Feasibility Rationale Description (FRD)

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

Version History

Date	Author	Version	Changes made
10/10/04	Hui-Hsien Chi	v1.0	• Initial draft
10/18/04	Hui-Hsien Chi	v1.1	• Correct some grammar mistakes • Change section 2.1.4 and 2.1.5 • Add one more risk in section 4
10/24/04	Hui-Hsien Chi	v2.0	• Change section 2.1.5 and 2.2.3 • Modified and clarified the risk in section 4 • Add section 2.2.5 and 2.2.6
11/14/04	Hui-Hsien Chi	v3.0	• Change section 2.2 • Change section 3.1 • Section 4, add contingency plan for risks • Change section 5
11/29/04	Hui-Hsien Chi	V3.1	• Modify section 2.1 • Modify section 4
12/05/04	Hui-Hsien Chi	V3.2	• Add COCOMO to estimate requirement for section 3.3 • Add COTS table on section 5.2 • Modify section 2.1.5 • Modify section 2.2 to match OCD, SSAD, and SSRD
1/24/05	Genesan Kim	V4.0	• Modified section 2.3 • Added stakeholder concurrence table in section 2.3
1/25/05	Genesan Kim	V4.1	• Modified section 3.2
1/30/05	Genesan Kim	V4.2	• Modified section 4 to update current risks
2/2/05	Genesan Kim	V4.3	• Modified section 2.1, added new ROI graph • COCOMO result, Section 6 Figure 2 updated
2/14/05	Genesan Kim	V4.4	• Updated section 2.2.1 to match OCD, SSRD, SSAD • Updated section 2.2.3 to match OCD, SSRD, SSAD • Updated section 2.2.4 to match OCD, SSRD, SSAD • Updated section 2.2.5
2/15/05	Genesan Kim	V4.5	• Updated section 3.1 to match OCD, SSRD • Updated section 3.3 to match OCD, SSRD
4/01/05	Genesan Kim	V4.6	• Updated section 2
4/22/05	Genesan Kim	V4.9	• Updated all sections to account for requirements change
4/23	Genesan Kim	V5.0	• Updated transition costs, information from transition plan update

Table of Contents

Table of Tables

Table of Figures

1. Introduction

1.1 Purpose of the Feasibility Rationale Description Document

The purpose of this document is to rationalize development decision of the Data Mining of Digital Library Usage Data application. It shall ensure that all the system elements have been demonstrated satisfying the success-critical stakeholders in the OCD and LCP, meeting the requirements described in the SSRD, and using the specified architecture defined in the SSAD.

The FRD represents that the life cycle process will:
-Satisfy the requirements
-Support the operational concept
-Match the budgets and schedules

The target audiences of the FRD are stakeholders described in OCD, and the people in the Architecture Review Board (ARB).

1.2 References

MBASE Guidelines version 2.4.1
http://cse.usc.edu/classes/cs577a_2004/guidelines/MBASE_Guidelines_v2.4.1.pdf

MBASE Version 2.4.1 Templates for FRD (version 1a)
http://cse.usc.edu/classes/cs577a_2004/guidelines/MBASEtemplates/FRD_Templatev1a.doc

"Data Mining of Digital Library Usage Data" Project Description
http://sunset.usc.edu/classes/cs577a_2004/projects/description/project7.htm

MBASE Electronic Process Guide
http://cse.usc.edu/research/MBASE/EPG

Fall 2004 CS 577a Project #7 LCO portion of OCD
http://seacliff.usc.edu/~team7/LCO/OCD_LCD_F04a_T07_V02.0.doc

Fall 2004 CS 577a Project #7 Prototype for LCO
http://seacliff.usc.edu/~team7/LCO/Prototypes/Prorotype_LCO_F04a_T07_V01.0.doc

Fall 2004 CS 577a Project #7 Prototype for SSRD
http://seacliff.usc.edu/~team7/LCO/SSRD_LCO_F04a_T07.doc

Fall 2004 CS 577a Project #7 Prototype for SSAD
http://seacliff.usc.edu/~team7/LCO/SSAD_LCO_F04a_T07.doc

Fall 2004 CS 577a Project #7 Prototype for LCP
http://seacliff.usc.edu/~team7/LCO/LCP_LCO_F04a_T07.doc

Fall 2003 CS 577a Project #3 LCO portion of FRD
http://ebase.usc.edu/eservices/cs577a_2003/team03a/LCO/FRD_LCO_F03a_T03.pdf

Fall 2003 CS 577a Project #8 LCO portion of FRD
http://ebase.usc.edu/eservices/cs577a_2003/team08a/LCO/FRD_LCO_F03a_T08.pdf

Fall 2003 CS 577a Project #11 LCO portion of FRD
http://ebase.usc.edu/eservices/cs577a_2003/team11a/LCO/FRD_LCO_F03a_T11.pdf

Fall 2003 CS 577a Project #17 LCO portion of FRD
http://ebase.usc.edu/eservices/cs577a_2003/team17a/LCO/FRD_LCO_F03a_T17.pdf

Fall 2003 CS 577a Project #32 LCO portion of FRD
http://ebase.usc.edu/eservices/cs577a_2003/team32a/LCO/FRD_LCO_F03a_T32.pdf

USC Information Services Division
http://www.usc.edu/isd/about/about.html

Dr. Bollen, Johan presentations
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

1.3 Change Summary

-Clarified the risks base on the ARB comments.
-Gave more specific contents of project benefit
-Changed the document to approach the change of OCD, SSRD, and SSAD.
-Changed the document to consist the priorities, process and resources.
-Modified the product features.
-Added contingency plan for risks
-Used COCOMO to estimate the high capability requirements

RLCA Changes

-Updated and fixed section 2.3 Stakeholder Concurrence

-Added stakeholder concurrence table 9
-Modified section 3.2 Process Match to System Characteristics and Priorities
-Updated previous risks and added new risks
-Updated business case analysis and generated new ROI graph
-Previous version did not include development cost
-Updated maintenance cost with respect to client feedback
-Fixed ROI graph based on RCLA ARB feedback

-Updated all of section 2.2 in terms of updating certain changes from other artifacts (mostly changes in level of service requirements and goals). All requirements, capabilities, interface requirements, and level of services mappings were all updated to reflect the changes through all documents, and also all incorrect mappings were identified and fixed.

-Section 2.2.1 Operational Concept Satisfaction, Table 3: OCD satisfaction – minor concepts, the level of service goals in this table were updated to map to the changed level of service requirements in the SSRD.

- Section 2.2.1 Operational Concept Satisfaction, Table 3: incorrect mappings to SSRD 2.1.1 were fixed to SSRD 3.2.1.

-Section 2.2.3 Capability Requirements Satisfaction, Table 5: incorrect mappings to SSAD for SR-2, SR-6 and SR-10 were fixed and updated due to changes in SSAD.

