

Bioinformatics Integration Support Contract (BISC), Phase II
SYSTEM AND SOFTWARE REQUIREMENTS DESCRIPTION

IMPORT

Immunology Database and Analysis Portal

Version 1.4

Period Of Performance: September 30, 2004—September 29, 2006

Developed Under Contract Number: HHSN266200400076C

ADB Contract Number: N01-AI-40076

Delivered: September 15, 2005

Project Sponsor:

National Institutes of Health (NIH)
National Institute of Allergy and Infectious Diseases (NIAID)
Division of Allergy, Immunology, and Transplantation (DAIT)

Prepared by:

NORTHROP GRUMMAN

Information Technology

Federal Enterprise Solutions
Health Solutions
2101 Gaither Rd, Suite 600
Rockville, Maryland 20850
(301) 527-6600
Fax: (301) 527-6401
kevin.biersack@ngc.com

Contents

1.0 Introduction 1

- 1.1 Scope..... 1
- 1.2 Purpose..... 1
- 1.3 Assumptions..... 1
- 1.4 Key Objectives..... 2
- 1.5 Identification..... 2
- 1.6 References..... 2
- 1.7 Change Procedures..... 3

2.0 Stakeholder and User Descriptions4

- 2.1 Stakeholder Summary 4
- 2.2 User Summary 4

3.0 Product Overview5

- 3.1 Summary of System Capabilities..... 5
- 3.2 Product Perspective..... 6
- 3.3 Product Priorities 6

4.0 System Requirements7

- 4.1 Security 7
 - 4.1.1 System Access..... 7
 - 4.1.2 Private Project Workspace/Collaborative Workspace Access 12
 - 4.1.3 Virus Protection 12
- 4.2 Data Archiving..... 13
- 4.3 Links 13
- 4.4 Interfaces..... 13
 - 4.4.1 User Interface 13
 - 4.4.2 Database Interface..... 14
- 4.5 System Availability and Continuity of Operations 14
- 4.6 Number of Concurrent Users 14
- 4.7 Scalability 15
- 4.8 Performance 15

5.0 Overview of BISC Phase II Functional Requirements.....16

6.0 Data Repository And Sharing Requirements.....17

- 6.1 Public Data Warehouse..... 17
- 6.2 Private Project workspaces 18
- 6.3 Collaborative Workspaces 20

7.0 Collection, Storage, and Management of Data Associated with DAIT-Funded Experimental Studies.....22

- 7.1 DAIT-Funded Contracts/Grants and Programs..... 22

7.2 Research Projects and Component Experiment Groups	23
7.3 Subject and Sample Information.....	24
7.3.1 Subject Information.....	25
7.3.2 Biological Sample Information	32
7.4 Individual Experiments.....	35
7.5 Biochemical Assay Data.....	38
7.5.1 Experimental Data for Genotyping Assays	40
7.5.2 Microarray for Gene Expression Analysis	42
7.6 Capture Results Data from the Study of Genetic Associations.....	44
8.0 Reference Data.....	46
8.1 Basic Gene Data.....	47
8.2 Basic Protein Data.....	48
8.3 Taxonomy Data.....	48
8.4 Genomic Sequence Data	48
8.4.1 Assembled Chromosomes	48
8.4.2 Chromosome Data (Contigs).....	49
8.5 Gene Expression Data.....	49
8.6 Polymorphism Data	50
8.6.1 Microsatellite Polymorphism Data.....	50
8.6.2 SNP Polymorphism Data	50
8.6.3 dbMHC Polymorphism Data.....	52
8.7 Protein/Gene Network Data.....	52
8.7.1 Pathways Data	52
8.7.2 Interaction Data.....	53
8.8 AfCS Molecule Pages.....	55
8.9 ImmPort Gene List.....	55
9.0 Loading Experimental and Related Data Into ImmPort.....	57
10.0 Data Manipulation	60
10.1 Data Query and Extraction	60
10.2 Data Analysis.....	65
11.0 Ontology	67
11.1 Ontologies and Controlled Vocabularies.....	67
11.2 Aggregation and Standardization of External Data	68
12.0 Semantic Data Mapping	70
13.0 User Assistance Requirements	72
13.1 Assistance	72
13.2 Technical Assistance	72
14.0 Other Requirements	73
14.1 Applicable Standards	73

14.2 Data Sharing Functions.....	73
14.3 Electronic Notebook	75
APPENDIX A	76
APPENDIX B Patient Phenotype Data	77
APPENDIX C USE CASES	89
1. Manage Users (MU).....	89
2. Log In/Off (LG).....	89
3. Manage Contract/Grant (CG)	89
4. Manage Programs (MP)	90
5. Use Public Data (UP)	90
6. Manage Private Data (PD)	91
7. Use Private Data (UD)	91
APPENDIX D Biochemical Assays	132

Tables

Table 2-1. ImmPort System Stakeholders	4
Table 2-2. ImmPort System Users	4
Table 3-1. Summary of ImmPort System Capabilities and Features	5
Table 4-1. Overview of User Community and User Roles	7
Table 4-2. Overview of ImmPort Capabilities and Permissions for each User Role	8
Table 4-3. System User Access Requirements	10
Table 4-4. System Access and Monitoring Requirements	11
Table 4-5. Private Project Workspace/Collaborative Workspace Access Requirements.....	12
Table 4-6. Virus Protection Requirements	12
Table 4-7. Data Archiving Requirements	13
Table 4-8. Links Requirements	13
Table 4-9. User Interface Requirements.....	13
Table 4-10. Database Interface Requirements.....	14
Table 4-11. System Availability and Continuity of Operations Requirements.....	14
Table 4-12. Number of Concurrent User Requirements.....	14
Table 4-13. Scalability Requirements	15
Table 4-14. Performance Requirements	15
Table 5-1. Summary of ImmPort System Functional Requirement Types.....	16
Table 6-1. Public Data Warehouse Requirements	17
Table 6-2. Private Project Workspace Requirements	18
Table 6-3. Collaborative Workspace Requirements.....	20
Table 7-1. DAIT-Funded Contract/Grant Information.....	22
Table 7-2. DAIT-Funded Contract/Grant Information.....	22
Table 7-3. DAIT Program Information	23
Table 7-4. Research Projects and Component Experiment Groups	23
Table 7-5. Research Project Information.....	23
Table 7-6. Experiment Group Information.....	24
Table 7-7. Subject Information Requirements	25
Table 7-8. Demographic Information	25
Table 7-9. Related Subject Information	26
Table 7-10. Personal Medical History Information	27
Table 7-11. Family Medical History	28
Table 7-12. Laboratory Test Results	29
Table 7-13. Clinical Exam Findings	30
Table 7-14. Treatment/Therapeutics Information	31
Table 7-15. Composite Diagnostic Evaluation Information	32
Table 7-16. Non-Human Subject Information	32
Table 7-17. Biological Sample Information Requirements.....	33
Table 7-18. Biological Sample Information	33
Table 7-19. Experiments	35
Table 7-20. Biochemical Assay Data	38

