VL-e Proof of Concept Environment

Abstract

The Proof-of-Concept Environment (PoC) is the shared, common environment for e-Science of the Virtual Laboratory for e-Science. In the PoC, the different tools and services used by and provided by the project are available, and bound together in a service oriented approach. It consists of both a software distribution and an actual infrastructure that allows scaling and validation in real-life applications. The VL-e Integration Team organises and advises on the structure of the PoC. This document describes the structure of the PoC and the VL-e Integration Team.

Table of Contents

Revision History

Version (date)	*Comment*
Version 1.3 (8 Nov 2005 14:22)	Released for distribution SP leaders

1 The VL-e Proof of Concept environment

The VL-e Proof-of-Concept (PoC) Environment is a shared, common environment, where different tools and services are both used and provided by the VL-e community to enhance their research.

It is clear that, to reach a PoC that fulfils this vision, all subprograms should provide tools for the PoC, and the PoC should have an agreed common "exchange point" to integrate tools and services from different sources. Only then can tools and services be exchanged and the synergy leveraged.

The best way of achieving this goal is by making the integration layer of VL-e explicit. This integration layer will be centred around the concept of services, based on a *Service-Oriented Architecture* (SOA). In terms of the project, this is the "P2" generic virtual lab methodology layer.

By adopting a service-oriented paradigm, the integration of services from different sources will be facilitated, and it will allow incremental development by adding components with well-defined and self-contained functionality. This approach also facilitates the matching of components with demand (technology push versus application pull).

2 The VL-e Integration Team (VLEIT)

The VL-e Integration Team (VLEIT) is a (small) group of people which steers the development of the VL-e Proof-of-Concept (PoC) environment. Its main task is to recommend to the directorate the set of tools and services to be provided in each release of the PoC, and to give advice on the allocation of human resources to realise this new release. The VLEIT team is chaired by the scientific project director.

The PoC environment of VL-e is an integrative environment, shared by all (application) groups in the project, for which support and maintenance is supplied. Changes to the PoC are managed according to a time-boxed release schedule, and the environment is appropriately tested and engineered.

2.1 Procedure

The release cycle of the PoC will be bi-annual. Feature freeze dates are April 1st and October 1st of each year. After the feature freeze, 3 months are allocated for integration, certification, knowledge transfer to the PoC support people, and packaging (see below). The new PoC environment will thus be released 3 months after the feature freeze date. This version will be available on the PoC for 6 months following the release date, including security fixes and support for critical patches.

Special Interest Groups (SIGs) exist within the project. These SIGs will map a specific area of interest for the project and operate within well-defined terms of reference, for a specified amount of time, and must document their findings in writing. On completion, a SIG must report to VLEIT. This report is informational only. There may be cases where a functionality of interest is so easily defined that the formation of a SIG is overkill. In that case, a one-page description of the desired functionality is sufficient. VLEIT collects and prioritizes these reports and organizes the development and implementation of tools that are to be part of the next PoC.

At every feature freeze date, VLEIT produces a written recommendation on proposed changes for the new version of the PoC to the directorate. This document contains the tools and services proposed to be part of the next release, a proposal on who are most suitable for the implementation and a list of deprecated services (if any).

The directorate makes a decision on which tools and services are to be developed for the next release and by who, based on this recommendation.