-Section 2.2.3 Capability Requirements Satisfaction, Table 5: all references to apache log files were updated due to the change in requirements of not using apache log files. SR-11 was also added due to the same addition in the SSRD.

-Section 2.2.4, Interface Requirements Satisfaction, IR-6 deleted from Table 6. This is due to change in the interface requirements as well as the architecture description of the system (Specifically the decision of not using the apache log data).

-Section 2.2.5 Level of Service Requirements Satisfaction
This was due to the change in level of service requirements in the SSRD.

-Section 3.1 of the system priorities and section 3.3 were also updated to update or fix the mappings between the requirements and satisfactions

-Section 3.1 Table 10; incorrect mapping to SSRD 3.1 SR-2 was changed to SSRD 3.2 SR-3 to match OCD 3.2 OG-1

-Section 3.3, this section was updated with respect to the requirements change of using apache log files. Since the usage data and log files will be provided by the client, all reference to using apache server log files was updated. Also incorrect mappings to SSRD 2.2 were changed to SSRD 3.2 when referencing the SR, system requirements.

-Updated Product Rationale

-Update all requirement mappings to be consistent with SSRD, SSAD, and OCD

-Updated Transition Cost due to change in training costs

2. Product Rationale

This section describes how the product develops and matches its specific architecture which will satisfy the stakeholders and reach the requirements. It also provides the cost estimate of the product.

2.1 Business Case Analysis

This section represents the analysis and estimation of the business case impact of monetary and non- monetary.

2.1.1 Development Costs

The development costs include hardware, software, personnel, and facilities costs. Since the project will be implemented by the students who are taking CSCI577a/b course, some costs are presented in terms of hours of effort.

Personnel Costs:

The total effort of the team:

The number of team members for Fall 2004(S): 6

The number of team members for Spring 2005(S2):4

Average time spent on the project per week (t): 20 hours

Total time available (T): 24 weeks

The total effort: S*t*(T/2) + S2*t*(T/2) = (6*20*12) + (4*20*12) = 2400 hours

Average pay per hour = $0: D*evelopment is done for credit*

Team personnel cost = $0

The total effort of the client:

The number of client(S): 1 (Jewel)

Average time spent on the project per week (t): 7.5 hours (*Client participates in development through weekly meetings with development team, evaluating prototypes, providing data, etc.)*

Total time available (T): 24 weeks

The total effort: S*t*T = 1*7.5*24 = 180 hours

Average client salary = $35

Total client personnel cost = 180 * $35 = $6300

Total personnel costs = $6300

The detail information of the personnel time cost is described in the [LCP 5.2].

In addition, this system is going to be developed using the existing software and hardware that belong to USC ISD/USC CES. Thus, the cost of software and hardware should not be an issue of this project. This relates to our win condition of the WinWin

Negotiation building the project with no budget, coding the system in-house, and using open-source/ freeware.

Hardware costs:
Development team will use personal workstations and computing facilities provided by the USC Information Services Division. The proposed application is standalone and therefore does not require hosting facilities to be tested.
Total hardware costs = $0

Software costs:
Project will use the following open source components which are available at no cost: SQLite, H3Viewer, Open Motif. The following free open source development tools will be used: gcc, make, autoconf etc. Project management will be done using software tools licenses for which are provided with no cost as part of the class: Rational Rose, Microsoft Project.
Total software costs = $0

Facilities:
Development team will use USC campus facilities for team meetings, and other development activities.
Total facilities cost = $0;

Total development costs = Personnel costs + Hardware costs + Software costs + Facilities costs = $6300

2.1.2 Transition Cost

The proposed system will be a brand new system developed for analyzing usage log and generating reports of usage log analysis. Thus, training time is one of the transition cost to be incurred in this project. The system will also use open-source, so there is no transition cost of COTS licenses. The system will not purchase any equipment because it is going to use existing equipment in USC library. The following is a breakdown of all costs that are a part of the transition process.

Training Costs:

- Trainer preparation time = 4 hours. Since training preparation will be done by the development team, cost = $0.
- Client will be trained as an administrator of the system, and will be given instructions on how to analyze log data using the system and distribute results of the analysis to regular users. Client training time = 4 hours; Average client salary = $35 per hour. Therefore cost = 4 * $35 = $140.
- Library staff users will be trained on how to visualize analyzed data received from the admin using the system. Number of users = 5; Training time = 4 hours; Average salary = $25 per hour. Thus, cost = 5 * 4 * 25 = $500.
- Maintainer training time = 4 hours; Average maintainer salary = $30 per hour. Cost = 4 * $30 = $120.

Total training costs = $140 + $500 + $120 = $760

Data preparation:
The system will use the log data provided by existing Digital Archive system. This data is provided as part of normal functioning of the DA system and therefore does not incur any costs.
Data preparation costs = $0;

COTS licenses:
Proposed system uses only open source and free components.
COTS licenses costs = $0

Operational readiness testing:
This testing will be done by the client through comprehensive evaluation of major functionality of the system. Estimated testing time = 2 hours; Average client's salary = $35. Cost = 2 * 35 = $70.
Total testing costs = $70

Site preparation:
The proposed system is a standalone application which can be installed on any workstation running Linux, which does not require any special preparations
Site preparation costs = $0

Facilities preparation:
Proposed system does not require use of any specialized facilities.
Facility preparation costs = $0

Equipment purchase:
The developed system will be installed on existing Digital Library equipment and also on personal workstation of the client; therefore no equipment will be purchased.
Equipment cost: = $0

Total transition cost = $760 + $70 = $830

2.1.3 Operational Cost

Since the proposed system is a stand-along application, there is no operational cost for the system. Moreover, there is no license cost for COTS software because the system uses the software provided by USC ISD and freeware or open-source.

2.1.4 Maintenance Cost

The system will be maintained by existing Digital Library staff. The proposed system will have one maintainer that is required to spend 5 hours per week on maintenance with a salary of $30 per hour.

Maintenance Cost Estimate:
Number of required maintainers = 1
Estimated maintenance time = 5 hours per week
Maintenance cost = 1 * 5 * 48 (weeks) * $30 = $7200 per year

2.1.5 Estimate of Value Added and Return on Investment

This section is to estimate the value added by the proposed system, and return on investment based on the input value. Since USC library is a non-profit organization, and currently, there is no plan of selling the proposed system. Therefore, the system will not generate any revenue, but the benefits of the project can be realized.

Estimate of Value Added

According to the result chain in OCD 2.1, we can recognize some value added of the proposed system. One of the many things the system will do is to analyze item characteristics, which will provide digital archive staffs better understanding of archive collection structure. Secondly, the system will analyze relationships between items and provides visualization relationships between them. This provides digital archive staffs better understanding of usage trends over time. With better understanding of usage trends combined with archive collection structure, it allows DA staffs to make better decision in improving digital archive collection and services.