Table 7-21. Experimental Data for Genotyping Assays	40
Table 7-22. Experimental Data for Microarray Assay for Gene Expression	42
Table 7-23. Experimental Results Types from Population Genetics Program Projects	44
Table 8-1. Reference Data Categories	46
Table 8-2. General Functional Requirements for Reference Data	46
Table 8-3. Basic Gene Information	47
Table 8-4. Basic Gene Information for Species.....	48
Table 8-5. Basic Protein Data.....	48
Table 8-6. Taxonomy Data	48
Table 8-7. Assembled Chromosomes	48
Table 8-8. Chromosome data (Contigs)	49
Table 8-9. Gene Expression Data	49
Table 8-10. Microsatellite Polymorphism Data	50
Table 8-11. SNP Polymorphism Data	50
Table 8-12. SNP Polymorphism Data for Species	51
Table 8-13. dbMHC Polymorphism Data	52
Table 8-14. Pathways Data	52
Table 8-15. Gene-Gene Interaction	53
Table 8-16. Protein-Protein Interactions.....	54
Table 8-17. Protein-Protein Interactions for Species	55
Table 8-18. AfCS Molecule Pages Data	55
Table 8-19. Publicly Available Data for Candidate Genes Comprising the ImmPort Gene List	56
Table 9-1. Experimental Data Load Requirements	57
Table 9-2. Support for Data Load Capabilities – Prioritization	58
Table 10-1. Data Query and Extraction Requirements with Priority	60
Table 10-2. Basic Gene Data Query Requirements	62
Table 10-3. Basic Protein Data Query Requirements.....	62
Table 10-4. Pathway Data Query and Extraction Requirements	63
Table 10-5. Protein Network Data Query and Extraction Requirements.....	63
Table 10-6. SNP Polymorphism Data Query Requirements	64
Table 10-3. Data Analysis Requirements with Priority	65
Table 11-1. Ontologies and Controlled Vocabularies Requirements	67
Table 11-2. Aggregation and Standardization of External Data Requirements	68
Table 12-1. Semantic Mapping Capabilities and Features of the ImmPort System	70
Table 13-1. User Assistance Requirements.....	72
Table 14-1. Applicable Standards	73
Table A1. Experimental Techniques To Be Supported by the ImmPort System	76
Table B1. DEMOGRAPHICS.....	77
Table B2. FAMILY RISK FACTOR HISTORY	78
Table B3. THERAPEUTICS.....	79
Table B4. DISEASE ACTIVITY & HEMATOLOGY (at time of sample).....	80

Table B5. DISEASE ACTIVITY - NEPHROLOGY/LFTs (at time of sample)	81
Table B6. HEMATOLOGIC FEATURES	82
Table B7. RENAL FEATURES	82
Table B8. DERMATOLOGY FEATURES	82
Table B9. RHEUMATOLOGY FEATURES	83
Table B10. CONCOMMITANT ILLNESSES	83
Table B11. CNS FEATURES	84
Table B12. ID FEATURES	85
Table B13. SEROLOGY & LABS	86
Table B14. DAS28	87
Table B15. SLEDAI (SLE patients only)	88
Table C1. Manage Users	89
Table C2. Log In/Off	89
Table C3. Manage Contract/Grant	90
Table C4. Manage Programs	90
Table C5. Use Public Data	90
Table C6. Manage Private Data	91
Table C7. Use Private Data	91
Table D1. Biochemical Assays To Be Supported by ImmPort	132

Figures

Figure 6-1. Private and Collaborative Workspaces within ImmPort 20

BISC System and Software Requirements Description Version History

Version	Date	Description
1.0	2/28/2005	Initial release of the BISC Phase II System and Software Requirements Description
1.1	4/6/2005	Modifications were made throughout the document. Assumptions in Section 1 were clarified. Requirements were gathered into tables in sections 3, 5, 7, 8, 10, and 12. Appendix B containing use cases was added.
1.2	6/17/2005	Assumptions in Section 1 were further clarified. Requirements were updated on the basis of feedback from the Project Officer and Population Genetics contractors. Priority was assigned to requirement groups, specifying requirements that will be implemented in ImmPort Version 1.0.
1.3	8/11/2005	Version 1.3 presents a reorganization of several sections since Version 1.2. Sections 4 through 14 were rearranged to present a more logical flow of the requirements. Section 4 in Version 1.2 is now Section 5. Section 5 in Version 1.2 is now Section 7. Section 6 in Version 1.2 is now Section 8. Section 7 in Version 1.2 is now Section 9. Section 8 in Version 1.2 is now Section 10. Section 9 in Version 1.2 is now Section 11. Section 10 in Version 1.2 is now Section 12. Section 11 in Version 1.2 was removed from Version 1.3. Section 12 in Version 1.2 has been divided in System Requirements in Section 4 and Data Repository and Sharing Requirements, which now appear in Section 6. Section 13 in Version 1.2 is now Section 14. Parts of Section 12 in Version 1.2 are now in Section 6. Other changes include refinement of requirements and their prioritization.

1.0 INTRODUCTION

1.1 SCOPE

The scope of the contract is to provide advanced information technology support in the production, analysis, archiving, and exchange of scientific data for a diverse community of life science researchers by: (1) conducting a requirements assessment of bioinformatics needs in a diverse community of basic and clinical researchers, (2) prototyping a system for the collection, storage, and analysis of data, (3) designing, implementing and maintaining a data warehouse of genomic, proteomic and all other related data relevant to the research of these programs, (4) developing or selecting specialized applications and providing technical assistance to participating centers in the capture, storage, management, query, and analysis of the data, and, (5) measuring performance and benefits resulting from these technical support activities and planning for their appropriate development and use in the future.

1.2 PURPOSE

The Immunology Database and Analysis Portal (ImmPort) system is being developed by a team led by Northrop Grumman Information Technology (IT) through the Bioinformatics Integration Support Contract (BISC). The ImmPort system is intended to serve as a long-term, sustainable archive of data generated by investigators funded through the Division of Allergy, Immunology and Transplantation (DAIT) of the National Institute of Allergy and Infectious Disease (NIAID), National Institutes of Health (NIH). The core component of the ImmPort system research database will be an extensive data warehouse containing an integration of experimental data supplied by DAIT-funded investigators and reference data extracted from a variety of public databases. The ImmPort system research database will also provide private project workspaces for the use of individual research projects. The ImmPort system will be freely accessible as a public resource to all investigators funded through DAIT.

This System and Software Requirements Description (SSRD) identifies all known user and system requirements that have been identified by the Northrop Grumman IT BISC Team (BISC Team) for the ImmPort system being developed for NIAID/DAIT. The key objective of the first year of the Base Period of Performance (POP) is to develop and field an operable system of databases that includes interfaces for data collection and data analysis and is supported by appropriate ontologies or structured vocabularies. The requirements identified in this document were primarily derived from the data storage and analysis needs of the scientists involved in the following two DAIT research programs (pilot projects).

- Population Genetics Analysis Program: Immunity to Vaccines/Infections
- Human Leukocyte Antigens (HLA) Region Genetics in Immune-Mediated Diseases

The DAIT Project Officer (PO) for the BISC specified the use of these pilot projects for definition of user requirements for ImmPort Version 1.0. While this version of the SSRD focuses on the requirements for the first year of the contract Base Period (years one and two), requirements are also identified that will not be implemented in the initial release of the ImmPort system. The user and system requirements identified for the two pilot projects are expected to be representative of the requirements of all DAIT-funded investigators.

1.3 ASSUMPTIONS

The following assumptions were made in defining the user and system requirements for the ImmPort system being developed under the BISC Phase II contract:

- Little, if any, interactive data entry (online transaction processing) will occur. The ImmPort system will reject files submitted for bulk loading that do not meet minimum data submission

requirements and generate an error report. The ImmPort system may provide interactive error correction of data during bulk loading into the database in future versions.

