

# **System and Software Requirements Definition (SSRD)**

## **Data Mining of Digital Library Usage Data**

### **Team 07**

#### **Clients**

Jewel Ward
------------

#### **Team Members**

Project Manager: Maxim Krivokon  
Developer: Bo Lee  
Developer: Genesan Kim  
Developer: Vu Nguyen  
IV&Ver: Shing-Cheung Chan  
IV&Ver: Marie Chi  
IV&Ver: Kristine Guevara

**<April 25<sup>th</sup> 2005>**

# Version History

Date	Author	Version	Changes made
9/28/04	Maxim Krivokon	1.0	<ul style="list-style-type: none"> <li>Initial draft</li> </ul>
9/30/04	Maxim Krivokon	1.1	<ul style="list-style-type: none"> <li>Incorporated client's suggestions</li> <li>Incorporated suggestions of team members</li> </ul>
10/01/04	Maxim Krivokon	1.2	<ul style="list-style-type: none"> <li>Added two new prototype screenshots</li> </ul>
10/10/04	Maxim Krivokon	1.3	<ul style="list-style-type: none"> <li>Replaced SR-1</li> <li>Removed PR-6</li> <li>Added Level Of Service and Evolution requirements</li> </ul>
10/24/04	Maxim Krivokon	1.4	<ul style="list-style-type: none"> <li>Added LR-5 = performance scalability as level of service requirement</li> <li>Elaborated off-nominal requirements in terms of availability and dependability</li> <li>Elaborated user front end to the analyzer and visualization modules</li> <li>Changed sections 4 and 6 to table format</li> </ul>
11/14/04	Maxim Krivokon	1.5	<ul style="list-style-type: none"> <li>Removed PR-3, Rephrased PR-4, PR-5. Split SR-1, IR-1</li> <li>Changes are pending client's approval</li> </ul>
11/30/04	Maxim Krivokon	1.6	<ul style="list-style-type: none"> <li>Moved PR-3 to PR-8; Added PR-9</li> <li>Introduced deployment requirement</li> <li>Removed DBA requirement</li> </ul>
12/3/04	Maxim Krivokon	1.7	<ul style="list-style-type: none"> <li>Introduced GNU compilation tools requirement</li> <li>Introduced corresponding deployment requirement</li> <li>Merged IR-1 and IR-2</li> <li>Removed LR-4</li> <li>Removed ER-2, 3 (because of stand alone architecture)</li> </ul>
1/30/05	Bo Lee	2.0	<ul style="list-style-type: none"> <li><b>[RLCA REVISION]</b></li> <li>5. Level of Service (L.O.S.) Requirements revised to be consistent with OCD 4.4 Level of Service (L.O.S.) Goals. (LR-1 – LR-5)</li> <li>Updated Table of tables for Table 32 – 35.</li> <li>Cover credits updated to 577b team.</li> <li>Entry for SR-11: Remove Analysis Report inserted.</li> </ul>
2/8/05	Bo Lee	2.0.1	<ul style="list-style-type: none"> <li><b>[RLCA REVISION]</b></li> <li>Apache log file -&gt; log file (Digital Archive usage log)</li> </ul>
2/9/05	Bo Lee	2.1	<ul style="list-style-type: none"> <li><b>[RLCA REVISION]</b></li> <li>Updated References</li> <li>Updated Section 5 Table 33, 35, 36 (from informal peer review)</li> </ul>

Date	Author	Version	Changes made
2/25/05	Bo Lee	2.2	<ul style="list-style-type: none"> <li>• <b>[RLCA after ARB]</b></li> <li>• (COMMENT) Entry for SR-11: Remove Analysis Report . Inserted in version 2.0 should be relocated but leave it there for this version, since it is too risky to do so for the references from other documents. it will be relocated for working set #1</li> <li>• (IV&amp;V) Revised Trace-back reference in Section 2</li> <li>• (IV&amp;V) Fixed Priority of LOS in table 34 and 36.</li> <li>• (IV&amp;V) Revised Trace-back reference in Section 5 LOS</li> </ul>
4/2/05	Bo Lee	3.0	<ul style="list-style-type: none"> <li>• <b>[IOC revision]</b></li> <li>• Fixed team website URL</li> <li>• (RLCA grading) Revised Table 2 PR-1 Limited Schedule</li> <li>• (RLCA grading) Revised Table 3 PR-2: No budget</li> <li>• (RLCA grading) Revised Table 9 PR-8: Target Operating Systems are Mac OS X and Linux  <b>-&gt; But this revision can be taken back since the target system is Mac OS X. Linux is an OS that the system supports to compile the source code</b></li> <li>• (RLCA grading) Revised Table 35 LR-3: Usability – Maximizing the usability of host resources</li> </ul>
4/3/05	Bo Lee	3.1	<ul style="list-style-type: none"> <li>• <b>[IOC revision]</b></li> <li>• Updated Table 16 SR-1: Importing usage data from log files.</li> <li>• Updated Table 35 LR-3</li> <li>• Updated Table 36 LR-4</li> </ul>
4/24/05	Bo Lee	4.0	<ul style="list-style-type: none"> <li>• <b>revised section 2 references and descriptions.</b></li> <li>• <b>revised section 3</b></li> <li>• <b>revised SR-1, SR-2, SR- 3.</b></li> <li>• <b>inserted SR-4, SR-5, SR- 6.</b></li> <li>• <b>revised SR-7, SR-8</b></li> <li>• <b>revised IR-1</b></li> <li>• <b>inserted IR-2, IR-3, IR-4, CLI-1</b></li> <li>• <b>revised LR-1, LR-2, LR-3, LR-4, LR-5</b></li> </ul>