The analysis result will also be available for researcher by request. Researchers can make use of the result by analyzing it and applying it to their own research.

The system is intended to give DA staffs a better mean to simplify organization procedure in improving the DA library. It will provide retraining procedures to ease difficulty of DA staffs in adjusting to the new system.

Estimation of ROI
The proposed system will be used by the client – Digital Archive project administrator and 5 Library employees, and it will make updating of Library Collections more efficient. Currently there is no automated system that helps Library employees to make decisions on which collections are important, which new areas of interest are emerging and which items could be removed from the collection, since they are not being used.

Currently 5 Library employees spend 5 hours per week for maintaining and developing digital collections. Also the Digital Archive manager spends 10 hours per month of her time for manual usage log analysis and coordination of development efforts of Library staff. Average salary of Library staff is $25 per hour. Average salary of Library project manager is $35 per hour. Thus current costs per year are = 5 * 5 * 48 * 25 + 10 * 12 * 35 = $34200 per year.

The proposed system will make this process more efficient in the following way: each of 5 library employees now will have to spend only 3 hours per week evaluating results of usage data analysis and making decisions on collection updates. Digital Archive manager will use the developed system to produce usage analysis data and will spend 5 hours per month doing that.

Thus improved cost will be = 5 * 3 * 48 * $25 + 5 * 12 * $35 = $20100 per year.
Thus savings = $14100 per year.
Total cost of the system = Development cost + Transition cost = $6300 + $830 = $7130.
Maintenance cost = $7200.

Table 1: < Return on Investment >

	1st year	2nd year	3rd year	4th year	5th year
Costs saved	$0	$14100	$14100	$14100	$14100
Cumulative costs saved	$0	$14100	$28200	$42300	$56400
Cost of the system	$7130	$7200	$7,200	$7,200	$7,200
Cumulative costs of the system	$7130	$14330	$21530	$28730	$35930
ROI	-1	-0.016	0.31	0.47	0.57

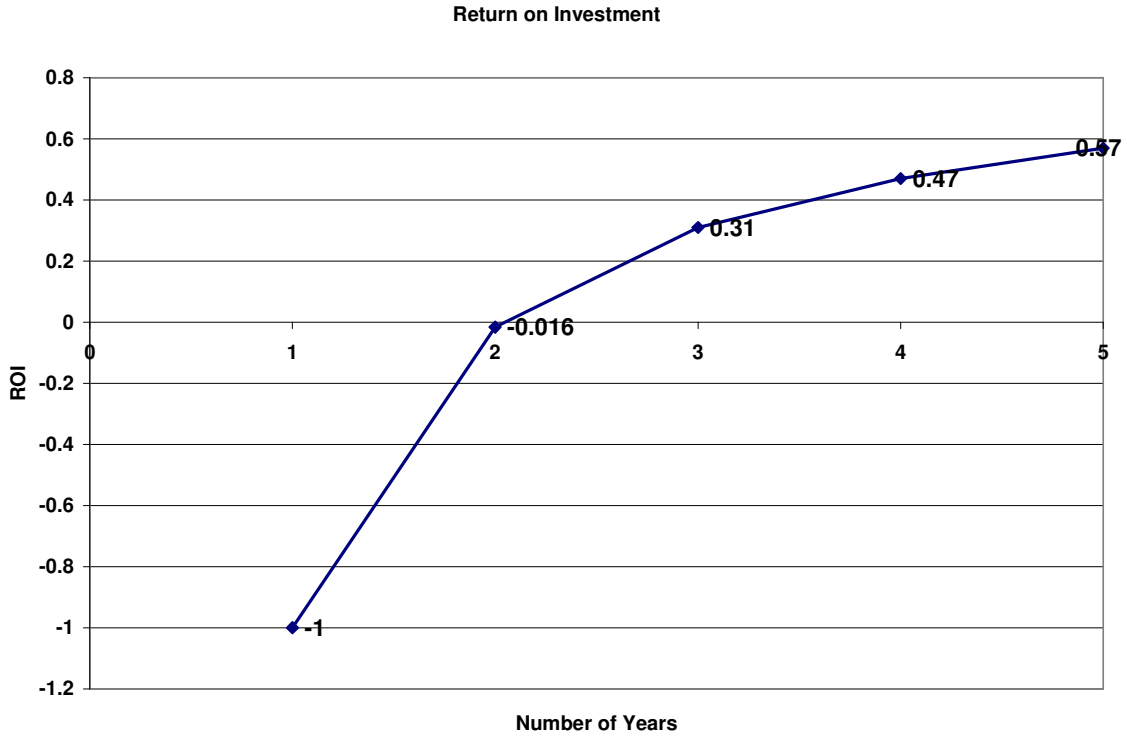


Figure 1 – Return on Investment

2.2 Requirements Satisfaction

This section explains how and why the proposed system will meet the specific architecture of SSAD and satisfy the system requirements of OCD and SSRD.

2.2.1 Operational Concept Satisfaction

This section shows how the project satisfies the operational concept which is listed in the [OCD 4]. Below is the list of the concepts from the [OCD 4]

Table 2: < Operational Concept Satisfaction – System Capabilities>

Operational Concept	C-01: Log Analysis
Description	The proposed system shall use the algorithm which proposed by Dr. Bollen to generate the relationships among item. The [SSRD 2.1.1 SR-3] defined the requirement of this system capability. Therefore, it completely satisfies the [OCD 4.3.1 C-01]
Reference	OCD 4.3.1 C-01 SSRD 3.2.1 SR-1

Operational Concept	C-02: Relation Clustering
Description	Cluster the relationship between objects to generate a relationship graph with levels.. The [SSRD 2.1.1 SR-1] defined the requirement of this system capability. Therefore, it completely satisfies the [OCD 4.3.2 C-02]
Reference	OCD 4.3.2 C-02 and SSRD 3.2.1 SR-2

Operational Concept	C-03: Visualization log analysis results
Description	The proposed system shall provide a visualization interface to display the relationships among the items. The [SSRD 3.2.1 SR-5] defined the requirement of this system capability. Therefore, it completely satisfies the [OCD 4.3.3 C-03]
Reference	OCD 4.3.3 C-03 and SSRD 3.2.1 SR-3

Table 3: < Operational Concept Satisfaction – minor concepts>

OCD Project Goal, Constraints, and Levels of Service	SSRD Reference
OCD 4.2.3 PC-1: Fixed Schedule	SSRD 1.4 PR-1: Limited schedule
OCD 4.2.4 PC-2: Limited Resources	SSRD 1.4 PR-1: No budget
OCD 4.2.1 PG-1: Better Understandings of Archive Collection Structure	SSRD 3.2.1 SR-4: the proposed system

	should be able to generate collection structure tree.
OCD 4.2.2 PG-2: Better Understandings of Archive Usage Trends	SSRD 3.2.1 SR-3: the system provides the relationship among the items.
OCD 4.4 LS-1 System dependability	SSRD 5 LR-1: Dependability
OCD 4.4 LS-2 Usability - Provide user friendly interface	SSRD 5 LR-2: Usability
OCD 4.4 LS-3 Performance – Data organized in meaningful way	SSRD 5 LR-4: Performance

2.2.2 Project Requirements Satisfaction

The project requirements were retrieved from the Winwin negotiation and the agreements with the client. All the requirements are documented in the [SSRD 2]. The [LCP 4] describes the plan of the project, and how it matches the questions in the [SSRD 2]. Below is a list of the project requirements from the [SSRD 2].