- The ImmPort system will provide an ontology specifically enhanced to include information on immunology, genetics, and the most widely used experimental procedures used by DAIT-funded research groups.
- By the end of the six-year period of performance of the BISC Phase II contract, the ImmPort system will serve as the primary data repository of record for raw and processed experimental data for the approximately 1100 DAIT-funded research groups.
- Each NIAID/DAIT research grant and contract will provide a maximum of 0.5 to 1.0 terabyte of data to ImmPort. By the end of the six-year period of performance of the BISC Phase II contract, the ImmPort system data storage capacity required to support approximately 1100 grants and contracts will be approximately 10 terabytes of online storage and up to 1000 terabytes of near-line and/or archival storage.
- Each DAIT research grant and contract supported by ImmPort will have an average of 5 ImmPort users, of which, one to two will be considered frequent and sophisticated users of the ImmPort system (power users). Each power user will spend an average of one hour per day using ImmPort.
- Although the system will be available 24 hours per day 7 days per week, except when maintenance is being performed, active use will only occur during 16 hours of each 24 hour day.
- The initial release of the ImmPort system, Version 1.0, will be implemented in October 2005 and will support between 10 and 30 concurrent users. By the end of the six-year POP of the BISC Phase II contract, the number of concurrent users will be between 100 and 200.
- Complex data queries will be handled.
- Analysis pipelines will be supported.
- The ImmPort system user interface will be web-based.

1.4 KEY OBJECTIVES

The key objective of this SSRD is to identify all system and user requirements for the ImmPort system, whether or not they will be implemented in Version 1.0 of the ImmPort system. This document is to be considered a “work in progress” and will evolve during the life of the BISC effort as new requirements are identified and others are modified or deleted. Additional versions will be released as needed to extend or clarify ImmPort system and software requirements.

1.5 IDENTIFICATION

Northrop Grumman Information Technology provides development of the BISC system (ImmPort) under Contract No. HHSN266200400076C, ADB Contract No. N01-AI-40076.

1.6 REFERENCES

- BISC Phase II Project Management Plan, version 1.3, July 11, 2005
- BISC Development Environment Description, version 1.0, January 31, 2005
- Northrop Grumman IT Contract No. HHSN266200400076C, ADB Contract No. N01-AI-40076
- Institute of Electrical and Electronics Engineers and the Electronic Industries Association (IEEE/EIA) 12207 Standard for Information Technology:
 - 12207.0 Software Life Cycle Processes
 - 12207.1 Software Life Cycle Processes-Life Cycle Data
 - 12207.2 Software Life Cycle Processes-Implementation Considerations

- Enhanced Asset-Based Logicon Engineering Resource (ENABLER) 5.0, Standard Project Procedures (SPP), June 1704
- Population Genetics Projects Site Visit Notes

1.7 CHANGE PROCEDURES

This document will be used throughout the life of the BISC project. It will be updated as changes are made to the development environment. When a significant change occurs, this document will be updated and a new version (e.g., 2.0) will be published. Minor changes may be published in an interim release (e.g., 1.2).

2.0 STAKEHOLDER AND USER DESCRIPTIONS

2.1 STAKEHOLDER SUMMARY

Table 2-1 lists those groups that are considered stakeholders in the development of the ImmPort system.

Table 2-1. ImmPort System Stakeholders

Stakeholder Group	Role
NIH/NIAID/DAIT	Provide funding for the BISC Phase II contract and for all investigators who will submit experimental data into the ImmPort system.
DAIT BISC Project Officer	Responsible for performance of the BISC effort.
DAIT-funded research scientists	Will use the ImmPort system as the repository for experimental data to comply with NIH data sharing policy and as a bioinformatics resource.
BISC Phase II Contractor Team	Developers of the ImmPort system.

Table 2-1 lists groups that are considered stakeholders in development of the ImmPort system.

2.2 USER SUMMARY

Table 2-2 lists those groups that are expected to be users of the ImmPort system.

Table 2-2. ImmPort System Users

User Group	Description of Use
Population Genetics research scientists	Will submit research data to the ImmPort system and will use the ImmPort system to analyze their data, along with data provided by other Population Genetics Program scientists, and other data gathered in the ImmPort system.
HLA Region Genetics in Immune-Mediated Disease research scientists	Will submit research data to the ImmPort system and will use the ImmPort system to analyze their data, along with data provided by other HLA Region Genetics Program scientists, and other data gathered in the ImmPort system.
Other DAIT-funded research scientists	Will submit research data to the ImmPort system and will use the ImmPort system to analyze their data, along with data provided by other DAIT-funded scientists, and other data gathered in the ImmPort system.
DAIT Project Officers	Will assess the status of research projects for which they are responsible by viewing research results in the ImmPort database.

Table 2-2 lists those groups that are expected to be users of the ImmPort system.

3.0 PRODUCT OVERVIEW

3.1 SUMMARY OF SYSTEM CAPABILITIES

The product of the BISC Phase II contract will be the ImmPort system. The key objective of the first year of the Base POP is to develop and field ImmPort Version 1.0, an operable system of databases that includes interfaces for data collection and data analysis and is supported by the ImmPort ontology. Table 3-1 summarizes all ImmPort system capabilities and features identified as user and system requirements, including many that will not be fully implemented in ImmPort Version 1.0. These capabilities and features may change based on the requirements identified in future versions of this SSRD. A general schedule is assigned to each capability/feature, indicating in which version of the system the capability/feature will be implemented. Capabilities/features that will be partially implemented in Version 1.0 or subsequent minor releases during the second year of the BISC, are designated “Version 1.” These capabilities/features will continue to be enhanced in subsequent quarterly releases, e.g., Version 1.1. Other capabilities/features scheduled for implementation in ImmPort Versions 2.0 and 3.0 are included in the table. Note that prioritization of individual-level requirements is discussed in Section 3.3.

Table 3-1. Summary of ImmPort System Capabilities and Features

Capabilities/Features	Capability/Feature Description	Schedule
Public Data Warehouse	A data warehouse that will contain raw and processed experimental data and reference data extracted from public sources. The data warehouse will contain data available to all ImmPort system users. The data warehouse will include near-line storage for data that is infrequently used and for large data sets that do not need to be instantly accessible.	Version 1
Private Project Workspaces (PPW)	Private project workspaces (PPWs) in which the individual investigators associated with DAIT-funded contracts or grants can store and analyze raw, unpublished data associated with a research project. The data in a PPW is shielded from the view of other ImmPort system users. When the experimental findings represented by data in the PPW are accepted for publication, the data is published to the Public Data Warehouse, thereby making it available to all ImmPort system users.	Version 1
Collaborative Workspaces (CW)	Collaborative workspaces (CWs) in which the investigators associated with DAIT-funded contracts or grants can securely share unpublished data (a subset of the data in the PPW of each investigator) with the data shielded from the view of other ImmPort system users.	Version 1
Data Marts	A series of data marts containing processed data extracted from the data warehouse and tailored to the needs of ImmPort system users.	Version 1
System Access Management	Access control and security of hardware, software, and data associated with the ImmPort system.	Version 1
Data Access Management	Access control and data security associated with PPWs and CWs.	Version 1
Data Sharing	Allow user-controlled sharing of results with other ImmPort system users or groups of users.	Version 1
Data Deposition	The process of loading experimental information into the data warehouse.	Version 1
Data Query	A simple data query capability (to be implemented in contract year one), along with some complex data query capabilities, which will be enhanced on a continuous basis in future releases of the ImmPort system. These query capabilities will allow investigators to extract and download specified data sets from the ImmPort system.	Version 1
Data Extraction	A data extraction capability tailored to the requirements of the dbMHC database.	Version 1
Data Analysis	A complex data analysis capability allowing direct analysis of data stored in the ImmPort system, including data mining capabilities and semantic architecture-based query construction. This capability will allow analyses using data in more than one PPW as well as data in the public warehouse.	Version 1

Capabilities/Features	Capability/Feature Description	Schedule
Ontology	An ontology that will contain standard terminology that will be used for stored experimental data across all data submitted to the ImmPort system (to be implemented during contract year one and expanded continuously in future years). The ontology will also be used to assist in standardizing data analysis queries so results are repeatable (to be implemented during contract year two and expanded continuously in future years).	Version 1
Data Transformation	Transformation of data into a set of common values based on the ontology. For example, all values for patient weight may be transformed from pounds into grams. This will facilitate data analysis.	Version 2
Data Integration	Merging experimental results and publicly available data from multiple sources into common locations by subject area. An integrated data model that will utilize shared data elements to link data from different experimental sources will facilitate the merging process.	Version 1
Journaling	A journaling function to catalog and store query code and/or results for reuse and for combination into complex queries and to maintain data access history.	Version 1
Data Archive	As necessary to maintain and enhance system performance, seldom-used data will be moved from online storage to "near-line" and/or offline storage.	Version 3
Tutorials	Data analysis tutorials and other educational support components.	Version 1
Technical Assistance	Technical assistance in the use of ImmPort capabilities.	Version 1

Table 3-1 summarizes the ImmPort system capabilities and features, with an implementation priority assigned to each.

