ADMIRE Platform includes new Eclipse workbench plugins of which all are described in detail in ADMIRE D5.4 [7]:

- Chart Visualiser (formerly Data Preparation Tool) (v1.0.0);
- Sharing Assistant (aka SKSA) (v1.1.0);
- Process Designer (v1.0.0);
- DMI Model Visualiser (v1.0.0);
- Gateway Client (v1.0.2).

As in the previous release, the functionality of the Eclipse Update Site to install and update its plugins across the web will be employed.

Note that, even though the DISPEL Editor Plugin is fully functioning, it has been deprecated and will not be maintained. Its functionality is now provided by the Process Designer, which allows the creation of DISPEL automatically in a drag and drop fashion as well as through a text-editing mode.

2.2.1 ADMIRE Update Site

Eclipse ADMIRE Update Site (described in ADMIRE D3.3 [1]) has been updated and populated with latest Workbench plugins (described in the following sections). It lets a user upgrade the plugins from the previous versions or install them for the first time in a new Eclipse in an easy and automated way.

The ADMIRE Update Site is hosted on the Testbed at:

`http://www.admire-project.eu/ADMIREUpdateSite/`.

2.2.2 Sharing Assistant (aka SKSA) Plugin

This tool allows users to add and share human-understandable knowledge in the context of the ADMIRE ontology developed in WP1. It is able to detect the context in which the user is working, based on the ontological description of the active platform and application components, and attach new knowledge in human-readable form to that context. Also, any previously recorded knowledge pertinent to the user's current context will be available to the user, thus making knowledge navigation and experience management easy for non-IT experts.

2.2.3 Process Designer Plugin

The Process Designer is a graphical environment tool that allows DMI experts to compose Processing Elements and create complex DMI processes. The PD creates DISPEL documents that can be submitted for processing to a Gateway or stored in the Repository.

2.2.4 DMI Model Visualiser Plugin

The DMI Model Visualiser is a tool integrated to the ADMIRE workbench for visualisation of the data mining models produced by the processing elements available on the ADMIRE Gateway. The visualiser does not strictly depend on any other workbench plug-ins as it can independently consume any PMML file. This file can be stored locally or available on an accessible URL.

2.2.5 Chart Visualiser (aka Data Preparation Tool) Plugin

The Chart Visualiser is a tool used to visualise row sets as charts. It can produce various kinds of charts, including pie charts, scatter plots and histograms to visualise the data.

2.2.6 Gateway Client Plugin

This plugin facilitates submitting DISPEL documents to the Gateway for processing.