Table 4: < Project Requirements Satisfaction>

Project Requirements	PR-1: Limited schedule
Requirements Satisfaction	Since this project is a part of cs577a course work, there is a limited schedule. The project needs to be implemented in 24 weeks, and it is distributed into four stages of development (inception, elaboration, construction and transition). The [LCP 2] has complete milestone and schedule of the system developing. Also, the [LCP 4] describes the plans of the project. It shows how the project approaches those milestone and schedule. Therefore, these plans and schedule demonstrate the project can be implemented in 24 weeks. The [LCP 2] and [LCP 4.1] completely satisfies the [SSRD 2.1 PR-1]
Reference	SSRD 2.1 PR-1, LCP 2 and LCP 4

Project Requirements	PR-2: No budget
Requirements Satisfaction	From the Winwin negotiation, one of the win conditions is no budget for this project, and it is also defined in the [SSRD 2.1 PR-2]. In the [LCP 5.2] and the [FRD 2.1.1], we described why there is no budget in this project. This is due to several reasons, such as no developer cost, no hardware cost, and no software cost.
Reference	SSRD 2.1 PR-2, LCP 5.2, and FRD 2.1.1

Project Requirements	PR-3:Use MS Project for scheduling and progress tracking
Requirements	The [LCP 4.2] described the system will use MS project for

Satisfaction	tracking and planning the project. This completely satisfies the [SSRD 2 PR-3]
Reference	SSRD 2.1 PR-3 and LCP 4.2

Project Requirements	PR-4: Use GNU compilation tools
Requirements Satisfaction	The [LCP 4.2] production stage, described the system will use GNU for developing the system. This completely satisfies the [SSRD 2.2 PR-4]
Reference	SSRD 2.2 PR-3, and LCP 4.2

Project Requirements	PR-5: Use efficient development language with 3d visualization capabilities
Requirements Satisfaction	In the [LCP 4.2] production stage, it has defined the developed team use C/C++ for coding user interface. This matches the [SSRD 2.2 PR-5].
Reference	SSRD 2.2 PR-5, and LCP 4.2

Project Requirements	PR-6: Use UML for modeling and design
Requirements Satisfaction	The [LCP 4.2] has described the develop team uses Rational Rose for building UML diagram. That shows the team will use UML for modeling and design of the project. This completely satisfies the [SSRD 2.2 PR-6].
Reference	SSRD 2.2 PR-6 and LCP 4.2

Project Requirements	PR-7: Use free components
Requirements Satisfaction	The [FRD 5.2] has listed all the open sources that will be used in this system. It satisfies this requirement.
Reference	SSRD 2.2 PR-7, and FRD 5.2

Project Requirements	PR-8: Target operating system is UNIX
Requirements Satisfaction	The proposed system shall be able to use on UNIX. The [LCP 2.2.2] has scheduled the transition preparation which defines that the system will be compiled for UNIX
Reference	SSRD 2.2 PR-8, and LCP 2.2.2

Project Requirements	PR-9: Use visualization component that provides interactive 3d interface.
Requirements Satisfaction	The result of the usage analysis should be present as 3D hyperbolic view. In the [LCP 4.2], the production stage has described that the develop team will use an open source, H3viewer, to implement the proposed system. Therefore, this plan matches [SSRD 2.2 PR-6]
Reference	SSRD 2.2 PR-9, and LCP 4.2

Project Requirements	PR-10: Use data storage that does not require maintenance.

Requirements Satisfaction	The [LCP 4.2] production stage, described the system will use SQLite for storing the data. Also, the [FRD 5.2] has the reason of why we chose this open source for storing the data. This completely satisfies the [SSRD 2.2 PR-10]
Reference	SSRD 2.2 PR-10 and LCP 3

Project Requirements	PR-11: Input data format compatibility
Requirements Satisfaction	The proposed system should be able to parse the usage data from log file. The [SSAD 3.1.7.1] has defined that the proposed system will provide an interface to import log data, and it can be able to check and parse the data.
Reference	SSRD 2.2 PR-11, and SSAD 3.1.7.1

Project Requirements	PR-12: Well documented code
Requirements Satisfaction	There are several peer reviews during the production stage. These reviews have been planned in the [LCP 4.2.2], and the purpose of these reviews are to ensure that the system code is documented and can be approved by the client. These review plans in the [LCP 4.2.2] completely satisfy the [SSRD 2.2 PR-12]
Reference	SSRD 2.2 PR-12 and LCP 4.1.2

Project Requirements	PR-13: Distribution consists of source code and binaries for UNIX
Requirements Satisfaction	The [LCP 2.2.2] has scheduled the transition preparation which defines that the system's source code will be compiled for LUNIX
Reference	SSRD 2.3 PR-13, and LCP 2.2.2

Project Requirements	PR-14: Maintainer training requirements
Requirements Satisfaction	Since there are some evolution requirements, the system needs a maintainer to maintain this system. The [LCP 3.1.3] has defined the responsibility for maintainer in the transition stage. Also, the [FRD 2.1.4] describes the possible cost of the maintainer.
Reference	SSRD 2.3 PR-13, and LCP 2.2.2

2.2.3 Capability Requirements Satisfaction

This section provides the high-risk or critical capability requirements. Moreover, we used COCOMO to evaluate the feasibility of these capability requirements and it is represent in the [LCP 6.3].