3.2 PRODUCT PERSPECTIVE

The BISC Project is a very ambitious undertaking. The goal is to produce a data warehouse in which all NIAID/DAIT-funded scientists will place their experimental data. The ImmPort system will also contain publicly available data that are gathered and stored in a manner that is tailored to the needs of immunology research. It will be supported by an ontology specifically extended to include human immunology, a unique semantic architecture-based query-building tool, and an artificial intelligence-based data analysis capability.

3.3 PRODUCT PRIORITIES

The following sections of the SSRD use a requirement prioritization numbering scheme ranging from 1 through 4. Requirements that are of extremely high importance are designated with a priority 1 level and will be implemented in ImmPort Version 1.0. Requirements of very high importance are designated with a priority 2 level and will be implemented during the second year of the BISC Base POP. Priority 3 requirements are considered very important, and they will be reevaluated after the release of ImmPort Version 1.0 to determine which will be implemented along with Priority 2 requirements and which will be implemented soon afterwards. Implementation of Priority 4 requirements will take place after the BISC 2-year base POP.

4.0 SYSTEM REQUIREMENTS

4.1 SECURITY

Security requirements address needs for controlling access to the ImmPort system and access to private data within the system.

Refer to Section 3.3 Product Priorities for a description of the requirement prioritization numbering scheme used in this section.

4.1.1 System Access

Access to ImmPort is restricted to three specific user communities: DAIT community, DAIT-funded research community, and ImmPort administration community. User role types within each community have been identified on the basis of requirements collected on how typical users will need to interact with the system in order to use data, manage data, and manage the system.

Table 4-1. Overview of User Community and User Roles

User Community	User Community Description	User Role	User Role Description
DAIT Community	DAIT employees	DAIT BISC Project Officer	The DAIT BISC Project Officer oversees the BISC II project. He/she performs system access control functions and/or delegates them to other DAIT employees or a Functional Administrator (see below for description on Functional Administrator).
		DAIT User	The DAIT user oversees specific grants/contracts that use ImmPort. He/she notifies the DAIT BISC Project Officer when new research grants/contracts are awarded.
DAIT-Funded Research Community	Users who work on research projects that are funded by DAIT	Principal Investigator (PI)	The PI serves as the point of contact for a research grant/contract. He/she manages research and administration activities associated with a grant/contract. He/she may also delegate these tasks to staff.
		Other Staff	Other Staff include any non-PI user associated with a research grant/contract. Other Staff perform administration and/or research duties (as determined by the PI) in support of the grant/contract.
ImmPort Administration Community	Members of the ImmPort development team	Functional Administrator	The Functional Administrator performs system access control functions in ImmPort in support of the DAIT BISC Project Officer.
		Systems Administrator	The Systems Administrator monitors the security of the system, ensures resource optimization, and performs troubleshooting and maintenance.

Table 4-1 summarizes the ImmPort system user community and user roles.

Table 4-2 provides an overview of ImmPort capabilities and permissions for each of the six user roles defined in Table 4-1. These capabilities and permissions (privileges) are designed to protect the data stored in ImmPort from unauthorized access. Data in ImmPort will be classified as either “public,” available to all users, or “private,” available to specific users. A check mark (✓) in the matrix indicates the role (column) is granted the capability/permission or privilege (row). A plus (+) indicates privileges that the DAIT BISC Project Officer or Functional Administrator may assign to DAIT users. An asterisk (*) indicates a privilege that a PI may assign to staff members of his/her research project or to other ImmPort users not associated with the research project. “CRUD” refers to create, read, update, or deactivate activities. The IDs of corresponding use cases, found in Appendix C, are included in the table for easy reference. Use cases were not developed for all capabilities and permissions shown in Table 4-2.

ImmPort Version 1.0 will provide very limited user management capabilities for the PPW. Each capability/permission designated with an asterisk in the “Other staff” column will be provided to all DAIT-funded scientists given access to the private research project data. The one exception will be the “Manage Data Access/Use” capability/permission. The PI will be allowed to designate a PPW administrator to support this capability. The capabilities and permissions designated with a plus in the DAIT user column will not be provided in ImmPort Version 1.0 to DAIT users. The capabilities and permissions designated with asterisks and pluses will be provided in ImmPort Version 2.0.

Table 4-2. Overview of ImmPort Capabilities and Permissions for each User Role

Use Case Packages	Capability/Permission	DAIT Community		DAIT Funded Research Community		ImmPort Administration Community		Use Case ID
		DAIT BISC Project Officer	DAIT User	PI	Other Staff	Functional Administrator	Systems Administrator	
Management/Administration of System Users, Projects, and Programs								
Manage Users	Register User (self)	✓	✓	✓	✓	✓	✓	MU-1
	Approve/Reject Registration	✓				✓	✓	MU-2
	Create User	✓				✓	✓	MU-3
	Query User	✓	✓	✓	✓	✓	✓	MU-4
	View/Update User (self)	✓	✓	✓	✓	✓	✓	MU-5
	Deactivate User	✓				✓	✓	MU-6
	Manage User Role	✓				✓	✓	
Log In/Off	Log In	✓	✓	✓	✓	✓	✓	LG-1
	Retrieve Login Information	✓	✓	✓	✓	✓	✓	LG-2
	Log Off	✓	✓	✓	✓	✓	✓	LG-3
	Change Password (self)	✓	✓	✓	✓	✓	✓	
	Change Password (others)						✓	
Manage Contracts/Grants	Create Contract/Grant	✓				✓	✓	CG-1
	Deactivate Contract/Grant	✓				✓	✓	CG-2
	Modify Contract/Grant	✓				✓	✓	CG-3
	Search for Contract/Grant	✓	✓	✓	✓	✓	✓	CG-4
Manage Programs	Create Program	✓				✓	✓	MP-1
	Deactivate Program	✓				✓	✓	MP-2
	Modify Program	✓				✓	✓	MP-3
	Search for Program	✓	✓	✓	✓	✓	✓	MP-4
Manage System Reports	Manage System Reports (CRUD)	✓				✓	✓	N/A
Public Data Functionality								
Manage Public Reference Data	Load Reference Data					✓	✓	N/A
	Manage Reference Data Organization					✓	✓	N/A
	Archive Reference Data					✓	✓	N/A
Manage Public Experimental Data	Load Experimental Data	✓		✓	*	✓	✓	N/A
	Manage Experimental Data Organization			✓	*	✓	✓	N/A
	Archive Experimental Data			✓	*	✓	✓	N/A
Use Public Data	View Public Data	✓	✓	✓	✓	✓	✓	UP-1
	Query Public Data	✓	✓	✓	✓	✓	✓	UP-2

