SEVENTH FRAMEWORK PROGRAMME
Research Infrastructure

FP7-INFRASTRUCTURES-2010-2 – INFRA-2010-1.2.3:
Virtual Research Communities

Combination of Collaborative Project and Coordination and Support Actions (CP-CSA)

LinkSCEEM-2
Linking Scientific Computing in Europe and the Eastern Mediterranean – Phase 2

Grant Agreement Number: RI-261600

D8.2
Specifications and Implementation Plan for Software Environment Providing Remote Access to Resources

Final

Version: 1.0.
Author(s): Terry McLaren, NCSA
Nick J. Sinanis,Cyprus Institute
Date: 15/09/2011

Project and Deliverable Information Sheet

LinkSCEEM Project	**Project Ref. №: RI-261600**	
	Project Title: LinkSCEEM-2	
	Project Web Site: http://www.linksceem.eu	
	Deliverable ID: <D8.2>	
	Deliverable Nature: <DOC_TYPE: Report / Other>	
	Deliverable Level: PU *	**Contractual Date of Delivery:** 30 / 8 / 2011
		Actual Date of Delivery: 16 / 9 / 2011
	EC Project Officer: Leonardo Flores Anover	

* - The dissemination level are indicated as follows: **PU** – Public, **PP** – Restricted to other participants (including the Commission Services), **RE** – Restricted to a group specified by the consortium (including the Commission Services). **CO** – Confidential, only for members of the consortium (including the Commission Services).

Document Control Sheet

Document	**Title:** Report on implemented data management software environment installation	
	ID: <**D8.2**>	
	Version: <1.0 >	**Status:** Final
	Available at: http://eniac.cyi.ac.cy	
	Software Tool: Microsoft Word 2007	
	File(s): LinkSCEEM-2-D8.2.docx	
Authorship	**Written by:**	Terry McLaren, NCSA Nick J. Sinanis, CyI - CaSToRC
	Contributors:	Mostafa Zoubi, SESAME Basman Elhadidi, NARSS Mohammed Abd El-Monem Gaafar, BA
	Reviewed by:	Christos Nicolaou, CaSToRC
	Approved by:	PMO/SC

Document Status Sheet

Version	Date	Status	Comments
0.1	12/8/2011	Draft	
1.0	15/9/2011	Draft	
1.1	16/9/2011	Final	

Document Keywords

Keywords:	LinkSCEEM-2, Computational Science, HPC, e-Infrastructure, Eastern Mediterranean

Table of Contents

PROJECT AND DELIVERABLE INFORMATION SHEET II
DOCUMENT CONTROL SHEET II
DOCUMENT STATUS SHEET II
DOCUMENT KEYWORDS III
TABLE OF CONTENTS IV
LIST OF TABLES IV
REFERENCES AND APPLICABLE DOCUMENTS IV
LIST OF ACRONYMS AND ABBREVIATIONS IV
EXECUTIVE SUMMARY 6
1 INTRODUCTION 6
2 SPECIFICATIONS OF REMOTE ACCESS TO RESOURCES 7
3 IMPLEMENTATION PLAN 8
4 NEXT STEPS 8

List of Tables

Table 1 Requirements for Data Management tools 7
Table 2 Computational access requirements 7
Table 3 Implementation plan 8

References and Applicable Documents

[1] http://eewrc.cyi.ac.cy/DARECLIMED
[2] http://www.3dcoform.eu

List of Acronyms and Abbreviations

API	Application Programming Interface
BA	Bibliotheca Alexandrina
CaSToRC	Computation-based Science and Technology Research Centre of the CyI
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture (NVIDIA)
CyI	The Cyprus Institute
DM	Data Management
DR	Data Repository
EC	European Community
EEWRC	Energy, Environment and Water Research Center of the CyI
GB	Giga (= 2^{30} ~ 10^9) Bytes (= 8 bits), also GByte
Gb/s	Giga (= 10^9) bits per second, also Gbit/s
GB/s	Giga (= 10^9) Bytes (= 8 bits) per second, also GByte/s
GFlop/s	Giga (= 10^9) Floating point operations (usually in 64-bit, i.e. DP) per second, also GF/s

GHz	Giga (= 10^9) Hertz, frequency =10^9 periods or clock cycles per second
GigE	Gigabit Ethernet, also GbE
GNU	GNU's not Unix, a free OS
GPGPU	General Purpose GPU
GPU	Graphic Processing Unit
HDD	Hard Disk Drive
HPC	High Performance Computing; Computing at a high performance level at any given time; often used synonym with Supercomputing
HPCC	HPC Challenge benchmark, http://icl.cs.utk.edu/hpcc/
HPL	High Performance LINPACK
HWA	HardWare accelerator
IB	InfiniBand
IBA	IB Architecture
I/O	Input/Output
JSC	Jülich Supercomputing Centre (FZJ, Germany)
KB	Kilo (= 2^{10} ~10^3) Bytes (= 8 bits), also KByte
LQCD	Lattice QCD
LinkSCEEM	Linking Scientific Computing in Europe and the Eastern Mediterranean
LinkSCEEM-2	Linking Scientific Computing in Europe and the Eastern Mediterranean – Phase 2
MB	Mega (= 2^{20} ~ 10^6) Bytes (= 8 bits), also MByte
MB/s	Mega (= 10^6) Bytes (= 8 bits) per second, also MByte/s
MFlop/s	Mega (= 10^6) Floating point operations (usually in 64-bit, i.e. DP) per second, also MF/s
MHz	Mega (= 10^6) Hertz, frequency =10^6 periods or clock cycles per second
MPI	Message Passing Interface
MPP	Massively Parallel Processing (or Processor)
NARSS	National Authority for Remote Sensing And Space Sciences
STARC	Science and Technology in Archaeology Research Center of the Cyl
TB	Tera (= 240 ~ 10^{12}) Bytes (= 8 bits), also TByte
TFlop/s	Tera (= 10^{12}) Floating-point operations (usually in 64-bit, i.e. DP) per second, also TF/s
WP	Work Package