Table 5: <Capability Requirements Satisfaction>

Title	**SR-1: Relationship generation**
Description	Generate relationship between the digital items and provide a graph that utilizes the nodes and links to represent the relationship among the digital items. Also, this graph shall be stored in the proposed system database. The [SSAD 4.1.7.3.2.1.1]] has provided the detail design of the generate report component which will produce item relationship, and this section also provides the activity diagram of the generate report component. This completely satifies the [SSRD 3.2.1 SR-1].
Criticality	High
Technical Issue	We need to create an algorithms to generate the relationship among the digital items. The algorthm will be based on Dr. Bollen's research.
Cost and Schedule	There is no cost and the schedule is on the production stage.
Dependencies	Depends on Dr. Bollen's algorthm
Side Effects	There may be a lot of large graphs in the database.
Risks	Since the algorthm is a new algorthm that was produced by Dr. Bollen, there are few resources that can be reference. Also, the proposed system database may not be able to handle these large numbers of the relationship graph.
Reference	SSRD 3.2.1 SR-1, SSAD 4.1.7.3.2.1.1

Title	**SR-2: Generate collection structure tree**
Description	Based on the detailed relationship graph a hierarchy of small sparse graphs / trees should be generated through grouping of nodes into sub-graphs at multiple levels based on their relatedness. The [SSAD 4.1.7.3.2.1.1]] has provided the detail design of the generate report component which will produce graph clustering, and store result in a tree structure, and this section also provides the activity diagram of the generate report component. This design and implementation completely satisfy the [SSRD 3.2.1 SR-2].
Criticality	Medium
Technical Issue	Used existing multiple graph clustering/normalization algorithm.
Cost and Schedule	There is no cost and the schedule is on the production stage.
Dependencies	Depended on a existing multiple graph clustering algorithm.
Side Effects	N/A
Risks	Since the DA data growth, the scalability of the graph clustering algorithm may not be able to handle the large amount of data.
Reference	SSRD 3.2.1 SR-2, and SSAD 4.1.7.3.2.1.1

Title	**SR-3: Visualization**
Description	The proposed system shall provide a 3D hyperbolic view of the item relationship graph. . The [SSAD 4.1.7.4.2.1] has provided the detail design of the visualization of analysis result module. That module is produced for displaying the analysis result in a 3D hyperbolic view. This design completely satisfies the [SSRD 3.2.1 SR-3].
Criticality	high
Technical Issue	Displayed 3D hyperbolic view
Cost and Schedule	Since we will use an open source to implement this system, there is no cost. The schedule is on production stage.
Dependencies	This will depend on H3viewer which is an open source that the client and the developed team had agreed to use.
Side Effects	Developers needs to study this open source.
Risks	It may have performance problems because it loads large graphs from a database, and then displays them on this 3D viewer
Reference	SSRD 3.2.1 SR-3 and SSAD 4.1.7.4.2.1

Title	**SR-4: Search**
Description	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. The [SSAD 4.1.7.4.2.1] has provided the detail design of the search tool. This design completely satisfies the [SSRD 3.2.1 SR-4].
Criticality	high
Technical Issue	Node traversal of tree structure
Cost and Schedule	Since we will use an open source to implement this system, there is no cost. The schedule is on production stage.
Dependencies	This will depend on H3viewer which is an open source that the client and the developed team had agreed to use.
Side Effects	Developers needs to study this open source.
Risks	It may have performance problems because number of nodes to be traversed can be a complexity risk
Reference	SSRD 3.2.1 SR-4 and SSAD 4.1.7.4.2.1

Title	**SR-5: Displaying Node Information**
Description	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. The [SSAD 4.1.7.3.2.1.1]] has provided the detail design of the the diplay of this node information. Also, this section provides the activity diagram of the search component. This design completely satisfies the [SSRD 3.2.1 SR-5].
Criticality	High
Technical Issue	Complexity of node traversal and displaying parent, siblings, and children.
Cost and Schedule	There is no cost and the schedule is on the production stage.
Dependencies	Tree collection structure.
Side Effects	Speed of displaying information.
Risks	Usability risks in terms of GUI design and addition of side panels.
Reference	SSRD 3.2.1 SR-5, SSAD 4.1.7.3.2.1.1

Title	**SR-6: Open the object webpage**
Description	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 opens the webpage for the object id, as shown in the [SSAD 4.1.7.1.2.1.2] activity diagram. This activity will solve and satisfy this requirement.
Criticality	High
Technical Issue	Serve status of the digital archive
Cost and Schedule	There is no cost and the schedule is on production stage.
Dependencies	That the object names will be consisted from data log files to the artifacts located on the server, including the url location.
Side Effects	Side effects include assumptions of consistent url.
Risks	If any of the consistencies listed above are changed, then the function will fail.
Reference	SSRD 3.2.2 SR-6, SSAD 4.1.7.1.2.1.2

Title	**SR-7: Omit mal-formatted retrieval records**
Description	If in provided usage log file a retrieval record is met that does not comply with log format, it should be skipped.. There is an activity, "check validity of data", in the [SSAD 4.1.7.1.2.1.2] activity diagram. This activity will solve and satisfy this requirement.

Criticality	High
Technical Issue	Usage data log format is specified in corresponding documentation.
Cost and Schedule	There is no cost and the schedule is on production stage.
Dependencies	Since this will develop by hand code, there is no dependency.
Side Effects	There is no side effect on this architecture design.
Risks	If log file format change is changed, this function also needs to be changed.
Reference	SSRD 3.2.2 SR-7, SSAD 4.1.7.1.2.1.2

Title	**SR-8: Omit usage log file with 0 valid records**
Description	If the provided usage log file contains no valid retrieval records, the system should stop the import action and should produce empty output files. There is an activity, “check validity of data”, in the [SSAD 4.1.7.1.2.1.2] activity diagram. This activity will solve and satisfy this requirement.
Criticality	Medium
Technical Issue	Log format is specified in corresponding documentation.
Cost and Schedule	There is no cost and the schedule is on production stage.
Dependencies	Since this will develop by hand code, there is no dependency.
Side Effects	There is no side effect on this architecture design.
Risks	There is no risk on this architecture design.
Reference	SSRD 3.2.2 SR-8, SSAD 4.1.7.1.2.1.2

Title	**SR-9: Do not visualize analysis report of unknown format**
Description	If the result is an unknown format, the proposed should not show the result and stop the visualization action. The [SSAD 4.1.7.4.2.1.1] has designed the activity of generate report and it also refers this requirement.
Criticality	Medium
Technical Issue	Format of analysis results could be explicitly defined in terms of field data types, data ranges and regular expressions.
Cost and Schedule	There is no cost and the schedule is on production stage.
Dependencies	Since this will develop by hand code, there is no dependency.
Side Effects	There is no side effect on this architecture design.
Risks	There is no risk on this architecture design.
Reference	SSRD 3.2.2 SR-9 SSAD 4.1.7.4.2.1.1

2.2.4 Interface Requirements Satisfaction

Table 6: <Interface Requirements Satisfaction>

Interface requirement:	IR-1: Visualize analysis report
Requirements Satisfaction:	The [SSAD 4.1.7.4] has provided the detail design of the visualization of analysis result module. That module is produced for displaying the analysis result in a 3D hyperbolic view
Reference:	SSRD 5.1 IR-1 SSAD 4.1.7.4