Use Case Packages	Capability/Permission	DAIT Community		DAIT Funded Research Community		ImmPort Administration Community		Use Case ID
		DAIT BISC Project Officer	DAIT User	PI	Other Staff	Functional Administrator	Systems Administrator	
(Reference and Experimental)	Download Public Data	✓	✓	✓	✓	✓	✓	UP-3
	View/Query Ontology	✓	✓	✓	✓	✓	✓	UP-4/UP-5
	Use Ontology to Query Public Data	✓	✓	✓	✓	✓	✓	UP-6
	Conduct Public Data Analyses	✓	✓	✓	✓	✓	✓	
	Execute/View Canned and Ad Hoc Reports	✓	✓	✓	✓	✓	✓	
	View Journal	✓	✓	✓	✓	✓	✓	
Management/Administration of Private Data Access and Use								
Manage Private Data	Create PPW and CW			✓		✓	✓	
	Manage Data Access/Use			✓	*	✓	✓	PD-1/PD-2
	Manage Data Organization			✓	*		✓	
	Load Experimental Data into PPW			✓	*		✓	
	Update Experimental Data			✓	*		✓	
	Remove Experimental Data from User's View			✓	*		✓	
	Publish Experimental Data to CW			✓	*			
	Publish Experimental Data to Public Data Area	✓		✓	*	✓	✓	PD-3
Private Data Functionality								
Use Private Data	View Data in PPW		+	✓	*		✓	UD-1
	Query Data in PPW			✓	*		✓	UD-2
	Save Private Data			✓	*		✓	
	Conduct Private Data Analyses		+	✓	*		✓	
	View Journal		+	✓	*		✓	

Table 4-2 provides an overview of the ImmPort system access privileges.

As seen in Table 4-2, there are two general areas of activity in ImmPort that are relevant to data security. The management and administration activities related to system access will ensure that only members of the three user communities described in Table 4-1 have access to ImmPort. Both the DAIT BISC Project Officer role (under DAIT Community) and the Functional Administrator role (under ImmPort Administration Community) will have the capability to authorize access to the system. It is anticipated that the Functional Administrator will provide administrative support to the DAIT BISC Project Officer in maintaining information on users, research contracts/grants funded by NIAID/DAIT, and research programs (e.g., Population Genetics, HLA Region Genetics). The ImmPort system will have the functionality to manage changes to user status and grants/contracts to ensure the continued security of the public data.

The second area of data security activity involves management and administration of private project data stored in the PPW associated with the research project and access to that data. The PI who creates a research project is responsible for controlling access to the research project’s PPW. The PI role may choose to share data security responsibilities with other research project staff members at his/her discretion by assigning a Project or PPW administrator. The PI always has access to the data, queries, and analyses stored by other members of the research project and overriding capabilities in determining who may have access to these data, queries, and analyses. Table 4-3 lists specific system user access requirements for ImmPort.

Table 4-3. System User Access Requirements

ID	Requirements	Priority
SUA_1	Access to the system shall be restricted to users of the following communities: DAIT community, DAIT-funded research community, and the ImmPort administration community. These communities are mutually exclusive.	1
SUA_2	<ul style="list-style-type: none"> ■ The system shall provide a registration process to collect required information from an individual requesting access as a DAIT user. ■ User Community (DAIT, DAIT-funded, ImmPort Admin) ■ User’s Last Name ■ User’s First Name ■ User’s Middle Initial (optional) ■ User’s Organization ■ Phone Number ■ E-mail Address ■ User Name ■ Password ■ Security Question with answer ■ NIH Employee ID ■ NIAID Contract/Grant Number(s) that the User Will Oversee (optional) 	1
SUA_3	<ul style="list-style-type: none"> ■ The system shall provide a registration process to collect required information from an individual requesting access as a DAIT-funded user. ■ User Community (DAIT, DAIT-funded, ImmPort Admin) ■ User’s Last Name ■ User’s First Name ■ User’s Middle Initial (optional) ■ User’s Organization ■ Phone Number ■ E-mail Address ■ User Name ■ Password ■ Security Question with Answer ■ NIAID/DAIT Contract/Grant Number(s) 	1
SUA_4	<ul style="list-style-type: none"> ■ The system shall provide a registration process to collect required information from an individual requesting access as an ImmPort administration user. ■ User Community (DAIT, DAIT-funded, ImmPort Admin) ■ User’s Last Name ■ User’s First Name ■ User’s Middle Initial (optional) ■ User’s Organization ■ Phone Number ■ E-mail Address ■ User Name ■ Password ■ Security Question with Answer 	3

SUA_5	The DAIT BISC PO and/or ImmPort Functional Administrator shall have access to the registration information to verify the individual's association with the DAIT community.	1
SUA_6	The DAIT BISC PO and/or ImmPort Functional Administrator shall have access to the registration information to verify the individual's association with the DAIT-funded community.	1
SUA_7	The system shall collect information from an individual seeking to become an ImmPort administrator user that the DAIT BISC PO or ImmPort Functional Administrator can use to verify the individual's association with the ImmPort administration community.	3
SUA_8	The system shall provide a range of access privileges appropriate to the DAIT BISC PO user role type (see Overview of ImmPort Capabilities and Permissions for each User Role).	1
SUA_9	The system shall provide a range of access privileges appropriate to the DAIT user role type (see Overview of ImmPort Capabilities and Permissions for each User Role).	2
SUA_10	The system shall provide a range of access privileges appropriate to the DAIT-funded PI role type (see Overview of ImmPort Capabilities and Permissions for each User Role).	1
SUA_11	The system shall provide a range of access privileges appropriate to the DAIT-funded Other Staff role type (see Overview of ImmPort Capabilities and Permissions for each User Role).	2
SUA_12	The system shall provide a range of access privileges appropriate to the ImmPort Functional Administrator role type (see Overview of ImmPort Capabilities and Permissions for each User Role).	2
SUA_13	The system shall provide a range of access privileges appropriate to the ImmPort Systems Administrator role type (see Overview of ImmPort Capabilities and Permissions for each User Role).	2
SUA_14	The experimental data stored in the public data warehouse shall be accessible to any user.	1
SUA_15	The reference data stored in the public data warehouse shall be accessible to any user.	1
SUA_16	When a user who has belonged to the DAIT-funded research community is no longer associated with an active DAIT contract/grant, the individual shall cease to have access to ImmPort.	1
SUA_17	The DAIT BISC PO and Functional Administrator shall have the capability to provide continued access to those DAIT-funded research users whose DAIT contract/grant ends by manually extending the ImmPort Access end date of the contract/grant. (This requirement highlights the potential for difference between the end date of a contract/grant's access to ImmPort, and the contract/grant's official period of performance.)	1
SUA_18	The DAIT BISC PO and Functional Administrator shall have the capability to reinstate access to an inactive DAIT-funded user by associating the user with a current contract/grant (i.e., a contract/grant whose access end date to ImmPort is active). For DAIT-funded users, "inactive" refers to a user who is not associated to any current contract/grant, and therefore, cannot access the system.	1
SUA_19	The DAIT BISC PO and Functional Administrator shall have full capabilities in managing users (see Overview of ImmPort Capabilities and Permissions for each User Role), including the abilities to add new users, rescind the access rights of existing users, and restore the access rights of former users.	1
SUA_20	The PI shall be able to de-associate a user directly from his/her contract/grant. (To maintain the data security of the system, it is important to capture users whose system access rights should be revoked as soon as possible.) (This is a separate process from exercising PPW/CW access control.)	2
SUA_21	When an individual seeking to become a DAIT-funded Other Staff registers using the front-end registration process, the system shall have the PI(s) associated with the contract/grant number(s) that the individual lists during the registration process validate that the individual is indeed associated with the DAIT contract/grant number(s).	2
SUA_22	The DAIT BISC PO or Functional Administrator shall have access to registration information (and PI validation for DAIT-funded Other Staff) in order to approve or reject requests for access.	1

Table 4-3 identifies system user access requirements.

Table 4-4 lists system access and system monitoring requirements for ImmPort.