# Table of Contents

<b>SYSTEM AND SOFTWARE REQUIREMENTS DEFINITION (SSRD)</b> .....	<b>I</b>
<b>VERSION HISTORY</b> .....	<b>II</b>
<b>TABLE OF CONTENTS</b> .....	<b>IV</b>
<b>TABLE OF TABLES</b> .....	<b>V</b>
1. Introduction .....	1
1.1 Purpose of the System and Software Requirements Definition Document.....	1
1.2 References .....	2
1.3 Change Summary .....	4
2. Project Requirements .....	6
2.1 Budget and Schedule .....	6
2.2 Development Requirements .....	7
2.3 Deployment Requirements .....	12
2.4 Transition Requirements.....	12
2.5 Support Environment Requirements .....	13
3. Capability Requirements .....	14
3.1 System Definition .....	14
3.2 System Requirements .....	15
5. System Interface Requirements.....	20
5.1 User Interface Standards Requirements .....	20
5.2 Hardware Interface Requirements .....	22
5.3 Communications Interface Requirements .....	22
5.4 Other Software Interface Requirements .....	22
6. Level of Service (L.O.S.) Requirements .....	23
7. Evolution Requirements.....	27
7.1 Capability Evolution Requirements.....	27
7.2 Interface Evolution Requirements .....	27
7.3 Technology Evolution Requirements .....	27
7.4 Environment and Workload Evolution Requirements.....	27
7.5 Level of Service Evolution Requirements .....	28
8. Common Definition Language for Requirements .....	29

# Table of Tables

<i>Table 1 Table 1 Changes Summary</i> .....	4
<i>Table 2 PR-1: Limited schedule</i> .....	6
<i>Table 3 PR-2: No budget</i> .....	6
<i>Table 4 PR-3: Use MS Project for scheduling and progress tracking</i> .....	7
<i>Table 5 PR-4: Use GNU compilation tools</i> .....	7
<i>Table 6 PR-5: Use efficient development language with 3d visualization capabilities</i> .....	8
<i>Table 7 PR-6: Use UML for modeling and design</i> .....	8
<i>Table 8 PR-7: Use free components</i> .....	9
<i>Table 9 PR-8: Target operating system is UNIX</i> .....	9
<i>Table 10 PR-9: Use visualization component that provides interactive 3d interface</i> .....	10
<i>Table 11 PR-10: Use data storage component that does not require maintenance</i> .....	10
<i>Table 12 PR-11: Input data format compatibility</i> .....	11
<i>Table 13 PR-12: Well documented code</i> .....	11
<i>Table 14 PR-13: Distribution consists of source code and binaries for Mac OS X and Linux</i> .....	12
<i>Table 15 PR-14: Maintainer training requirements</i> .....	12
<i>Table 16 SR-1: Relationship generation</i> .....	15
<i>Table 17 SR-2: Generate collection structure tree</i> .....	16
<i>Table 18 SR-3: Visualization</i> .....	17
<i>Table 19 SR-4: Search</i> .....	17
<i>Table 20 SR-5: Displaying Node information</i> .....	17
<i>Table 21 SR-6: Open the object webpage</i> .....	18
<i>Table 22 SR-7: Omit mal-formatted input file</i> .....	19
<i>Table 23 SR-8: Omit usage log file with 0 valid records</i> .....	19
<i>Table 24 SR-9: Do not visualize analysis report of corrupted format</i> .....	19

*Table 25 IR-1: Visualize analysis result* .....20

*Table 26 IR-2: Search Interface* .....20

*Table 27 IR-3: Object family navigation interface* .....20

*Table 28 IR-4 Webpage view interface*.....21

*Table 29 CLI-1: Running components with file name argument* .....22

*Table 30 IR-5: log format support*.....22

*Table 31LR-1: System Dependability - Stable data import/export* .....23

*Table 32 LR-2 : Usability – User-friendly interface for navigating the graph and selecting a node.* .....24

*Table 33 LR-3: Usability – Maximizing the usability of host resources*.....25

*Table 34 -4: Performance – Organizing data meaningfully for users*.....25

*Table 35 LR-5: Performance- data of current scale*.....26

*Table 38 ER-1: Advanced analysis*.....27

*Table 39 ER-2: Interface with other usage log formats*.....27

*Table 40 ER-3: Accommodating collection growth*.....28

*Table 41 ER-4: Workload growth should affect system performance in linear proportion*.....28

# 1. Introduction

## 1.1 Purpose of the System and Software Requirements Definition Document

The purpose of this document is to express the common vision among the project stakeholders in form of specific statements on what the Digital Archive (DA) Usage Data-Mining system should be, what it should do, how well it should do it and how it should be implemented and deployed. These specific statements establish clear objectives for future development process and guide the choices of architecture and third party components. For example, user interface requirements that have been developed as a result of prototyping effort and that are described in this document determine the standalone architecture of the system, and C/C++ as the main development environment. Another purpose of this document is to draw a clear line between optional and required features of the system in terms of functionality and quality of service, thus, defining the scope of this project more specifically. Namely this document distinguishes basic and advanced analysis capabilities that will be developed in two stages given successful transition between them.

This document extends and brings details to the requirements and constraints defined in the Operational Concept Description. As such it defines specifically the system capabilities and features that provide necessary contributions according to the system's Results Chain diagram presented in OCD. For example this document contains interface requirements providing for visualization of Digital Archive collection structure. The requirements defined in this document express agreements reached among the project stakeholders during regular meetings and Win-Win negotiations. Changes introduced to the project requirements by new agreements among stakeholders should be reflected as changes to this document.

## 1.2 References

“Data Mining of Digital Library Usage Data” Project Description

[http://sunset.usc.edu/classes/cs577a\\_2004/projects/description/project7.htm](http://sunset.usc.edu/classes/cs577a_2004/projects/description/project7.htm)

MBASE Guidelines version 2.4.2

[http://sunset.usc.edu/classes/cs577b\\_2005/guidelines/MBASE\\_Guidelines\\_v2.4.2.pdf](http://sunset.usc.edu/classes/cs577b_2005/guidelines/MBASE_Guidelines_v2.4.2.pdf)

MBASE Version 2.4.1 Templates for SSRD (version 1a)

[http://cse.usc.edu/classes/cs577a\\_2004/guidelines/MBASEtemplates/SSRD\\_Templatev1a.doc](http://cse.usc.edu/classes/cs577a_2004/guidelines/MBASEtemplates/SSRD_Templatev1a.doc)

MBASE Electronic Process Guide

<http://cse.usc.edu/research/MBASE/EPG>

Fall 2004 CS 577a Team 7 Easy Win-win report

[http://seacliff.usc.edu/~team7/LCO/EWW\\_LCO\\_F04a\\_T07.doc](http://seacliff.usc.edu/~team7/LCO/EWW_LCO_F04a_T07.doc)