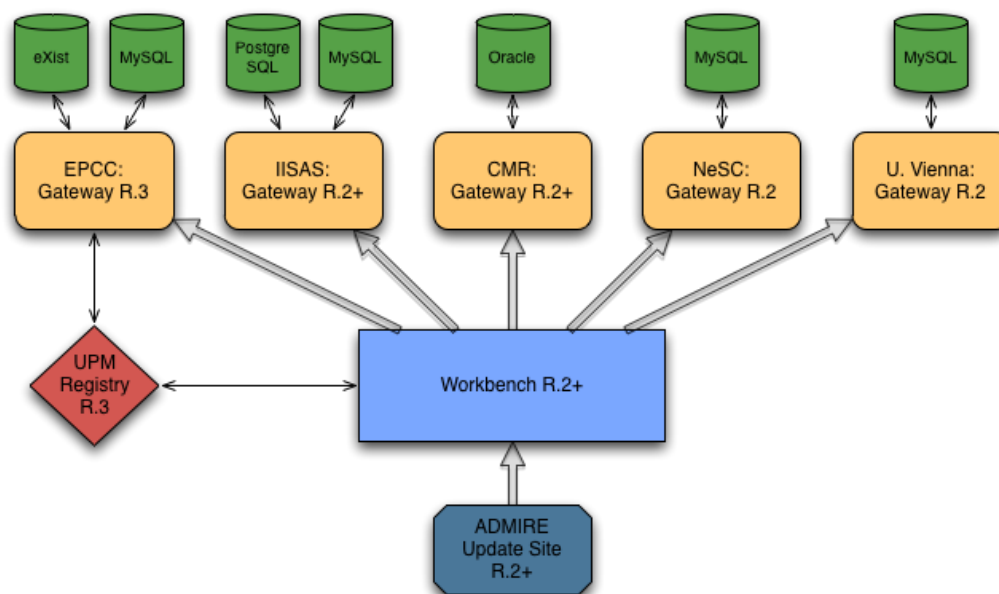


Figure 1: Current ADMIRE Testbed (at project month 24), spanning five sites: EPCC, IISAS, Comarch, NeSC, U.Vienna and UPM.

3 ADMIRE Testbed

At the time of writing, there are five ADMIRE Gateways and a main instance of ADMIRE Registry deployed on the ADMIRE Testbed. They form the Development Platform which can be used by all consortium members. When the platform reaches a maturity level suitable for providing it to the external world a Production Platform will be created. Figure 1 offers a sketch of its current state.

3.1 EPCC

- Gateway at version 1.0.2 running at `admire4.epcc.ed.ac.uk`. Available services include:
 - OGSA-DAI
 - DSMON
- Data Resources available:
 - DbAdmire3Resource – MySQL:
 - * Standard OD tables: ‘littleblackbook’ and ‘extblackbook’
 - * Generated data for datamining:
 - clean data – ‘customer_clean’, ‘contract_clean’
 - ‘polluted’ – ‘customer’, ‘contract’
 - * Some other tables like: ‘weather’

- DbAdmire4Resource – MySQL:
 - * Census information: ‘stat’
 - * UK Postcode mappings: ‘postcode’
 - * Other used for Decision Trees: ‘measurements’, ‘flights’ etc.
- XMLDBResource – eXist XMLDB:
 - EPRs of ADMIRE Gateways:
 - * gateway-epr-epcc.xml
 - * gateway-epr-iisas.xml
 - * gateway-epr-comarch.xml
 - * ...

3.2 IISAS

- Gateway at version 1.0.1 running at hudson.ui.sav.sk. Available services include:
 - OGSA-DAI
- Data Resources available (in one MySQL database and mirrored in other PostgreSQL):
 - DbSvpResource:
 - * SVP data (Vah reservoirs), tables ‘vodne_diela’ and ‘backup_vodne_diela’
 - DbGribMetaResource:
 - * Metadata of GRIB files, tables ‘grib_meta’, ‘grib_meta_c’, ‘grid_coords’, ‘grid_coords2’
 - DbOravaWaterStationsResource:
 - * Hydrological data (water height, water discharge) from several hydrological stations in the Orava area. Tables ‘Hh’, ‘Qh’, ‘Stations’, ‘Thod’
 - DbSKCGMSSAV1Resource:
 - * Various pedological and crop-related statistics for Slovakia. Too many tables to list, examples: ‘CROP’, ‘CROP_YIELD’, ‘INITIAL_SOIL_WATER’.
 - DbSKCGMSSAV2Resource:
 - * A backup copy of DbSKCGMSSAV1Resource.

3.3 Comarch

- Gateway at version 1.0.1 running at rd.comarch.pl. Available services include:
 - OGSA-DAI
- Data Resources available (in one Oracle database):
 - crm:
 - * CRM resource. Includes around 200 different tables, for example ‘CDM_CUSTOMERS_T’

3.4 NeSC

- Gateway at version 1.0.0 running at nesc-red-009.nesc.ed.ac.uk. Available services include:
 - OGSA-DAI
- Data Resources available (in one MySQL database):
 - DbTestNeSCResource:
 - * Holds EURExpressII data

3.5 U. Vienna

- Gateway at version 1.0.0 running at lela.gridlab.univie.ac.at. Available services include:
 - OGSA-DAI
- Data Resources available (in one MySQL database):
 - DbViennaTestResource:
 - * Testing data for preprocessing algorithms.