Table 4-4. System Access and Monitoring Requirements

ID	Requirements	Priority
SAM_1	The system shall uniquely identify each user.	1
SAM_2	A user shall authenticate his/her identity at logon time by supplying an authenticator (e.g., password) in conjunction with his/her user identification (ID) prior to the execution of any application or utility on the system.	1
SAM_3	The user's ID shall be associated with all auditable actions taken by the user.	3
SAM_3a	The system shall maintain an audit record of login/logoff actions.	3

SAM_3b	The system shall maintain an audit record of user activities to create a contract/grant record.	3
SAM_3c	The system shall maintain an audit record of user activities to update a contract/grant record.	3
SAM_3d	The system shall maintain an audit record of user activities to deactivate a contract/grant record.	3
SAM_3e	The system shall maintain an audit record of user activities to create a user record.	3
SAM_3f	The system shall maintain an audit record of user activities to update a user record.	3
SAM_3g	The system shall maintain an audit record of user activities to associate a user from a contract/grant.	3
SAM_3h	The system shall maintain an audit record of user activities to disassociate a user from a contract/grant.	3
SAM_4	The system shall support encrypted data transmission over the Web.	1
SAM_5	The system shall safeguard against unauthorized retrieval of data being exported from the system.	1
SAM_6	The system shall maintain an audit trail on deposition of experimental data in the public data warehouse.	1
SAM_7	The system shall maintain an audit trail on access of experimental data in the public data warehouse.	2
SAM_8	The system shall journal automated activities, examples of which may include data versioning, transfer of data to near-line storage, and data archiving.	2
SAM_9	The system shall journal data versioning.	2
SAM_10	The system shall journal transfer of data to near-line storage.	3
SAM_11	The system shall journal data archiving.	3
SAM_12	The system shall use a consistent time zone for the basis of time stamping activities.	1
SAM_13	The system shall time stamp when a user's registration is approved/rejected.	1
SAM_14	The system shall time stamp when a user's account is created/modified.	1
SAM_15	The system shall make user passwords stored in the database unintelligible to unauthorized parties.	1

Table 4-4 summarizes the ImmPort system access and monitoring requirements.

4.1.2 Private Project Workspace/Collaborative Workspace Access

This section addresses the access control requirements surrounding the PPWs and CWs described in Sections 4.2 and 4.3, respectively. Table 4-5 provides specific requirements related to workspace access.

Table 4-5. Private Project Workspace/Collaborative Workspace Access Requirements

ID	Requirements	Priority
PPC_1	The PI shall assign access privileges to the workspace to other users.	1
PPC_2	The system shall retain the flexibility for the PI with administrator rights to the workspace (the PI who created the workspace) to grant administrator rights to another user.	1
PPC_3	A user with authorized access to a workspace shall have read-write privileges unless access restrictions are imposed by the PI.	2
PPC_4	The PI (or designee) shall review a list of users accessing data in the workspace on at least a quarterly basis.	2
PPC_5	A journal shall record user access to a workspace. The journal should include the user ID, date and time of access.	2

Table 4-5 summarizes the ImmPort system PPW access requirements.

4.1.3 Virus Protection

This section addresses anti-virus protection requirements.

Table 4-6. Virus Protection Requirements

ID	Requirements	Priority
VIP_1	The system shall scan for viruses at least on a weekly basis.	1
VIP_2	Critical security patches to the system shall be tested in accordance with the Northrop Grumman IT Testing Plan.	1

Table 4-6 summarizes the ImmPort system virus protection requirements.

4.2 DATA ARCHIVING

It is estimated that work performed for a research grant/contract may produce up to one terabyte of data in total over the complete course of the grant/contract's period of performance. To ensure the system's continued performance, archiving will store historical data that do not require real-time access. Table 4-7 lists ImmPort system requirements for data archiving.

Table 4-7. Data Archiving Requirements

ID	Requirements	Priority
DAA_1	The system shall provide archiving capabilities to store historical data that do not require real-time access.	2
DAA_2	The archive shall extend the system's storage capacity to store up to one terabyte of total data for each contract/grant by the end of the full BISC 4-year option period..	4
DAA_3	The system shall provide either a manual archive function or an automated archive function (driven by time or storage capacity).	3
DAA_4	The system shall enable archiving of data associated with a given contract/grant.	3
DAA_5	The system shall enable archiving of data within a given contract/grant associated with a specified experiment group, experiment, or biochemical analysis.	4
DAA_6	Backup copies of data archives shall be stored in a secure offsite location.	3

Table 4-7 summarizes the ImmPort system data archiving requirements.

4.3 LINKS

The system will serve as a portal to external bioinformatics resources, including data repository resources, analytical resources, and related publications. Table 4-10 lists ImmPort system requirements for linkage to external data stores.

Table 4-8. Links Requirements

ID	Requirements	Priority
LIN_1	The system shall provide links to research publications that used experimental data generated by DAIT-funded research and submitted to ImmPort.	4
LIN_1a	The system shall provide a mechanism for a user to voluntarily submit information on his/her research publication that used experimental data generated by DAIT-funded research and submitted to ImmPort.	4
LIN_2	The system shall provide links to external data and analytical resources that representative users indicate are of interest to them.	2

Table 4-8 summarizes the ImmPort system links requirements.

4.4 INTERFACES

User interfaces and database interfaces are addressed in this section. No external hardware interface requirement has been identified.

4.4.1 User Interface

The ImmPort user interface will emphasize organization, feedback, and a consistent design. A style guide will document specific interface guidelines. Table 4-9 lists ImmPort system user interface requirements.

Table 4-9. User Interface Requirements

ID	Requirements	Priority
INT_1	The system interface shall comply with the Northrop Grumman IT Bioinformatics Interface Style Guide Document.	1
INT_2	The system shall accommodate both the computer expert and non-expert user in navigating information content.	1
INT_2a	The system shall enable the user to browse a directory catalogue identifying the contents of the public	2

	data warehouse.	
INT_2b	The system shall enable the user to navigate a directory catalogue identifying the contents of the public data warehouse.	2
INT_2c	The system shall provide a section that informs the user of additions/changes that have taken place since the last release.	2
INT_3	The system shall notify the user when data input is inconsistent with constraints.	1
INT_4	The system shall comply with Section 508 standards for accessibility.	1

Table 4-9 summarizes the ImmPort system user interface requirements.

4.4.2 Database Interface

The ImmPort system will provide appropriate interface capabilities with diverse applications, as enumerated in Table 4-10.

Table 4-10. Database Interface Requirements

ID	Requirements	Priority
INT_5	The ImmPort system shall support interfaces utilizing diverse applications to the Relational Database Management System, including open database connectivity (ODBC), Java Database Connectivity (JDBC), Oracle Call Interface (OCI), Oracle C++ Interface (OCCI), and Oracle Object for OLE (OO4O), and XML.	1
INT_6	The system shall support diverse database APIs that will access the system's databases.	3

Table 4-10 summarizes the ImmPort system database interface requirements.

4.5 SYSTEM AVAILABILITY AND CONTINUITY OF OPERATIONS

The section identifies requirements for making the ImmPort system available during the normal course of business operations as well as for recovery from unexpected events. Table 4-11 lists ImmPort system requirements for system availability and continuity of operations.

Table 4-11. System Availability and Continuity of Operations Requirements

ID	Requirements	Priority
SAC_1	The ImmPort system shall be available to its users at all times (24 hours per day, 7 days per week) when hardware or software maintenance is not being performed or new releases are being installed.	1
SAC_2	The ImmPort system database shall be available to DAIT-funded investigators a minimum of 16 hours per day, 250 days per year (approximately two-thirds of all available time).	1
SAC_3	In case of catastrophic failure of any type, the ImmPort system shall be restorable to the previous day's configuration, including all data.	1
SAC_4	The system shall operate in accordance with a data backup/recovery plan.	1
SAC_5	System backup data shall be maintained in a secure offsite location.	1

Table 4-11 summarizes the ImmPort system availability and continuity of operations requirements.