Fall 2004 CS 577a Project #7 LCA version of OCD

[http://seacliff.usc.edu/~team7/LCO/OCD\\_LCO\\_F04a\\_T07\\_V5.3](http://seacliff.usc.edu/~team7/LCO/OCD_LCO_F04a_T07_V5.3)

USC Copyright Agreement

[http://cse.usc.edu/classes/cs577a\\_2004/forms/CopyrightAgreementV1.html](http://cse.usc.edu/classes/cs577a_2004/forms/CopyrightAgreementV1.html)

LCO ARB Prototype Demo

[http://seacliff.usc.edu/~team7/LCO/Prototypes/Prototype\\_V01.19.zip](http://seacliff.usc.edu/~team7/LCO/Prototypes/Prototype_V01.19.zip)

LCO ARB Presentation

[http://seacliff.usc.edu/~team7/LCO/ARB\\_LCO\\_F04\\_T07.ppt](http://seacliff.usc.edu/~team7/LCO/ARB_LCO_F04_T07.ppt)

LCA ARB Presentation

[http://seacliff.usc.edu/~team7/LCA/ARB\\_LCA\\_F04\\_T07.ppt](http://seacliff.usc.edu/~team7/LCA/ARB_LCA_F04_T07.ppt)

Dr. Bollen, Johan presentations

[http://www.cs.odu.edu/~jbollen/presentations/facstaff\\_02\\_28\\_03.pdf](http://www.cs.odu.edu/~jbollen/presentations/facstaff_02_28_03.pdf)

<http://www.cs.odu.edu/~jbollen/presentations/ecdl02.pdf>

<http://www.cs.odu.edu/~jbollen/presentations/aisti04.pdf>

Negotiations with the client

[http://seacliff.usc.edu/~team7/CMN/CMN\\_09\\_16\\_F04a\\_T07.doc](http://seacliff.usc.edu/~team7/CMN/CMN_09_16_F04a_T07.doc)

[http://seacliff.usc.edu/~team7/CMN/EMAIL\\_11\\_23\\_F04\\_T07.doc](http://seacliff.usc.edu/~team7/CMN/EMAIL_11_23_F04_T07.doc)

[http://seacliff.usc.edu/~team7/CMN/EMAIL\\_11\\_24\\_F04\\_T07.doc](http://seacliff.usc.edu/~team7/CMN/EMAIL_11_24_F04_T07.doc)

Team Website

<http://seacliff.usc.edu/~team7b>

## 1.3 Change Summary

Table 1 Table 1 Changes Summary

Version	Changes Made
1.0	Initial draft.
1.1	Complete the rest of Operational Concept Descriptions (OCD) document
1.2	Added two new prototype screenshots
1.3	Replaced SR-1, Removed PR-6, Added Level Of Service and Evolution requirements
1.4	Added LR-5 = performance scalability as level of service requirement Elaborated off-nominal requirements in terms of availability and dependability Elaborated user front end to the analyzer and visualization modules Changed sections 4 and 6 to table format
1.5	Removed PR-3, Rephrased PR-3, PR-5. Split SR-1, IR-1 Changes are pending client's approval
1.6	Moved PR-3 to PR-8; Added PR-9 Introduced deployment requirement Removed DBA requirement
1.7	Introduced GNU compilation tools requirement Introduced corresponding deployment requirement Merged IR-1 and IR-2 Removed LR-4 Removed ER-2, 3 (because of stand alone architecture)
2.0	Revised 5. Level of Service (L.O.S.) Requirements to be consistent with OCD 4.4 Level of Service (L.O.S.) Goals. (LR-1 – LR-5)
2.1	Revised 5 Level of Service (L.O.S.)
2.2	Revised 2 System Requirements trace back references Revised 5 Level of Service priority and trace back references.
3.0	Fixed team website URL Revised Table 2 PR-1 Limited Schedule for better description Revised Table 3 PR-2: No budget for measurable and achievable. Revised Table 9 PR-8: Target Operating Systems are Mac OS X and Linux to include Linux as a target system. Revised Table 35 LR-3: Usability – Maximizing the usability of host resources to resolve vague expression
3.1	Updated Table 16 SR-1: Importing usage data from log files for better description.
4.0	revised section 2 references and descriptions. revised section 3 revised SR-1, SR-2, SR- 3. inserted SR-4, SR-5, SR- 6. revised SR-7, SR-8 revised IR-1 inserted IR-2, IR-3, IR-4, CLI-1

	revised LR-1, LR-2, LR-3, LR-4, LR-5
--	--------------------------------------

## 2. Project Requirements

This section contains the mandates which were agreed on during Win-Win and other negotiations with the client and completion of which will determine project's success. These mandates are expressed as constraints on budget, schedule, specific features and capabilities, development and transition processes.

### 2.1 Budget and Schedule

**Table 2 PR-1: Limited schedule**

<b>Project requirement:</b>	PR-1: Limited schedule
<b>Description:</b>	Development of LCO and LCA packages has to be completed by the end of the Fall semester. The system should be finished at the end of the Spring semester.
<b>Measurable:</b>	LCO package is scheduled for delivery on 10/25/2004 and LCA package on 12/06/2004. IOC should be delivered by the early middle of May 2005. All should be delivered on time.
<b>Achievable:</b>	The project uses Schedule as Independent Variable process to achieve the deliverables and comply with deadlines. Development proceeds according to a defined schedule with several intermediate deliverables.
<b>Relevant:</b>	Realizes: OCD 4.2.1, PG-1, Win condition 1.1, OCD 4.2.4 PC-1
<b>Specific:</b>	Specifies deliverables and dates
<b>Priority:</b>	High

**Table 3 PR-2: No budget**

<b>Project requirement:</b>	PR-2: No budget
<b>Description:</b>	The design and development efforts will not be financed by the client and will be accomplished by students for credit.
<b>Measurable:</b>	No money spent for the development itself, but the effort of the developers and the client may count.
<b>Achievable:</b>	Using open source library or freeware.
<b>Relevant:</b>	Realizes: OCD 4.2.3, Win condition 1.2, OCD 4.2.5 PC-2
<b>Specific:</b>	Specifies what parts of the project will be affected by the lack of budget and how uncompensated development is motivated
<b>Priority:</b>	High