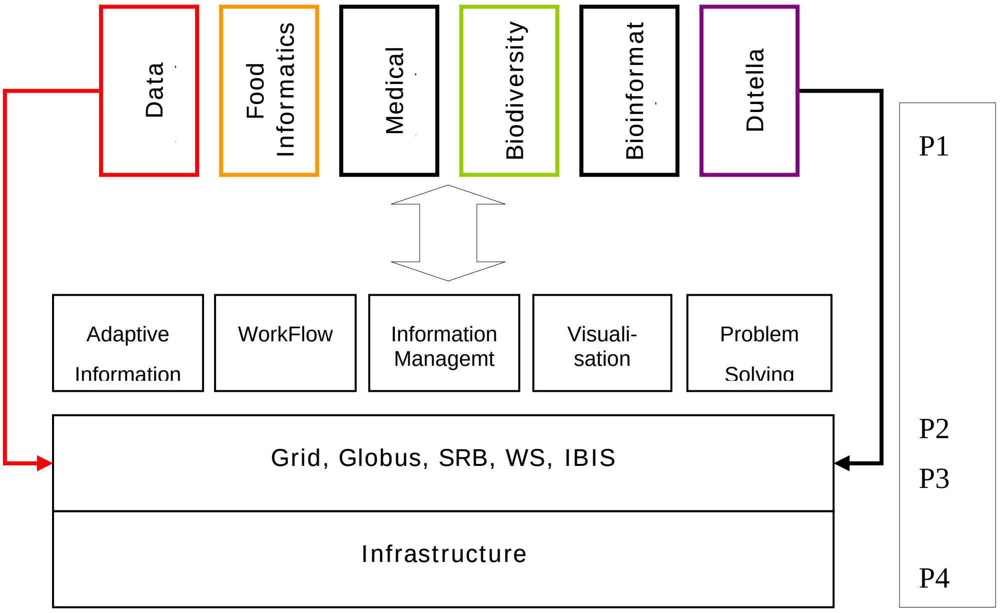

Figure 1. Structure of the VL-e proof-of-concept environment.

2.2 Human Resources

VLEIT consists of representatives for the different activities in the generic VL layer, the foundation middleware layer and for the PoC (resp. P2, P3 and P4) shown in Figure 1 and is chaired by the scientific director. The application subprograms are represented by a technical liaison. The members of VLEIT, as of November 1st, are:

Scientific Director (chair)	Bob Hertzberger
P1 representative	Robert Belleman
Adapt. Inform. Disclosure	Machiel Jansen
Workflow	Zhiming Zhao, Paul van Hooft, Adam Belloum
Information Management	Nader Mirzadeh, Hakan Yakali
Visualization	Robert van Liere
Problem Solving Environments	Breanndán Ó Nualláin
HPCN P3 representative	Kees Verstoep
Scaling and Validation	David Groep
PoC infrastructure	Maurice Bouwhuis
Industry representatives	*to be specified in detail in a forthcoming document*

Within the VL-e project, several *Scientific Programmers* have been hired as part of specific subprograms. Programmers within program lines P2 and P4 are obliged to spend a specific, predetermined amount of their effort on activities directly beneficial to the PoC. These are the resources that VLEIT can avail of to do the implementation and development work for the PoC. It must be noted that this work for the PoC can therefore be outside the research topics of the organisational unit to which that person has been appointed (the remainder of the time of said programmer can of course be spent according to the priorities of the organisational unit).

In addition, there can be situations where the implementation and development work for the PoC matches the area of interest of scientific programmers in P1 and/or PhD students (AIOs). When this happens, VLEIT should involve these people as much as possible and include this in their recommendations to the directorate.

3 Content of the PoC: Principles and Directions

The PoC is the shared common environment of the Virtual Laboratory. It consists of:

- physical infrastructure (such as computers, storage, network hardware),
- central services (database hosting, repository services, code development systems),
- the PoC software suite.

For convenience, the following elements of this suite can be distinguished:

1. *Grid foundation middleware ('P3')*; the basic software that is based on interfaces and concepts that are globally adopted and standardized. This includes elements such as the security model, resource allocation interface, access to (mass) storage, access to (but not the manipulation of) databases, network monitoring & reservation, the information system, and accounting.
2. *Generic Virtual Laboratory software ('P2')*; the software developed within the project for the PoC.
3. Services *imported from outside*; given that not all higher-level services are necessarily developed within VL-e, a significant fraction of these will be imported from elsewhere.
4. Associated *installation and deployment tools*; the PoC suite is installed on the central facilities and (where applicable) also available for distributed installation.

All these will evolve incrementally, according to the needs of the PoC users, the available (manpower) resources and the VLEIT prioritization procedure outlined above.

The PoC environment should also be firmly embedded in the current e-Science development trends world wide and respond in a timely manner to global events (new standards, promising new tools and solutions, shifts in software development paradigms). It will be the remit of the VLEIT team to assess such developments in terms of their effect on the PoC and potential for adoption by stimulating the awareness of these developments.

Closely linked to the PoC is the Certification Test-Bed (CTB). This system includes (almost) all the elements that are to be found in the PoC, but will run the expected next release of the software (grid foundation middleware, generic VL-e software, services). The Certification Test-Bed will act both as a testing and preview ground for new software, as well as the handover point for code, integration and deployment testing, and the point of knowledge transfer from the developers to the PoC support people (based on documented knowledge). The certification period takes almost three months, as stated earlier. The viewgraph in Figure 2 illustrates the certification and release procedure for the PoC.