Executive Summary

In the context of the software environment that will provide the remote access to the resources, the tasks comprising this deliverable will define the necessary specifications and will conceive the implementation plan. For the purposes of defining the specifications of the system that will allow remote access to resources, we couple on the results obtained by the questionnaire described in deliverable D8.1. This is a first approximation of the specific needs and it provides only a crude input to the task of making an implementation plan.

The scientific communities engaged in LinkSCEEM-2 (Climate Research, High Energy Physics, Archeology, etc.) are all projecting data intensive activities either in the form of digital repositories or in data transfers and analyses. The exact timelines for satisfying the above needs vary and they mostly depend on external to this task and work-package (WP) factors e.g., Cy-Tera, SESAME facility.

Recognizing the current state of affairs, we record the current list of requirements together with providing a rough plan to implement a remote data access solution. It has become apparent that a second iteration will be required to complete the specifications and the implementation plan. It is therefore suggested that the work on this deliverable's objectives is continued for an additional six months to adjust with the status of the external dependencies and give additional time to evaluate the community needs.

1 Introduction

This second deliverable of WP8 aims to collect the specifications of a software system that will provide a distributed and remote access to data resources. Those resources can reside in Digital Repositories (DR) or other data storage means, at partner sites.

Partner sites, with their corresponding scientific communities, have an expressed need to effectively manage their data resources. This process requires DR to be put in place that will satisfy a diverse set of requirements of each scientific community.

As a starting point, the scientific communities participating in the project are approached to collect the requirements and based on this input the specifications of the system that will provide the remote access to resources will be assembled. Those communities are:

- Archeology
- High Energy Physics (HEP)
- Climate Research

The Archeology community has most of the need for remote access to resources and also a long research history on the topic of DR and remote access. To a lesser extent, the climate researchers and HEP could directly profit from the infrastructure offered by this WP.

2 Specifications of Remote Access to Resources

The suggested strategy to assemble the list of specifications is to start with a basic requirements collection and analysis. The questionnaire responses collected with the D8.1 work provided us with an initial list of requirements, which can be found in Table 1.

R1	Linked repositories
R2	Access control
R3	Sharing options
R4	Visualization and preview options (3D objects, images, etc...)
R5	Metadata options (Add, Edit)
R6	Data classification and categorization
R7	Capturing and integrating high-throughput data
R8	Integrating distributed software and hardware systems into a single application
R9	Data compression techniques
R10	Large scale data transfers

Table 1 Requirements for Data Management tools

From the same exercise the requirements on the computational management have been gathered and they are listed in Table 2.

CR1	Compatibility with current computation and visualization packages
CR2	Intuitive and user-friendly interfaces
CR3	Access control for both authentication and authorization
CR4	Status and monitoring (system utilization before and during jobs, current allocation, & job status)
CR5	Provision of a graphic interface to view system status, e.g. myJam
CR6	Ability to pre-empt jobs (hold, cancel)
CR7	Ability to prioritize jobs (LIFO, FIFO, RR, etc...)
CR8	Ability to exploit idle processors
CR9	Availability of the myJam tool on the Sun Grid Engine (SGE)

Table 2 Computational access requirements

The next step is to perform the requirements analysis in order to arrive to the list of specifications. For this, our suggestion is to proceed with further contacts with the relevant communities and also couple with similar projects active in the region, e.g., DARECLIMED [1]

Given that additional input, it will become possible to formulate a close to complete list of specifications for the remote access software system that will serve the scientific communities of the region.

3 Implementation Plan

The implementation plan has the objective to ensure that a proper system capable for remote access to resources will be deployed amongst the interested partner sites. As it has been mentioned before, a complete plan has as prerequisite a complete set of specifications to be defined. At the current state of the project, the input that has been collected allows us to work on an initial version of this plan.

The projected implementation plan foresees the following steps:

Collect user requirements (ongoing)
Assemble list of system specifications (ongoing)
Organize a technical specifications review amongst partner sites and respective communities with the objective to decide upon a tool or a set of tools
Lay out concrete deployment plan
Organize a technical review of the deployed system

Table 3 Implementation plan

An updated and expanded version of this plan will be provided as the user requirements and specifications definition processes advance.

4 Next Steps

With the gathered input from the scientific communities, we conclude that further work is required to reach the goals of this deliverable. Our expectation is to continue gathering the user requirements and expand the reach to more communities and other EU funded projects. This will result into concrete use cases that will easily satisfy the completeness of the gathered requirements and system specifications. Once this process completes, we will be confident that the implementation plan will meet the expectations and reach the scientific goals set by the communities.