## 2.2 Development Requirements

### 2.2.1 Tools Requirements

Table 4 PR-3: Use MS Project for scheduling and progress tracking

<b>Project requirement:</b>	PR-3: Use MS Project for scheduling and progress tracking
<b>Description:</b>	MS Project should be used for creating a schedule for the project's main tasks, monitoring the project's progress, evaluation of alternative project plans and for tracking of individual performance.
<b>Measurable:</b>	Produced schedules, reports, evaluations should comply with standards / formats established by MS Project documentation
<b>Achievable:</b>	All team members have accomplished MS Project tutorial and a related homework assignment. LCP document already contains planning for inception and elaboration phases that was done using MS Project.
<b>Relevant:</b>	Realizes Win condition 1.6. Using MS Project will provide for more productive project management, and will allow producing documentation artifacts that are compliant with industry accepted standards. OCD 4.2.4 PC-1 can be relieved.
<b>Specific:</b>	Addresses reasons to use MS Project
<b>Priority:</b>	Medium

Table 5 PR-4: Use GNU compilation tools

<b>Project requirement:</b>	PR-4: Use GNU compilation tools
<b>Description:</b>	Configuration and compilation of the system source code should be automated using GNU tools. GNU Make of version 3.80 and up should be used. Use of other tools such as autoconf, automake is optional.
<b>Measurable:</b>	The produced code is compiled by running the corresponding GNU utility program.
<b>Achievable:</b>	All open source components to be used by the system are developed using GNU tools. Several team members have experience using them.
<b>Relevant:</b>	Ensures ease of system installation. Realizes win condition 3.3, OCD 4.2.6 PC-3
<b>Specific:</b>	Addresses how the compilation should be automated
<b>Priority:</b>	Medium

## 2.2.2 Language Requirements

Table 6 PR-5: Use efficient development language with 3d visualization capabilities

<b>Project requirement:</b>	PR-5: Use efficient development language with 3d visualization capabilities
<b>Description:</b>	Major system capabilities should be implemented using efficient, compiled language that provides extensive 3d visualization and graphical user interface capabilities.
<b>Measurable:</b>	Examples invoking 3d visualization and GUI capabilities of the language should be available. The language should have a compiler that produces binary executable code.
<b>Achievable:</b>	C++ is fast, compiled language that has extensive 3d graphics Application Programming Interfaces (API) (e.g. "OpenGL", "DirectX"). Several team members have development experience in C++. H3viewer visualization library is implemented in C++. Also, clustering component MCL is written in C.
<b>Relevant:</b>	Realizes OCD 4.2.4, OCD 4.2.5
<b>Specific:</b>	Addresses which features of development language are most important.
<b>Priority:</b>	High

Table 7 PR-6: Use UML for modeling and design

<b>Project requirement:</b>	PR-6: Use UML for modeling and design
<b>Description:</b>	System models and diagrams should be designed in compliance with definitions of UML language.
<b>Measurable:</b>	Produced artifacts can be compared with examples from UML specifications
<b>Achievable:</b>	UML tutorials and specifications are widely available
<b>Relevant:</b>	Realizes Win condition 1.6, Relieves OCD 4.2.3 PC-1
<b>Specific:</b>	Addresses the modeling standards that should be followed
<b>Priority:</b>	Medium

## 2.2.3 Computer Hardware Requirements

The system is a regular user application that does not access hardware directly and does not utilize any special hardware devices. The system will use interfaces for hardware access provided by the host operating system.

## 2.2.4 Computer Hardware Resource Utilization Requirements

The proposed system will be using hardware resources through interfaces provided by the host operating system and therefore will be limited in the amount of resources it can use. No extraordinary hardware resource usage is expected.

## 2.2.5 Computer Software Requirements

**Table 8 PR-7: Use free components**

<b>Project requirement:</b>	PR-7: Use free components
<b>Description:</b>	Third-party software components reused for developing the system must be open source or freeware.
<b>Measurable:</b>	For each third-party component its license should provide for free non-commercial use in distributed or modified form.
<b>Achievable:</b>	Availability of the following open source components has been confirmed: implementations of graph clustering algorithms, graph visualization libraries, RDBMS.
<b>Relevant:</b>	Realizes PG-1, Win condition 1.2, Relieves OCD 4.2.5 PC-2, OCD 4.2.6 PC-3.
<b>Specific:</b>	Specifies which functionality can be implemented with open source components.,
<b>Priority:</b>	High

**Table 9 PR-8: Target operating system is UNIX**

<b>Project requirement:</b>	PR-8: Target operating system is UNIX
<b>Description:</b>	The developed system should be operational on UNIX server
<b>Measurable:</b>	Developed system provides functionality that satisfies all system requirements when run onUNIX.
<b>Achievable:</b>	Prototyping and evaluation confirmed that identified third-party components are compatible with Unix based systems. Using standard C/C++ that is compilable in UNIX.
<b>Relevant:</b>	Realizes Win condition 1.3. Realizes additional agreements with the client (see references), Relieves OCD 4.2.3. PG-3
<b>Specific:</b>	Specifies main operating system
<b>Priority:</b>	High

**Table 10 PR-9: Use visualization component that provides interactive 3d interface**

<b>Project requirement:</b>	PR-9: Use visualization component that provides interactive 3d interface.
<b>Description:</b>	The project should use visualization component that provides interactive 3d graph visualization capabilities.
<b>Measurable:</b>	Examples of three dimensional visual representations of data should be available for the component used by the project. Interface provided by the component accepts user input and dynamically changes visualization behavior based on that.
<b>Achievable:</b>	H3Viewer library provides 3d hyperbolic interactive visualization of arbitrary input graphs. Sample visualization tool based on H3Viewer library was demonstrated at the LCO Architecture Review Board (ARB). Tailored graphical interface using H3Viewer was presented at the LCA ARB.
<b>Relevant:</b>	Realizes: OCD 4.2.1, 4.2.2, 4.2.3, 4.2.4, 4.2.5, 4.2.6
<b>Specific:</b>	Specifies requirements for visualization platform to be used
<b>Priority:</b>	High

**Table 11 PR-10: Use data storage component that does not require maintenance.**

<b>Project requirement:</b>	PR-10: Use data storage that does not require maintenance.
<b>Description:</b>	The system should use data storage component that does not require additional maintenance except regular user input.
<b>Measurable:</b>	System installation does not require additional setup of data storage. The system stores / retrieves data seamlessly during operation not requiring any assistance for those actions.
<b>Achievable:</b>	Data stored in files does not require maintenance, since file system is maintained by the operating system. Database components that do not require database server installation are available (e.g. SQLite).
<b>Relevant:</b>	Ensures ease of system use. Realizes win condition 3.3 OCD 4.2.5 PC-2
<b>Specific:</b>	Addresses which component of the system is critical for usability
<b>Priority:</b>	High

## 2.2.6 Computer Communication Requirements

The proposed system is a standalone application. All data necessary for performing usage analysis will be provided by the user and will be accessed by the system from the local computer. Thus no communication with other computers will be necessary.