Interface requirement:	IR-2: Search Interface
Requirements Satisfaction:	The [SSAD 4.1.7.1.2] shows the design of the search tool included for the visual tree collection structure. This design can satisfy this requirement.
Reference:	SSRD 5.1 IR-2, SSAD 4.1.7.1.2

Interface requirement:	IR-3: Object family navigation interface
Requirements Satisfaction:	The [SSAD 4.1.7.3.2] shows the design of presenting hierarchy information with respect to selected objects. Those designs can satisfy this requirement.
Reference:	SSRD 5.1 IR-3 SSAD 4.1.7.3.2

Interface requirement:	IR-4: Webpage view interface
Requirements Satisfaction:	The [SSAD 4.1.7.3.2] shows the system will provide a component that links an object to a url location with object information. These designs can satisfy this requirement.
Reference:	SSRD 5.1 IR-4 SSAD 4.1.7.3.2

2.2.5 Level of Service Requirements Satisfaction

Table 7: <Level of Service Requirements Satisfaction>

Title	**LR-1: System Dependability – Stable data import/export**
Criticality	High
Criteria of the Requirement Satisfaction	The system should not crash while importing and exporting data.
Satisfaction	The [SSAD 4.3] shows this requirement applys to SCO-01, SCO -02, SCO -03, SCO -04, and it provides the projected value. Also, the [SSAD 4.1.7.1.3], [SSAD 4.1.7.2.2], [SSAD 4.1.7.3.3], and [SSAD 4.1.7.4.3] have detail description of how the system will satify this requirment
Reference	SSRD 5 LR-1, SSAD 4.3, 4.1.1.3, 4.1.7.2.2, 4.1.7.3.3, and 4.1.7.4.3

Title	**LR-2: Usability – User-friendly interface for viewing item relationships and updating data**
Criticality	Medium
Criteria of the Requirement Satisfaction	The proposed system should be user-friendly and easy to use.
Satisfaction	The [SSAD 4.3] shows this requirement applys to SCO -04, and it provides the projected value. Also, the [SSAD 4.1.7.4.3] have detail description of how the system will satify this requirment.
Reference	SSRD 5 LR-2, SSAD 4.3, and 4.1.7.4.3

Title	**LR-3: Usability – Maximizing the usability of host resources**
Criticality	Medium
Criteria of the Requirement Satisfaction	User can perform routine tasks through the system with minor performance degradation.
Satisfaction	The [SSAD 4.3] shows this requirement applys to SCO -04, and it provides the projected value. Also, the [SSAD 4.1.7.3.3] have detail description of how the system will satify this requirment
Reference	SSRD 5 LR-3, SSAD 4.3, and 4.1.7.3.3

Title	**LR-4: Performance – Organizing data meaningfully for users**
Criticality	Medium
Criteria of the Requirement Satisfaction	Data should be organized in a way that is meaningful to both archive staff and researchers.
Satisfaction	The [SSAD 4.3] shows this requirement applys to SCO - 03, and it provides the projected value. Also, the [SSAD 4.1.7.3.3] have detail description of how the system will satify this requirment
Reference	SSRD 5 LR-4, SSAD 4.3, and 4.1.7.3.3

Title	**LR-5: Performance – data of current scale**
Criticality	Medium
Criteria of the Requirement Satisfaction	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. The system should be able to generate, cluster and visualize for graphs with desired number of nodes.
Satisfaction	The [SSAD 4.3] shows this requirement applys to SCO - 04, and it provides the projected value. Also, the [SSAD 4.1.7.4.3] have detail description of how the system will satify this requirment.
Reference	SSRD 5 LR-5 , SSAD 4.3, and 4.1.7.4.3

2.2.6 Evolution Requirements Satisfaction

Table 8: <Evolution Requirements Satisfaction>

Title	**ER-1: Advanced analysis**
Criticality	Medium
Criteria of the Requirement Satisfaction	Base on the basic analysis extending its functionality additional types for advanced analysis.
SSAD Reference	The [SSAD 4.1.7.3.1] has described that the log analysis module is designed in a way that allow the future evolution.
Reference	SSRD 6.1 ER-1 SSAD 4.1.7.3.1

Title	**ER-2: Interface with other usage log formats**
Criticality	Medium
Criteria of the Requirement Satisfaction	Follow the standard module fashion so that it can be easy to write an additional module to link with other usage log formats
SSAD Reference	The [SSAD 4.1.7.1.1] has described that the log analysis module is designed in a way that allow the future evolution.
Reference	SSRD 6.2 ER-2 SSAD 4.1.7.1.1

Title	**ER-3: Accommodating Collection growth**
Criticality	Medium
Criteria of the Requirement Satisfaction	Tested the proposed system by 10% increase of existing data numbers.
SSAD Reference	The [SSAD 4.1.7.4.1] has described that the log analysis module is designed in a way that allow the future evolution.
Reference	SSRD 6.5 ER-4 SSAD 4.1.7.4.1

Title	**ER-4: Workload growth should affect system performance in linear proportion**
Criticality	Medium
Criteria of the Requirement Satisfaction	Tested the proposed system by 10% increase of existing data numbers.
SSAD Reference	The [SSAD 4.1.7.4.1] has described that the log analysis module is designed in a way that allow the future evolution.
Reference	SSRD 6.5 ER-4 SSAD 4.1.7.4.1

2.3 Stakeholder Concurrence

All the stakeholders, including the student developers and client, attended scheduled Win-Win negotiations. The concurrence of the stakeholders involved was preserved through these Win-Win negotiations. During these negotiations, win conditions were created and identified by all stakeholders and agreements were reached through a process of stating issues and resolving them through certain options. The following table shows the stakeholder concurrence and how the team and client came to resolve issues that arose during the negotiations

Table 9: <Stakeholder Concurrence>

Issue	Resolution
Conflict between client's desire to integrate their personal visualization tool into the	Spend two weeks to research the tool, and client will provide relevant research papers,

proposed system, and lack of knowledge and research of the tool by development team.	guidance, and meetings with people in the area of study.
Potential conflict between the team wanting formalized and set requirements, and client stating that the requirements are susceptible to change over time during the semester due to open source integration and unfamiliarity with all components.	Research more related open source programs at an early stage, and consistently discuss requirements during client meetings to finalize and stabilize them. Perform enough research on all requirements in order to ensure a win-win environment.
Client stated 'No Crashes' as a win condition with respect to the end product, however team thought this was not a guarantee but rather dependent on the maintenance and transition success.	Provide a recover plan incase of system crash and provide all support and maintenance information. Perform all appropriate research on design and development to provide a system that is measurably stable.
Conflict between team's schedule and semester length constraint, and client's desire to add more evolution requirements along the way that were not clearly agreed upon or stated in the negotiations.	Have frequent team meetings with client and discuss these possible evolution requirements. Properly analyze and agree upon further options and the feasibility of the additions.