4.6 NUMBER OF CONCURRENT USERS

The number of ImmPort users is expected to grow substantially over six years. Table 4-12 identifies the expected number of concurrent users during the initial ImmPort release and the projected number of concurrent users to be supported during the sixth and final option year of the BISC II contract.

Table 4-12. Number of Concurrent User Requirements

ID	Requirements	Priority
NCU_1	Based on assumptions in Section 1.3, the ImmPort system version 1.0 shall support 10 to 30 concurrent users.	1
NCU_2	Based on assumptions in section 1.3, at completion of the six-year period of performance of the BISC contract, the ImmPort system shall support 100 to 200 concurrent users.	4

Table 4-12 summarizes the ImmPort system concurrent user requirements.

4.7 SCALABILITY

The ImmPort system will use scalable technology to support an increase in users over the course of the BISC II contract. Table 4-13 lists ImmPort system scalability requirements.

Table 4-13. Scalability Requirements

ID	Requirements	Priority
SCA_1	The system shall be able to accommodate the anticipated growth of the ImmPort user community to 100 to 200 concurrent users and data storage requirements of up to 1000 terabytes over the six-year contract period.	3
SCA_2	In order to accommodate this growth, storage and processing capabilities shall be scalable in the following areas: Additional processing Additional memory Additional disk storage Additional database instances, if required Increased network bandwidth capacity	2
SCA_3	When changes are made in any of these areas, for any reason, the ImmPort database shall not encounter any significant system downtime (i.e., downtime more than 2 business days).	2

Table 4-13 summarizes the ImmPort system scalability requirements.

4.8 PERFORMANCE

It is assumed that a small number of concurrent users will access ImmPort in the first year. Performance statistics will help the BISC Team adjust system capabilities to meet the growing number of users after the first year. Table 4-14 lists ImmPort system performance requirements.

Table 4-14. Performance Requirements

ID	Requirements	Priority
PER_1	For purposes of initial system performance requirements, the system shall provide acceptable levels of performance when accessed over a broadband Internet connection with initial system load of 10 to 30 concurrent users, based on a 16-hour per day average usage.	1
PER_2	Exclusive of data transmission time, 95 percent of screens, excluding reports and queries, shall return results within 5 seconds.	2
PER_3	Based on the assumptions in Section 1.3, over the contract period, overall system performance shall be maintained during the incremental growth of users to a maximum of 100 to 200 total concurrent users.	4
PER_4	The system shall capture information on system usage, data utilization, and performance statistics.	2
PER_5	The system shall be able to journal user activities, including session start and end times, type of actions performed, query types executed, and resources used by each query.	2
PER_6	The system shall generate reports on system usage, data utilization, and performance statistics upon request.	3

Table 4-14 summarizes the ImmPort system performance requirements.

5.0 OVERVIEW OF BISC PHASE II FUNCTIONAL REQUIREMENTS

The following table lists the types of functional requirements that will be implemented in the BISC Phase II ImmPort system and their respective sections in the SSRD.

Table 5-1. Summary of ImmPort System Functional Requirement Types

Requirement Type	Section Reference
Data repository and sharing	6.0
Public data warehouse	6.1
Private project workspace	6.2
Collaborative workspace	6.3
Data defining DAIT-funded contracts, grants, and programs	7.1
Data defining a research project and its component experiment groups	7.2
Data associated with the subjects and biological samples used by a research project	7.3
Data defining a research experiment	7.4
Biochemical assay data	7.5
Data related to analysis of genetic associations	7.6
Reference data from public databases	8.0
Experimental data submissions	9.0
Data query and extraction	10.1
Data analysis	10.2
Ontology	11.0
Semantic data mapping	12.0
User assistance requirements	13.0

Table 5-1 summarizes the types of user requirements to be implemented by the ImmPort system.

6.0 DATA REPOSITORY AND SHARING REQUIREMENTS

6.1 PUBLIC DATA WAREHOUSE

The ImmPort system will provide a public data warehouse that will contain raw, minimally processed, and processed experimental data as well as reference data extracted from public sources. These data will be available for all ImmPort system users to view and use in queries and analyses. The public data warehouse will include near-line storage for data that are infrequently used and/or that do not need to be instantly accessible. Table 6-1 lists requirements for a public data warehouse.

Refer to Section 3.3 Product Priorities for a description of the requirement prioritization numbering scheme used in this section.

Table 6-1. Public Data Warehouse Requirements

ID	Requirements	Priority
PDW_1	The public data warehouse shall store minimally processed experimental data. These experimental data are identified in Section 7. Collection, Storage, and Management of Data Associated with DAIT-Funded Experimental Studies.	1
PDW_2	The public data warehouse shall store reference data from public sources. These reference data and their sources are identified in Section 8. Reference Data.	1
PDW_3	The public data warehouse shall organize the publicly available experimental data and reference data into data categories (e.g., proteins, genes, gene expression) to facilitate end user access. The categories of reference data are identified in Section 8. Reference Data.	1
PDW_4	The system shall require that data submitted to ImmPort meet an acceptable technical standard (e.g., file format, structural format, etc.) to ensure the data can be integrated in the public data warehouse.	1
PDW_5	The reference data in the public data warehouse shall be available for all system users to view. User access privileges to the public data beyond read-only are addressed in Section 4.4. Security.	1
PDW_6	The public data warehouse shall have the capacity to store 4 terabytes of interactively retrievable data for ImmPort v.1.0. (This minimum requirement will be revisited after initial release to ensure optimized performance as new contracts/grants beyond the Population Genetics projects and HLA projects begin to use the system.)	1
PDW_7	The public data warehouse shall provide near-line storage for data files containing experimental data that do not require real-time access.	3
PDW_8	The system shall retrieve a data file up to one gigabyte in size from near-line storage within one hour of the user's request. (This requirement does not apply to archived data.)	3
PDW_9	The system shall have the capability to direct a data file exceeding a parameter-driven size limit to near-line storage.	3
PDW_10	The system shall enable data extracted from public databases to be refreshed periodically with new or modified values.	1
PDW_11	The system shall enable data in the public data warehouse that originated from work performed by a research contract/grant to be updated or corrected with accompanying comments on the reason for change by an authorized user. User access privileges to the public data beyond read-only are addressed in Section 4.4. Security.	2
PDW_12	Data versioning shall occur when public data are refreshed.	1
PDW_13	Data versioning shall occur when public data are corrected.	1
PDW_14	When an experimental data set submitted by a contract/grant is corrected, the incorrect version of the data set shall be flagged as incorrect.	2
PDW_15	When an authorized user corrects a data set in the public data warehouse previously submitted by a contract/grant with which the user is associated, the system shall require the user to provide a textual explanation of why the data set was corrected.	2
PDW_16	When an experimental data set submitted by a contract/grant is corrected, the system shall not permit the use of the incorrect version of that data set.	2
PDW_17	When a data set is flagged as being incorrect, a pointer to the location of the correct version of that data set shall be maintained.	2
PDW_18	The system shall provide a notification available to all users of a correction to experimental data in	2

the public data warehouse.

Table 6-1 summarizes the ImmPort system requirements for a public data warehouse.

6.2 PRIVATE PROJECT WORKSPACES

The requirements associated with private project workspaces (PPWs) represent an attempt to reflect the real-world functioning of the scientific community, providing maximum flexibility using the simplest possible requirements. Before enumerating the specific requirements associated with PPWs, some data-related concepts are defined in the following paragraphs.