## 2.2.7 Standards Compliance Requirements

**Table 12 PR-11: Input data format compatibility**

<b>Project requirement:</b>	PR-11: Input data format compatibility
<b>Description:</b>	The developed system must be able to parse usage logs of the format that is used by the Digital Library web server.
<b>Measurable:</b>	The system is able to parse usage log file of specified format and store the retrieved data to the system's data storage(file system)
<b>Achievable:</b>	Parsing facilities are widely available in C++ and other languages. The team has access to sample log files.
<b>Relevant:</b>	Realizes PG-2
<b>Specific:</b>	Specifies the format of input data that must be supported
<b>Priority:</b>	High

**Table 13 PR-12: Well documented code**

<b>Project requirement:</b>	PR-12: Well documented code
<b>Description:</b>	The code developed or tailored by the project team should be reasonably documented to provide for ease of future development and maintenance.
<b>Measurable:</b>	Adequacy of documentation will be determined through customer reviews of each major deliverable.
<b>Achievable:</b>	Customer reviews will ensure proper level of documentation detail
<b>Relevant:</b>	Realizes Win condition 1.7, OCD 4.2.4 PC-1, 4.2.5 PC-2
<b>Specific:</b>	Addresses the mechanism of achieving proper documentation
<b>Priority:</b>	High

## 2.3 Deployment Requirements

Table 14 PR-13: Distribution consists of source code and binaries for UNIX

<b>Project requirement:</b>	PR-13: Distribution consists of source code and binaries for UNIX.
<b>Description:</b>	The system distribution package should include system source that can be configured and compiled on UNIX(LINUX) using GNU compilation tools. The distribution should also include a binaries executable for UNIX (LINUX)
<b>Measurable:</b>	Delivered source code compiles on UNIX(LINUX) and produces a functional binary. Distributed binaries are fully functional.
<b>Achievable:</b>	The system will be developed using GNU compilation tools and therefore final development snapshot will be transformed into distribution package.
<b>Relevant:</b>	Ensures ease of system use. Realizes win condition 3.3 OCD 4.2.3 PG-3
<b>Specific:</b>	Addresses contents of distributable package
<b>Priority:</b>	Medium

## 2.4 Transition Requirements

Table 15 PR-14: Maintainer training requirements

<b>Project requirement:</b>	PR-14: Maintainer training requirements
<b>Description:</b>	Potential maintainer for the system should have experience designing and debugging programs in C/C++. Also the maintainer should have enough background knowledge about operations of the Digital Archive and rationale behind this project.
<b>Measurable:</b>	C/C++ experience can be established from previous experience. Background knowledge can be evaluated during interview
<b>Achievable:</b>	Among Digital Archive information technology staff there are employees who have sufficient C/C++ experience. Archive and project background knowledge can be provided through educational presentations
<b>Relevant:</b>	Realizes Win condition 1.9 OCD 4.2.3 PG-3
<b>Specific:</b>	Specifies requirements for maintainer
<b>Priority:</b>	Medium

## **2.5 Support Environment Requirements**

The system is expected to be stable and will not require extensive support.

## 3. Capability Requirements

### 3.1 System Definition

The system defined in this document is a Digital Archive usage data-mining tool with visualization capabilities. It is a standalone application operational with minimal dependencies. The proposed system will have only one class of users and therefore one operational mode. (OCD Figure 9)

The system takes usage log files as input, parses, analyzes, and normalizes usage records contained in the input files and stores the results in files, and displays the results in graphical interface.

There are three major modules of the system:

- Log Analysis: analyzing the input log data and generate relationships among the items in the input file.
- Relation Clustering: clustering the relationships.
- Visualization: presenting graphical and interactive analysis results.

Log Analysis module capabilities:

- parsing the input data file.
- generation of weighted relations between items based on their co-retrievals
- relative centrality/impact ranking of items

Relation Clustering module capabilities:

- clustering of items based on their relations
- construction of structure tree based on the clustering and centrality analysis

Visualization module capabilities:

- 3d hyperbolic view of the structure tree
- navigation capabilities within the graph view: rotation, panning
- searching
- displaying node information
- opening a webpage for an object.

## 3.2 System Requirements

### 3.2.1 Nominal requirements

**Table 16 SR-1: Relationship generation**

<b>System requirement:</b>	SR-1: Relationship generation
<b>Description:</b>	<p>Provided that usage log data was properly imported and analysis options were set the analyzer module performs the analysis of relationships among Digital Archive items based on their co-retrievals. As a result the module produces a graph representation of the DA collection with the following features:</p> <ul style="list-style-type: none"> <li>- a node represents a DA item</li> <li>- a link and its weight represents strength of relatedness between two items</li> </ul> <p>The produced graph is stored in files.</p>
<b>Measurable:</b>	Generated relations are stored in files and can be reviewed in a text file.
<b>Achievable:</b>	Examples of this analysis are provided in Johan Bollen's papers (See references)
<b>Relevant:</b>	Realizes: Win condition 2.2 , SC-2
<b>Specific:</b>	Describes how item relations are generated
<b>Priority:</b>	High