.

The agreements above and additional agreed upon win conditions can be found in the Win-Win Negotiation Report at the following URL:

http://seacliff.usc.edu/~team7/LCO/EWW_LCO_F04a_T07_V02.0.pdf

3. Process Rationale

This section analyzes how the development will satisfy the stakeholders' cost and schedule base on Life Cycle Plan and System and Software Requirements Description.

3.1 System Priorities

The system priorities were discussed in the Winwin negotiations. However, some of the priorities have been changed in the client meeting after the Winwin negotiation. One main goal of the organization is to estimate the usage of Digital Archive. After evaluating the DA usage log, the application shall produce a graphical user interface of the result, which is defined in the SSRD. The second main goal of the project is to implement the system with a limited schedule. The developers of this project are graduate students in cs577, so there are only 24 weeks to complete the whole project.

The system priorities are:

Table 10: < System Priorities >

High Priorities:

OCD Goal	SSRD
Evaluate digital library usage. [OCD 3.2 OG-1]	Relationship generation [SSRD3.2 SR-3]
Limited Schedule according to. [OCD 4.2 PG-1]	Limited schedule [SSRD2.1 PR-1]
Service Compatibility [OCD 4.2 PG-2]	Usage log compatibility [SSRD2.2.7 PR-10]
Package Compatibility [OCD 4.2 PG-3]	Use open source visualization library [SSRD 2.2.1 PR-4]

Medium Priorities:

OCD Goal	SSRD
Gather relevant data for researches. [OCD 3.2 OG-2] Research Data Generation [OCD 4.2 PG-4]	Generate new analysis report[SSRD 4.1.1]
Make improvements to the digital library collections [OCD 3.2 OG-2]	Graph clustering / normalization [SSRD 3.1 SR-3]

3.2 Process Match to System Characteristics and Priorities

The choice of process model is the Spiral model, which ensures concurrency in development of different artifacts and supports a MBASE approach. The Spiral Model will allow the team to successfully match the system priorities by getting constant client feedback through each cycle while staying in our time constraint. It is actually the WinWin Spiral Model that has been decided to develop the proposed system. This "WinWin" addition includes emphasis on the critical stakeholders prior to stages that already exist in the Spiral Model. It is through this process that we can obtain project objectives, constraints, and alternatives which are crucial system priorities in general. Also, the deadlines of LCO and LCA have been set as project milestones, which represent the inception and elaboration stages of development.

The WinWin Spiral Model was also used because the team understandings of the requirements were low at the beginning due to open source integration. And since the system priorities deal directly with the client's objectives, development team's constraints and goals, the WinWin Spiral Model provides a way to incrementally develop the system, concentrating on the most crucial requirements and leaving options to drop low priority requirements. It also allows the developers and the client to gradually understand the system and for the team to design the architecture, affected by regular client feedback, which satisfies the stakeholders at every cycle.

For more details, the descriptions are provided in the [LCP 2].

3.3 Consistency of Priorities, Process and Resources

There is no direct development cost associated with the project [LCP 5.2], which is consistent with the no budget requirement described in the [SSRD 2]. The requirements of the project shall be met following the priorities which were decided in the WinWin negotiations or the client meetings. The high priority requirements have been implemented first.

Following modules are the high priority task of the system:

- Relationship Generation
  This module is for requirements in the [SSRD 3.2 SR-1], Generation the relationship matrix in order to generate the tree collection structure
- Generate Tree Structure
  This module is for the requirement, “Generate collection structure tree”, in the [SSRD 3.2 SR-2]
- Visualization of log analysis:
  This module is for the requirements, “Visualization” and “Graph node statistics”, in the [SSRD 3.2 SR-3 and SR-5]
- Search Tool
  This module is for the requiremen “Search” in SSRD SR-4. This was marked as high priority after several meetings with the client.

There is a medium priority task, “Open the object webpage”, for the proposed system. We had also used COCOMO II to estimate the system which includes this module and the person months of this project will become 8.9 person-months. It is described in the [LCP 6] or see [FRD 6].

4. Project Risk Assessment

This section identifies the major sources of risk in the project. It also provides the description of the risks which includes risk exposure quantities. Moreover, these risks were retrieved from DART, and have been rated by the key stakeholders of this project.

Table 11: <Project Risk Assessment>

Title	**RSK-01 Change of Personnel**	
Description	Only one returning member from csci577a of fall 2004 semester. There are 4 new people that were introduced to this project. On top of that, one person decided to leave the team, therefore these results in a 4 person team.	
Risk Exposure	Probability of Loss	7 (high)
	potential Magnitude(size of Loss)	7 (medium)
	RE=PL*SL=42	
Risk Reduction Leverage	Have frequent client meetings with the whole team to familiarize everyone with background information and leave session open for questions.	
Actions to Mitigate Risk	Constant feedback and meetings with returning team member and new members. Ensure understanding of proposed system through client meetings and documents from the previous semester.	
Contingency Plan	The original developers give new developers extra time and personal trainning to help them quickly get familiar with the project.	

Title	**RSK-02 Modification of open source code**	
Description	Specifically the modification of h3viewer. Modifying the software to interface and integrate it with our system may lead to unexpected delay and conflicts. There may be overlooked assumptions that can cause incompatibility with other components.	
Risk Exposure	Probability of Loss	7 (high)
	potential Magnitude(size of Loss)	6 (high)
	RE=PL*SL=49	

Risk Reduction Leverage	Frequent referral to documentation and feedback from someone familiar with h3viewer.
Actions to Mitigate Risk	Continuous in-depth research and analysis on the open source software while implementation is going on. Test completed components each step of the way and ensure problems or risks are not unnoticed.
Contingency Plan	Consult with our client Ms. Ward to get someone that is more familiar with h3viewer to help the team with the implementation.

Title	**RSK-03 Scalability of visualization clustering algorithm**	
Description	Since the data will keep increasing in DA's database, the scalability of visualization clustering algorithm may not be able to handle the large numbers of data.	
Risk Exposure	Probability of Loss	6 (high)
	potential Magnitude(size of Loss)	6 (high)
	RE=PL*SL=42 (high)	
Risk Reduction Leverage	Reduce the number of the display nodes on one view, and then use large data to test visualization clustering algorithm.	
Actions to Mitigate Risk	The system will not display all nodes on one view.	
Contingency Plan	Reduce the number of display nodes on one view, and also search for other related algorithm.	

Title	**RSK-04 Inappropriately modify the open source component**	
Description	The open source component needs to be changed to match the requirements of the proposed system, but some requirement functions may not correctly build on the component since developers might be unfamiliar with the component.	
Risk Exposure	Probability of Loss	5 (medium)
	potential Magnitude(size of Loss)	6 (medium)
	RE=PL*SL=30	

Risk Reduction Leverage	Read the document to get familiar with open source and using the component package to build the prototype.
Actions to Mitigate Risk	The developers self-study, and also have group study.
Contingency Plan	Do the coding inspection, or the peer review of codes.