3.6 UPM

- Registry at version 0.0.2 running at regulo.dia.fi.upm.es. Available services include:
 - The registry file is at <http://delicias.dia.fi.upm.es/cbuil/rdf/admire/PlatformOntology.rdf>

3.7 ADMIRE Libraries Dependency Management

In every distributed project there is a risk of running into problems due to the use of different versions of libraries. To mitigate against this, we have introduced a dependency manager for ADMIRE projects. We chose Ivy (<http://ant.apache.org/ivy>) as it is light-weight, integrates well with Ant build scripts (widely used in ADMIRE) and, simply, does the job.

We also created a consortium-wide ADMIRE “jars repository” to use with the dependency manager. It is located here: <http://admire3.epcc.ed.ac.uk/ADMIRE-jars-repo/>.

This introduced solution makes updates to libraries/projects very easy, as well as reducing the size of code projects and helping to keep the projects clean.

3.8 Integrated Test and Monitoring Platform

As the ADMIRE Testbed has been growing, a need was identified to be able to monitor the availability and proper functioning of the Gateways and the Registry. This resulted in the creation of a set of tests which “ping” the Gateways and the Registry to check if they are up and responding as expected. Additionally, there are tests which verify the processing of data and availability of the databases on the Gateway side and the integrity of the Processing Elements’ descriptions stored in the Registry. Tests are scheduled to run at a regular 30 minutes intervals and publish results to the *ADMIRE Dashboard* website:

`http://admire3.epcc.ed.ac.uk/dashboard/`

So far, the Dashboard has proved to be very useful. It has already helped to discover problems caused by frequent access. Additionally, at a glance, it allows users to identify Gateways or Registries which are down or not passing tests and speeds up the process of maintenance and support.

3.9 Testbed Usage

It is expected during PM24-30 to utilise the Testbed to perform a number of scale-up experiments. These will be run using distributed and large (terabytes, hundreds of terabytes) amounts of data. To achieve this, the ACRM churn prediction scenario data set will be divided into four databases, each hosted by different ADMIRE collaborators. The main entities of the ACRM database are customers, contracts, contacts and calls. Groups of tables for these entities and supporting tables, i.e. dictionaries, will be deployed on different servers running different DBMS. Large data volumes for each part will be generated using the Data Generation Tool, a utility able to generate arbitrarily large synthetic CRM data based on the Comarch ACRM database schema and known patterns of records.

Figure 2 depicts the planned usage of the Testbed servers. More details about the “ACRM scale-up experiment” can be found in ADMIRE D6.4 Pilot Applications Development and Evaluation Suite Report [8].