Table 17 SR-2: Generate collection structure tree

<b>System requirement:</b>	SR-2: Generate collection structure tree
<b>Description:</b>	Based on the detailed relationship graph collection structure tree should be produced using centrality analysis and graph clustering. For each tree level the item with highest centrality rank in the current graph is selected to be the local root. All of remaining items of the current graph are clustered based on strengths of their relations and resulting sub-graphs are made children of the local root. Then the procedure is repeated for each child sub-graph containing more than N nodes. The resulting tree is stored in files.
<b>Measurable:</b>	After the generation is complete the tree with the following properties should be stored in files: - the produced collection structure tree has items as regular nodes and sets of items as leaf nodes. Leaf nodes contain up to N items - for each tree level items in different sub-trees on average have weaker relations than items in the same sub-tree. – all parents of an item in the tree have higher centrality rank. -all children of an item in the tree have lower centrality rank.
<b>Achievable:</b>	Markov Cluster Algorithm has been evaluated and successfully used for prototype demo at the LCO ARB. Centrality ranking is done by simple statistic based on item relations.
<b>Relevant:</b>	Realizes SC-2, Win condition 2.3. Ensures visualization / browsing of the item relationship graph
<b>Specific:</b>	Describes hierarchical structure of clustering
<b>Priority:</b>	Medium

**Table 18 SR-3: Visualization**

<b>System requirement:</b>	SR-3: Visualization
<b>Description:</b>	Provided that the analysis was done and results stored in the files. The visualization module should be able to load them and provide a 3d hyperbolic view of the collection structure tree. The view should provide the following capabilities: - navigation of the tree (rotation / panning) - display usage statistics associated with a selected node
<b>Measurable:</b>	The collection structure tree and associated statistics is presented according to specifications and prototype examples
<b>Achievable:</b>	Major visualization capabilities were demonstrated by the prototype at the LCO ARB. Extended graphical user interface allowing display of node statistics was demonstrated by the prototype at the LCA ARB.
<b>Relevant:</b>	Realizes: Win condition 3.1, 3.4, SC-3
<b>Specific:</b>	Specifies the required visualization capabilities
<b>Priority:</b>	High

**Table 19 SR-4: Search**

<b>System requirement:</b>	SR-4: Search
<b>Description:</b>	Provided that the analysis was done and the collection structure tree is displayed in the 3d hyperbolic view. The search capability should find a node in the tree, and focus and highlight it.
<b>Measurable:</b>	Search function works fine with exact object id which is in the tree, all other invalid keyword does not make any difference to the visualization.
<b>Achievable:</b>	The fully functional search capability was demonstrated at TRR ARB.
<b>Relevant:</b>	Realizes: OCD 4.2.3
<b>Specific:</b>	Specifies the visualization capabilities
<b>Priority:</b>	High

**Table 20 SR-5: Displaying Node information**

<b>System requirement:</b>	SR-5: Displaying Node information
----------------------------	-----------------------------------

<b>Description:</b>	Provided that the analysis was done and the collection structure tree is displayed in the 3d hyperbolic view. The family nodes' object ids of selected or found node are displayed in the right-hand side panel of the screen.
<b>Measurable:</b>	The four boxes on the right-hand side panel hold parents(ancestors), siblings, and children nodes' object ids.
<b>Achievable:</b>	The fully functional displaying node information capability was demonstrated at TRR ARB.
<b>Relevant:</b>	Realizes: OCD 4.2.3
<b>Specific:</b>	Specifies the visualization capabilities
<b>Priority:</b>	High

Table 21 SR-6: Open the object webpage

<b>System requirement:</b>	SR-6: Open the object webpage
<b>Description:</b>	Provided that the analysis was done and the collection structure tree is displayed in the 3d hyperbolic view. On a selected or found object id, the system launches a web browser window and open the webpage for the object id.
<b>Measurable:</b>	The system uses hard-coded URL to open a website.
<b>Achievable:</b>	The fully functional opening the object webpage capability was demonstrated at TRR ARB.
<b>Relevant:</b>	Realizes: OCD 4.2.3
<b>Specific:</b>	Specifies the visualization capabilities
<b>Priority:</b>	High

## Off-nominal requirements

**Table 22 SR-7: Omit mal-formatted input file**

<b>System requirement:</b>	SR-7: Omit mal-formatted input file.
<b>Description:</b>	If in provided usage log file a retrieval record is met that does not comply with log format, it should be skipped.
<b>Measurable:</b>	Importing log file with mal-formatted entries should produce empty output files.
<b>Achievable:</b>	log format is specified in corresponding documentation
<b>Relevant:</b>	Addresses possible usage log data corruption
<b>Specific:</b>	Details actions of the system
<b>Priority:</b>	High

**Table 23 SR-8: Omit usage log file with 0 valid records**

<b>System requirement:</b>	SR-8: Omit usage log file with 0 valid records.
<b>Description:</b>	If the provided usage log file contains no valid retrieval records, the system should produce empty output files.
<b>Measurable:</b>	The output files should be empty.
<b>Achievable:</b>	log format is specified in corresponding documentation.
<b>Relevant:</b>	Addresses possible usage log data corruption
<b>Specific:</b>	Details actions of the system in case of mal-formatted file
<b>Priority:</b>	Medium

**Table 24 SR-9: Do not visualize analysis report of corrupted format**

<b>System requirement:</b>	SR-9: Do not visualize analysis report of corrupted format
<b>Description:</b>	If in the analysis results of the chosen report are of corrupted format then visualization module should stop visualization action and should provide user with corresponding error message.
<b>Measurable:</b>	Visualization stops if corrupted analysis report is chosen
<b>Achievable:</b>	Format of analysis results could be explicitly defined in terms of field data types, data ranges and regular expressions.
<b>Relevant:</b>	Addresses possibility of analysis data corruption
<b>Specific:</b>	Details how format for analysis results could be defined
<b>Priority:</b>	Medium

## 5. System Interface Requirements

### 5.1 User Interface Standards Requirements

#### 5.1.1 Graphical User Interface Standards

**Table 25 IR-1: Visualize analysis result**

<b>Interface requirement:</b>	IR-5: Visualize analysis result
<b>Description:</b>	After opening analysis report user should be provided with the following: <ul style="list-style-type: none"> <li>- 3d hyperbolic view of the collection structure tree</li> <li>- navigation of the tree using mouse: rotation pivoted at the center of the current view, panning (left, right, up, down).</li> <li>- centering of the view on a clicked item and orienting a tree such that all item ancestors are on the left, and all descendants on the right</li> <li>- displaying associated usage statistics for selected item</li> </ul>
<b>Relevant:</b>	Realizes: Win condition 3.4
<b>Priority:</b>	High