A special arrangement exists for software that is obtained from an external source. In cases where external software is needed and this software has no dependencies on the current software configuration on the PoC, this software may be directly deployed on the PoC provided the software complies with the packaging, installation and configuration regulations defined by the CTB.

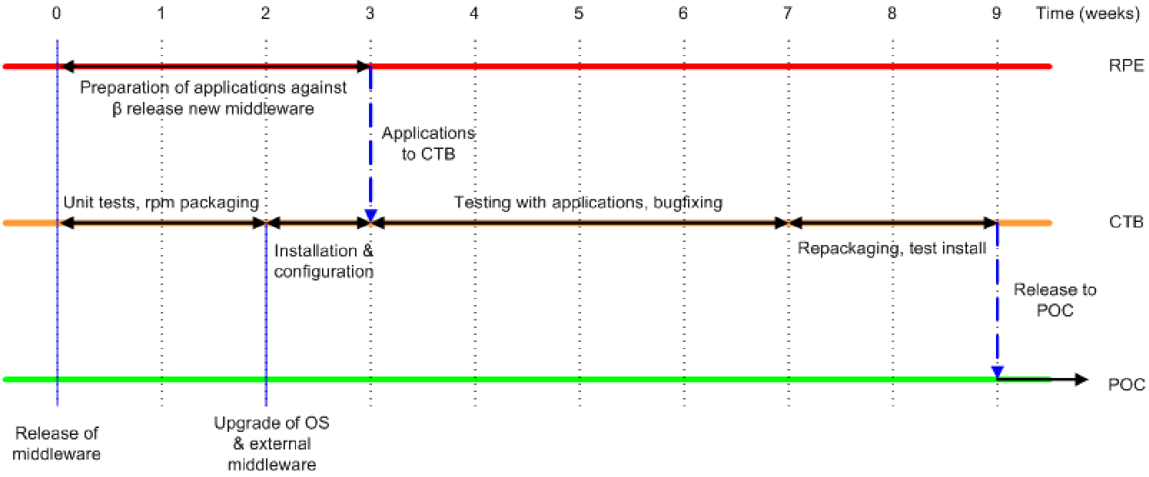
0
1
2
3
4
5
6
7
8
9
Time (weeks)
RPE
Preparation of applications against β release new middleware
Applications to CTB
Unit tests, rpm packaging
Testing with applications, bugfixing
CTB
Installation & configuration
Repackaging, test install
Release to POC
POC
Release of middleware
Upgrade of OS & external middleware

Figure 2: Certification and release procedure for the PoC. RPE: Rapid Prototyping Environment (or external source), CTB: Certification Test Bed, PoC: Proof-of-Concept facility.

3.1 Structure of the PoC Environment

Following the global trends in software development and composition, the functionality of the PoC should be centred on *services*. These services should represent basic functionality blocks and expose this functionality through a well-defined interface. Basic services can subsequently be combined into higher-level services that in turn expose a well-defined interface. This is referred to as a *Service Oriented Architecture (SOA)*. The service layer will be the integration and convergence point of the PoC.

3.1.1 Definition of Services: protocols and standards

Services in a SOA can be modelled in different ways. Currently the most popular way of representing services (at least those that are not time-critical) is as *Web Services*. Given the large international momentum behind this representation, it is both logical and advisable to render most of the services in the PoC as web services. For specific protocols that are critical to system performance, alternatives can be chosen that are widely recognised as standards or de-facto standards for the service domain (*e.g.*, file transport using GridFTP, inter-process communication in parallel programs using MPI or MPICH-G2).

Also, and in particular for the grid foundation middleware, the interfaces exposed by the services should comply with internationally recognised standards. Where available and tested, those standardized by bodies like GGF, OASIS and W3C should be adopted. Elsewhere, *de facto* standards (such as those set by efforts like OGSA-DAI, the Globus Toolkit, the Apache foundation, MyGrid, SRB, OSG, EGEE, NAREGI, HealthGrid) should be used where appropriate and sufficiently mature. It is the responsibility of the VLEIT team to assess and recommend the standards to be adopted, whereas monitoring and tracking of standards should be done wherever possible and relevant throughout the whole VL-e project.

3.1.2 Type of Services: Base and Value added

Base services provided by PoC will be the building blocks for value added services of PoC. These base services will stimulate the collaboration, help building the user community and encourage its use.

Each SP in P2 layer, considering the common basic needs of the applications, can provide few generic base services, as well.