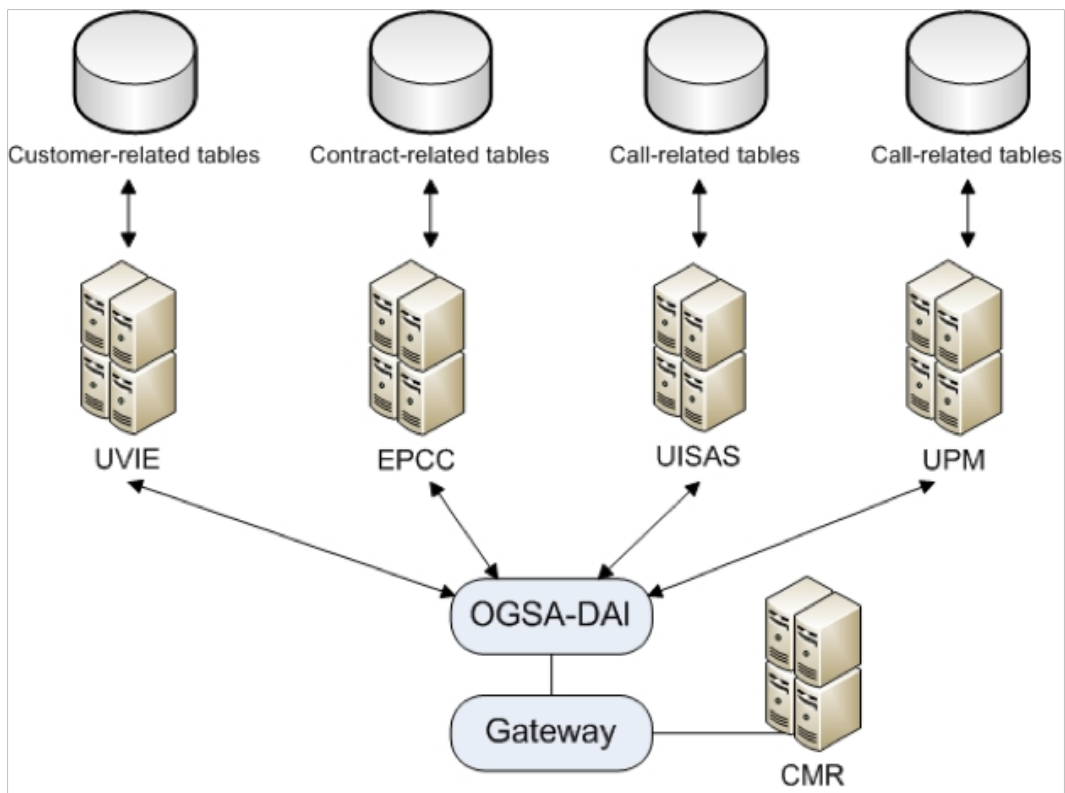


Figure 2: Planned Testbed setup for the Scale-up Experiment

4 Future Plans

ADMIRE Platform Release 3 is the third 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.

4.1 Project months 19-24

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

1. ADMIRE Platform Release 4
 - Improve particular Workbench plugins as well as their integration, also introduce Gateway Process Manager.
 - Improve the ADMIRE Gateway, including adding optimisations.
 - Improve the ADMIRE Registry, including adding insert operations.
 - Deploy the ADMIRE Repository.
 - Make the installation of the server-side software easier.
2. ADMIRE Testbed infrastructure
 - Roll out ADMIRE Platform Release 3 onto the nodes of all partners.
 - Create system tests based on use cases from WP6 which will exercise the whole platform and add them to the Dashboard.
 - Extend the Dashboard, including integrating with the ADMIRE Portal.
3. Deploy and run the ACRM scale-up experiment.

References

- [1] The ADMIRE Project. ADMIRE Platform Release 2, Sep 2009.
- [2] The ADMIRE Consortium. ADMIRE: Description of Work , Feb 2008.
- [3] The ADMIRE Project. ADMIRE Platform Release 1, Mar 2009.
- [4] Vivian Lee and work package partners. ADMIRE – Development and Deployment Report for USMT V3. Deliverable report D4.4, The ADMIRE Project, Feb 2010.
- [5] 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.
- [6] Vivian Lee and work package partners. ADMIRE – Development Progress Report. Deliverable report D4.3, The ADMIRE Project, Sep 2009.
- [7] Amy Krause, Ivan Janciak, Michal Laclavik, Branislav Simo, Maciej Jarka, Marek Lenart, Yuzhang Han, and Adrian Mouat. ADMIRE – Tools Development and Tools Integration Report. Deliverable report D5.4, The ADMIRE Project, Feb 2010.
- [8] Ondrej Habala, Maciej Jarka, Andrzej Biernacki, Maciej Gańczyk, and Viet Tran. ADMIRE – Pilot Applications Development and Evaluation Suite Report. Deliverable report D6.4, The ADMIRE Project, Feb 2010.