**Table 26 IR-2: Search Interface**

<b>Interface requirement:</b>	IR-2: Search Interface
<b>Description:</b>	Given 3d view of structure tree, search field should be on the top of right-hand side panel of the screen. After filling in the search field, pressing “enter” key will trigger search.
<b>Relevant:</b>	Realizes: OCD 4.2.3
<b>Priority:</b>	High

**Table 27 IR-3: Object family navigation interface**

<b>Interface requirement:</b>	IR-3: Object family navigation interface
<b>Description:</b>	Given 3d view of structure tree, object family members –

	parents, siblings, and children – will be displayed on the right-hand side of the screen as object ids. Users can click on the object ID to focus and highlight the node in the view and to fill the search field with the object id.
<b>Relevant:</b>	Realizes: OCD 4.2.3
<b>Priority:</b>	High

**Table 28 IR-4 Webpage view interface**

<b>Interface requirement:</b>	<b>IR-4 Webpage view interface</b>
<b>Description:</b>	Given 3d view of structure tree, and a node in the view or an object id on the right-hand side panel is selected. On clicking “view” button which is located on the top of the right-hand side panel, the system will launch a default web browser.
<b>Relevant:</b>	Realizes: OCD 4.2.3
<b>Priority:</b>	High

## 5.1.2 Command-Line Interface Requirements

Table 29 CLI-1: Running components with file name argument

<b>Interface requirement:</b>	<b>CLI-1 Running components with file name argument.</b>
<b>Description:</b>	Given all the input files, launching and feeding the input/output file name is done by command-line interface. A simple one line command such as: \$[executable filename] [inputfile name] <enter> Giving more controllability to users.
<b>Relevant:</b>	Realizes: OCD 4.2.3
<b>Priority:</b>	High

## 5.1.3 Diagnostics Requirements

User does not require any specific debugging information to be provided. Thus debugging capability will be determined during the development phase and tailored to the needs of the development team and system maintainer.

## 5.2 Hardware Interface Requirements

No hardware devices are accessed directly by the system

## 5.3 Communications Interface Requirements

The proposed system is a standalone application that does communicate with remote computers.

## 5.4 Other Software Interface Requirements

Table 30 IR-5: log format support

<b>Interface requirement:</b>	IR-5: log format support
<b>Description:</b>	The analyzer module should support the logging format that is used by the web server.
<b>Relevant:</b>	Necessary for SR-1
<b>Priority:</b>	High

## 6. Level of Service (L.O.S.) Requirements

Table 31 LR-1: System Dependability - Stable data import/export

<b>Level of Service:</b>	LR-1: System Dependability - Stable data import/export
<b>Description:</b>	The system should be able to perform importing/exporting data as many and frequent as desired without crash or unexpected restart.
<b>Measurable:</b>	System dependability should be measured as the rate of successful importation and exportation. The output files correctly reflect the input files and the processing of the component, and the system eventually.
<b>Achievable</b>	Valid format check before and after the importing/exporting process. Robust error handling mechanism that suggests alternative way to achieve a task when the first way failed.
<b>Relevant:</b>	SR-1: Relationship generation SR-2: Generate collection structure tree SR-3: Visualization SR-7: Omit mal-formatted input file SR-8: Omit usage log file with 0 valid records SR-9: Do not visualize analysis report of corrupted format OCD 4.4 LS-1: Dependability OCD 3.2 OG-1
<b>Specific:</b>	SR-1 the system should import data from ORACLE-formatted file and produce output in the format compatible to clustering component MCL For SR-2, The system should import files for MCL clustering component and produce output in the format of H3Viewer compatible. For SR-3, the visualization component imports a file in h3Viewer format(.h3v) Format check for SR-7 and file integrity check for SR-8 In case of user abortion, analysis report is not created and the data remain unchanged. (SR-9)
<b>Desired level</b>	The system should not crash while importing and exporting data with the successful importation and exportation rate of 95% of all times when running constantly with the 10 average times of importing and exporting a day for 5 days. No loss of data on unexpected crash while exporting. No program crash with any fatal error on importing/exporting modules.
<b>Accepted level</b>	The system should not crash while importing and exporting data with the successful importation and exportation rate of 90% of all times when running constantly with the 10 average times of importing and exporting a day for a day. No loss of data 90% of time on unexpected crash while exporting.
<b>Priority:</b>	High

**Table 32 LR-2 : Usability – User-friendly interface for navigating the graph and selecting a node.**

<b>Level of Service:</b>	Usability – User-friendly interface for navigating the graph and selecting a node.
<b>Description:</b>	Provide user-friendly interfaces for navigating the graph and selecting a node. Moving the graph, selecting a node, searching, selecting object id.
<b>Measurable:</b>	User will provide usability feedback Correctness in visualizing the item relationships as a graph. Visualization of the item relationship as a tree in 3D. Data integrity when loading several data files consequently. User can be accustomed to the system use in a trivial amount of time.
<b>Achievable</b>	Internal and external usability test. Algorithm verification mechanism for correctness in item relationships. Using Graphic API for 3D display.
<b>Relevant:</b>	SR-3: Visualization SR-4: Search SR-5: Displaying Node information SR-6: Open the object webpage IR-1: Visualize analysis result IR-2: Search Interface IR-3: Object family navigation interface IR-4 Webpage view interface OCD 4.4 LS-2: Usability OCD 3.2 OG-1, OG-3
<b>Specific:</b>	User should feel easy to move and drag the graph and select a node (SR-3, IR-1) The search field and searching procedure should be done without any instruction (SR-4, IR-2) Selecting an object id on the node information should be easy and user should instantly know the results relevant to his/her action (SR-5, IR-3) Opening a web browser should be performed by one click (SR-6, IR-4)
<b>Desired level</b>	Usability feedback with overall score 8 points out of 10. 99% correctness in visualizing the relationships. Flawless 3D view. No loss or corrupted data.
<b>Desired level</b>	Usability feedback with overall score 6 points out of 10. 90% correctness in visualizing the relationships. No loss or corrupted data 90% of time.
<b>Priority:</b>	High