Most of the added value of the PoC will be in generic high-level (‘virtual laboratory’) services, which either provide new functionality or combine new methodologies with existing grid foundation middleware or base (grid) services into new services. *What* functionality is to be developed and provided *when* is to be assessed by the VLEIT team. VLEIT should also be able to make recommendations on importing tools from outside the project that fill gaps identified in the PoC.

3.1.3 Integration: Workflow tools

To make the PoC useful for applications, integrating these services into coherent groups is needed. The current view of this layer of integration, build on top of the services layer, is that this integration should be realized via *workflow* tools. There are various workflow tools out there (currently being assessed) that are relevant to (parts of) the application domains. The vast majority of these workflow tools have adopted web services and the standard description language WSDL as the integrative interface. This will (at least initially) allow several workflow tools to co-exist within the PoC, and ultimately enable all of the tools to access the same services. The installation and maintenance of workflow environments on the PoC will be supported.

As long as no single common workflow environment has been agreed upon, it should be considered as an experimental service. When more than one workflow environment is used, not all will be given full-service support and additional support will be needed from the scientific programmers in P1. Thus, the concept of a single environment for sharing tools, information and knowledge is not impaired.

3.1.4 Data Integration Tools

Most of scientific applications depend on intensive access to data sources. The need of data source access can be for different purposes, e.g., for retrieving or storing experimental data.

SP2.4, Information Management group (IM) in VL-e will define a set of services for ease of data access/storage that can be used by scientists in their scientific experiments. IM has already defined a few generic base services that will be provided through PoC environment:

- Service-based on-demand storage repository: This service will enable any authorized person to create a specific storage area as needed. Creator of the storage area will have the full control over it, i.e., add, remove and update, as well as removing the repository itself.
- Service-based generic query interface to any database using JDBC drivers.

3.2 VL-e bi-weekly meetings

To increase awareness in both the generic services that are desired by application developers but also of international developments that are relevant for VL-e, the *VL-e bi-weekly meetings* will be re-organized. Representatives of research groups will be given the opportunity to present arguments on the importance and relevance to the VL-e project in general. These meetings will be announced to everybody in VL-e. The VLEIT team will make an effort to organize that representatives with a matching interest (i.e. service requesters and service providers and v.v.) are present.

4 Development Direction of the PoC

4.1 Infrastructure

The infrastructure will consist of central compute and storage facilities (accessible via various interfaces), storage meta-data servers (SRB), central database services (Oracle, PostgreSQL), software engineering services (gForge, CruiseControl) and a set of repository services and web sites.

The PoC central infrastructure will grow in terms of compute resources but especially in terms of data storage, by means of extensions (via grants by NCF) to Matrix and by extended storage capacity.

4.2 Access to the software and services

Access to the PoC will be preferably from the end-user's work place. This implies that the user will access the services though either software on the local workstation, installed and updated using tools provided by the PoC, through a (web) portal, or though a dedicated environment (e.g. for VR-style applications).

User interfaces will be distributed throughout the country (cf. systems currently at University of Utrecht, Philips Research, AMOLF). If software is installed on the local workstation, it is feasible to also offer services to the community through that workstation, but these services will not be considered part of the supported PoC infrastructure unless they are certified and approved by the PoC support group (at SARA).

It is intended, although not yet realised by release 2 (April 2006), that software that is installed on the PoC central facilities is available for installation at distributed locations around the country, using an installation/upgrade tool, where applicable.

4.3 Software suite architecture

The PoC will move to a service-based environment, rendered as web services. Unfortunately, *web services* as a definition is not sufficient for interoperability and the realisation of a shared environment. Therefore, wherever feasible, the service interface exposed should comply with the WS-Interoperability basic profile. For specific needs, WS draft standards can be used (such as WS-Addressing, WSRF), provided that the interface is extensible and easily modifiable, and the interface is not proprietary.

The lifetime of the XML-based, web services rendering of a service oriented architecture is likely to be reasonably long-lived, given the large amount of industrial support for the web services concept (but note that this does not extend to individual WS-* 'standards').

The PoC will not be exclusively based on web services. The PoC aims for interoperability with other projects, especially for the grid foundation middleware. Interoperability is essential to ensure that VL-e can effectively work with related projects, both today and in the future, in collaborations that may not yet be envisaged. This means that not all future developments and updates to the PoC on this layer can be foreseen on the 2-3 years timescale.

But also for performance reasons, the XML-based rendering as web services may not always be the best choice. In either case, the arguments (either for interoperability or for performance) should be documented and used as the basis for a decision by the VLEIT team.