Title	**RSK-05 Users misuse the proposed system**	
Description	This system provides a hyperbolic 3D viewer of the item relationship graph tree. Since it is a complex GUI, users may misuse the proposed system.	
Risk Exposure	Probability of Loss	4 (low)
	potential Magnitude(size of Loss)	4 (low)
	RE=PL*SL=16	
Risk Reduction Leverage	Give users training and clearly develop the user manual.	
Actions to Mitigate Risk	Planed to give user training and user manual. The detail schedule is described in the [LCP 2.2].	
Contingency Plan	provided error message when users misuse the system.	

5. Analysis Results

5.1 Product Features

5.1.1 Advantages

These are some advantages of the proposed system:

- The results of the analysis help Digital library staff to improve the Digital Archive.
- The results of the analysis aid Digital library staff to decide which digital collections should be added to the archive next.
- It can augment other the Digital Library services, such as provide a recommender service.

5.1.2 Limitations

The proposed system will provide a hyperbolic 3D view that attracts users. However, it may have performance problems because of loading large graphs from files that store our current information and then displays them on a 3D viewer.

5.1.3 Tradeoffs Considered

The client and developers have decided to change the system from a client/server system to a stand-alone system. The advantage is this stand-alone system will not have network traffic problem. It makes the proposed system more stable and reliable. However, the disadvantage of this change is the user may have minimal knowledge of downloading log files from Apache server. Therefore, the developer team will provide training to solve this risk or the client will provide the log file for the users.

5.1.4 Changes Considered

Since the open source library used to build the system is not the web based system and client agreed with the changing of the user interface, this specific win condition of user interface decided on the Winwin negotiation will not be considered.

Moreover, the proposed system changed from a client/server system to be a stand-alone system. Therefore, the system will not connect to DA database, and the user needs to download the log file from DA's Apache Server, and then the proposed system will import these files' data into the local database. Furthermore, during preparation for the RLCA milestone, another change on top of this was considered, where the log file is no longer obtained from the apache server log files, but provided by the client and the DA personnel. With this change, the developers can be provided with necessary data to run through the data mining algorithm. Thus, the client and developers have agreed with this change.

5.2 Commercial-Off-The-Shelf Solutions

The following table lists the open source/COTs that were selected or rejected for using in the project.

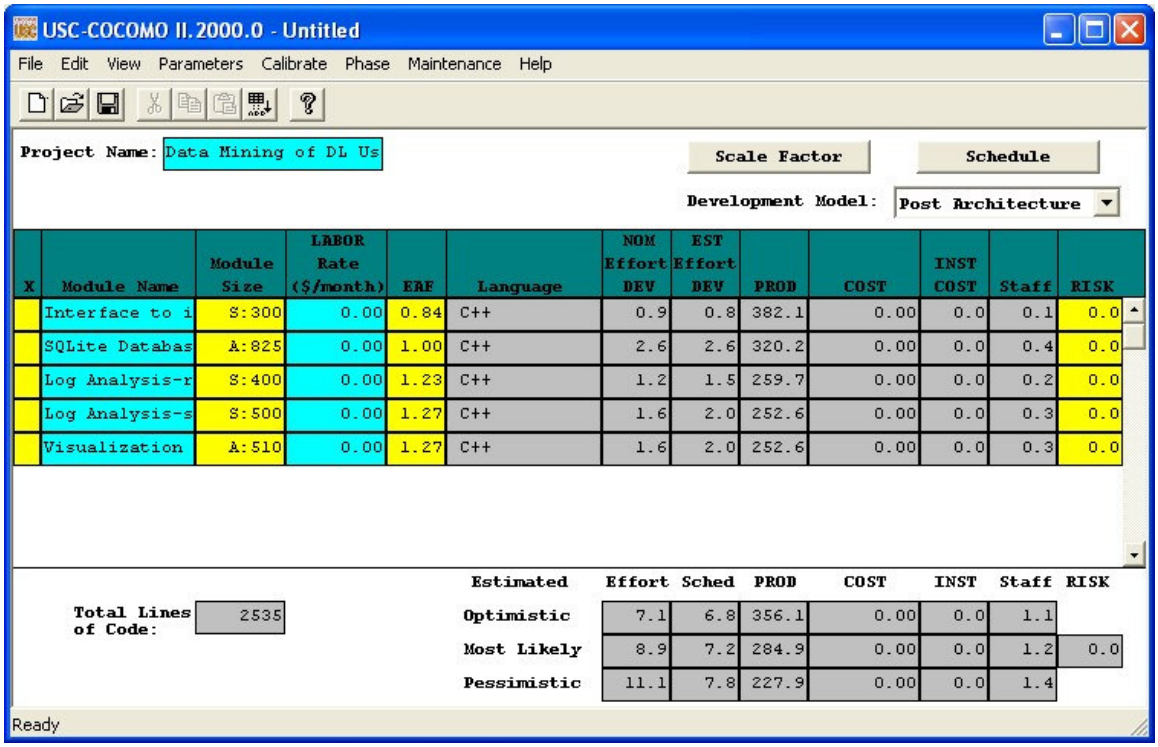
Table 12: <COTS Table>

Component Name	Status	Rationale for selection/rejection, and relevant references	Related components	
			Dependencies	Interacting Components
Visualization (SSRD SR-3)				
H3Viewer	Selected	This open source visualization library can provide a 3D hyperbolic view. Also, the client likes the interface of this component, and had agreed with using it to display the log analysis result. Since there is a scalability of using this	Linux, Windows, Unix	H3Viewer Library

		component, we have evaluated how to reduce this risk. The action to mitigate the risk is the system will not display all the nodes on one view. It will use leaf nodes to represent N items, and display content of a leaf node. Reference link: http://graphics.stanford.edu/~munzner/h3/		
Displaying Node information (SSRD SR-5)				
Open Motif	Selected	This open source is a graphical user interface that provides the intermediary mechanisms for communication between the application and the user. It provides a set of functions for creation of windows with common user interface controls like test fields, lists, buttons, etc. This feature helps the developers reduce their developing time. Reference link: http://www.opengroup.org/openmotif/	Linux, Windows	Open Motif Library

6. Appendices

Figure 2 – COCOMO Result



7. Glossary

DL
Definition: Digital Library

DA
Definition: Digital Archive

ARB
Definition: Architectural Review Board

Freeware
Definition: The software is free for people to use it.

Open-source

Definition: People free to use or modify the source code which is Open-source.

ISD
Definition: Information Services Division.

Usage Log
Definition: The log data includes the usage of digital item.