**Table 33 LR-3: Usability – Maximizing the usability of host resources**

<b>L.O.S. Requirement:</b>	LR-3: Host Resource Availability – Maximizing the host resources
<b>Description:</b>	Maximize the host resources by which all the computationally / time heavy system procedures which do not require user participation should be done in background allowing parallel usage of the host computer for light / routine tasks (internet, email, word editing etc.)
<b>Measurable:</b>	Availability of the resources of the host machine while performing a heavy system task can be tested experimentally, and measured by number of routine tasks user can perform per unit time.
<b>Achievable:</b>	The proposed system will not interfere with standard OS task-switching policies. Non-busy waits might be introduced to release host resources for other tasks.
<b>Relevant:</b>	OCD 4.4 LS-3: Resource Availability OCD 3.2 OG-1, OG-3
<b>Specific:</b>	It conforms to OCD 4.4 LG-2 Usability providing better usability of host resources.
<b>Desired level:</b>	User can proceed with routine tasks with less than 50% of performance degradation
<b>Accepted level:</b>	User can accomplish main routine tasks with less than 80% of performance degradation.
<b>Priority:</b>	High

**Table 34 -4: Performance – Organizing data meaningfully for users**

<b>L.O.S. Requirement:</b>	LR-4: Performance – Organizing data meaningfully for users.
<b>Description:</b>	Data should be organized in a way that is meaningful to both archive staff and researchers.
<b>Measurable:</b>	Archive staff will provide usability feedback Data successfully apply to research publications
<b>Achievable:</b>	Export file in the format that is conventional among digital library related people.
<b>Relevant:</b>	SR-5: Displaying Node information IR-3: Object family navigation interface OCD 4.4 LG-4: Performance OCD 3.2 OG-1, OG-2, OG-3

<b>Specific:</b>	Showing how the object id of the selected node is related with other items, and what kinds of objects are around the node. (SR-5, IR-3)
<b>Priority:</b>	Medium

**Table 35 LR-5: Performance- data of current scale**

<b>L.O.S. Requirement:</b>	LR-5: Performance - data of current scale
<b>Description:</b>	The system should be able to produce analysis reports and visualize them for input data of size and complexity as for the Digital Archive collection in the current state.
<b>Measurable:</b>	The system should be able to generate, cluster and visualize for graphs with desired number of nodes.
<b>Achievable:</b>	H3viewer visualization framework is claimed to be able to visualize graphs of up to 300000 nodes. The total number of the objects the Digital Archive is dealing with is 120,000. This fact and possibilities for increasing the number of nodes will have to be researched.
<b>Relevant:</b>	SR-1: Relationship generation SR-2: Generate collection structure tree SR-3: Visualization OCD 4.4 LG-3: Performance OCD 3.2 OG-1
<b>Specific:</b>	The number of relationship generated (SR-1), the size of structure tree (SR-2), and the scale of visualization are all dependent on the capability of the three components combined – log analysis, relation clustering, and visualization. Current scale of data set of digital archive is 120,000, and the system should support at least 120,000
<b>Desired level:</b>	The system should be able to handle graphs with up to 150,000 nodes which is 30,000 more than 120,000.
<b>Accepted level:</b>	System is able to handle most of the analysis report generation requests submitted by the user. (constrained analysis period might reduce number of nodes in the graph and its complexity)
<b>Priority:</b>	Medium

## 7. Evolution Requirements

### 7.1 Capability Evolution Requirements

Table 36 ER-1: Advanced analysis

<b>Interface requirement:</b>	ER-1: Advanced analysis
<b>Description:</b>	Based on successful implementation of basic analysis functionality additional types of analysis will be implemented: - extracting of user relationships based on similarities in their retrieval patterns - determining the most 'important' / influential users of the collection
<b>Relevant:</b>	Realizes: Win condition 5.2
<b>Priority:</b>	Medium

### 7.2 Interface Evolution Requirements

Table 37 ER-2: Interface with other usage log formats

<b>Interface requirement:</b>	ER-2: Interface with other usage log formats
<b>Description:</b>	The system should be designed in modular fashion so that it would be easy to write additional module that would parse usage log files of different formats but containing all necessary information
<b>Relevant:</b>	Provides for system extensibility, and future reuse
<b>Priority:</b>	Medium

### 7.3 Technology Evolution Requirements

The standards and interfaces that the system is relying on are relatively static (Apache server log format, SQL) and even if changed require minimal tailoring.

### 7.4 Environment and Workload Evolution Requirements

Table 38 ER-3: Accommodating collection growth

<b>Interface requirement:</b>	ER-3: Accommodating collection growth
<b>Description:</b>	The system should accommodate collection growth that is most probable in near future. With current Digital Archive collection of about 300000 system should be able to function properly with collection increase of 10%.
<b>Relevant:</b>	Collection is predicted to grow.
<b>Priority:</b>	Medium

## 7.5 Level of Service Evolution Requirements

Table 39 ER-4: Workload growth should affect system performance in linear proportion

<b>Interface requirement:</b>	ER-4: Workload growth should affect system performance in linear proportion
<b>Description:</b>	Projected workload growth of 10% should deteriorate system performance in linear fashion – that is by 10% * some scalar. System should not exhibit exponential or worse performance deterioration, since that will make it less usable.
<b>Relevant:</b>	Realizes Win Condition 3.3, LG-4, LG-5
<b>Priority:</b>	Medium

## 8. Common Definition Language for Requirements

### **Digital Archive (DA)**

Collection of electronic and multi-media documents, such as images, photos, audios, and videos

### **SQLite**

a small C library that implements a self-contained, embeddable, zero-configuration SQL database engine.

### **GNU**

An on-going project of the Free Software Foundation to create a complete, freely distributable, POSIX-compliant computing environment.

### **Operating System (OS)**

Software that shares a computer system's resources (processor, memory, disk space, network bandwidth, and so on) between users and the application programs they run.

### **Application Program Interface (API)**

the interface (calling conventions) application programs use for accessing services provided by some lower-level module (such as the operating system or JVM)

### **Architecture Review Board (ARB)**

developer or property owner controlled committees responsible for the interpretation and application of property covenants to proposed construction and remodeling projects in a given community

### **Apache**

An open source Web server originally formed by taking all the "patches" (fixes) to the NCSA Web server and making a new server out of it