A “research project” is a concept in which a series of experiments are performed and analyzed by a group of investigators to address a specific scientific question or to test a specific scientific hypothesis. A research project is directed by a principal investigator. Preliminary, pre-publication data generated by a research project can be stored in a PPW in ImmPort, allowing the members of the research project to use the tools and public data provided by the ImmPort system to analyze their data. When the experimental data is ready for publication, it can then be released to the public data warehouse.

An “experiment group” is a concept in which a series of individual experiments are grouped together because they utilize the same constant variables, responding variables, experimental approaches and experimental properties, but differ based on defined conditions/manipulated variables. These attributes of an experiment, in combination with data defining the experimental sample and the sources of the sample, are referred to as the metadata defining the experiment.

An “experiment” includes the metadata defining the experimental procedure and the samples used, a primary result data set, and one or more secondary result data sets, which are defined below. An experiment group consists of one or more experiments. Metadata, with the exception of the variable that is different for each experiment in the experiment group, are associated with each individual experiment, but metadata common to all experiments are rolled-up to the experiment group level. The work products of a research project consist of many experiments and experiment groups.

A “primary data set” reflects the minimum information about a specific experimental sample and includes the primary results from an experimental evaluation of a specific sample (raw data), the experimental design, and the experimental setup. A primary result file normally contains data that can be interpreted without need for proprietary tools or software.

A “secondary data set” reflects the processing and analysis of one or more primary data sets and includes the processed data combined with the metadata that describes essential details about the analytical processing approaches, algorithms, and variables used. Secondary data sets may also be the product of processing one or more secondary data sets.

PPWs will enable DAIT-funded researchers to selectively share prepublication experimental data with other DAIT-funded researchers. Access control requirements related to these secure workspaces are addressed under Section 4.4. Table 6-2 lists requirements for PPWs.

Table 6-2. Private Project Workspace Requirements

ID	Requirements	Priority
PPW_1	The system shall provide a PPW into which the user can deposit experimental data results.	1
PPW_2	The user must be a PI to establish a PPW.	1
PPW_3	The system shall require that the PI establishing a PPW must first accept the terms and conditions for using the PPW feature.	1
PPW_4	The system shall grant PPW administrator rights to the PI establishing the workspace.	1

PPW_5	The contents of this workspace shall be inaccessible to any user who is not authorized by either the PI establishing the workspace or the PPW administrator assigned by the PI.	1
PPW_6	The PI shall have the capability to create more than one PPW. (No maximum number of workspaces is defined at this time; however, a maximum limit on the workspaces that a user can initiate may be needed in the future to conserve system resources.)	1
PPW_7	The system shall allow the execution of a query using the data in the PPW along with any data in the public data warehouse.	2
PPW_8	The system shall allow the execution of analyses using the data in the PPW along with any data in the public data warehouse.	2
PPW_9	The system shall enable an authorized user to save in the PPW a data set resulting from execution of an analysis or query operation.	1
PPW_10	The system shall allow the PI to authorize specific functions to users given access to a PPW, including the ability to load data sets into the PPW, edit and delete data in the PPW, and promote data sets from the PPW to the public data warehouse (see Overview of ImmPort Capabilities and Permissions for each User Role).	2
PPW_11	If the owner of a PPW (i.e., the PI who established the PPW) is scheduled to lose access to ImmPort, then the system shall give that user advance notice (2 weeks) that all data within a PPW will be subject to deletion from the system 2 weeks after the user loses access to the system.	2
PPW_12	If there is a personnel change at the PI-level for a research project, the system shall permit the transfer of PPW ownership to the new PI. In this situation, the data stored in the PPW will not be deleted if the original PI loses access to the system.	2
PPW_13	The authorized user shall have the capability to publish data from the PPW to the public data warehouse (i.e., transfer data from the PPW to the public data warehouse.)	1
PPW_14	The system shall persist data sets that a user deletes from his PPW.	2

Table 6-2 summarizes the PPW requirements.

Figure 6-1 illustrates two separate research projects controlled by two separate principal investigators (PI-1 and PI-2). Each research project contains data from three experiment groups. The data generated by research project 1 is housed in PPW 1, while data generated by research project 2 is housed in PPW 2. The contents of PPW-1 cannot be viewed or analyzed by PI-2 and the contents of PPW-2 cannot be viewed or analyzed by PI-1. PI-1 and PI-2 can create Collaborative Workspace A (CW-A) as described in Section 6.3 below, and deposit experiment groups 2 and 8 into CW-A. Both PI-1 and PI-2 can then view and analyze the data in both of these experiment groups.

Figure 6-1. Private and Collaborative Workspaces within ImmPort

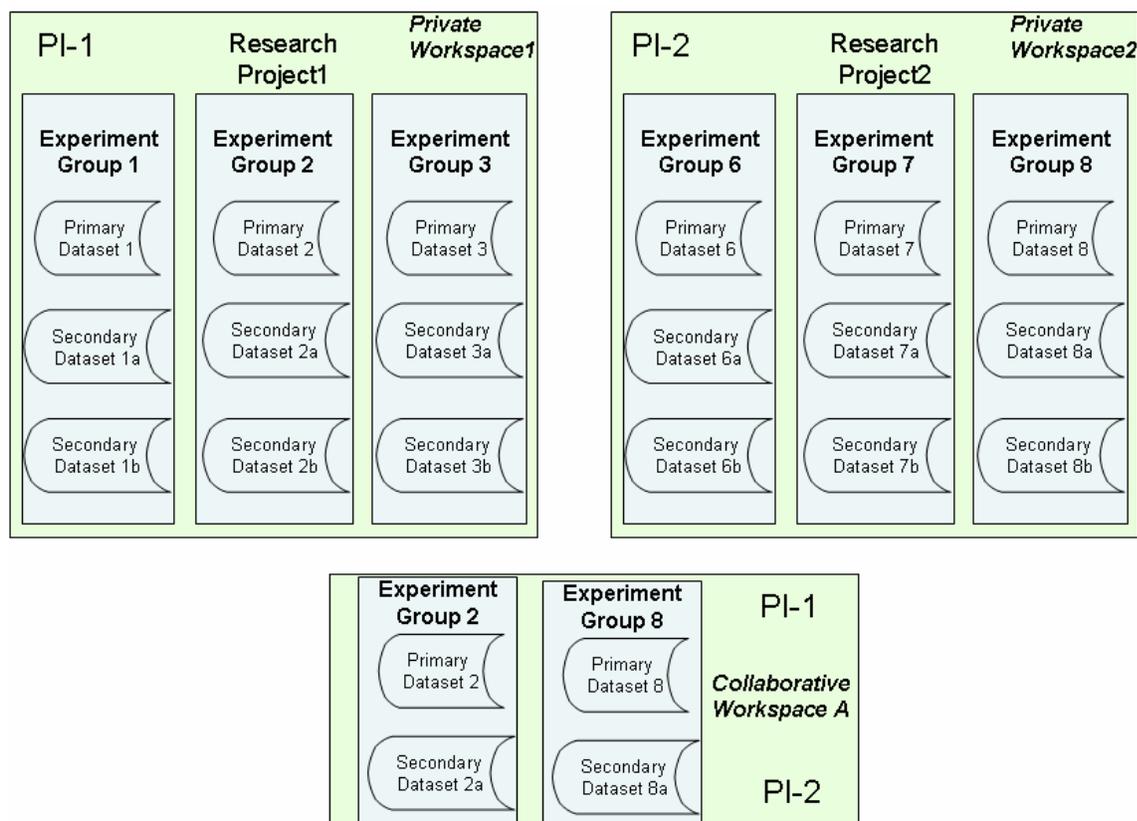


Figure 6-1 illustrates how a PPW houses the data of a research project and the assignment of experiment groups by research projects to a CW.

6.3 COLLABORATIVE WORKSPACES

NIAID/DAIT and DAIT-funded researchers have expressed interest in increased collaboration among researchers. The ImmPort system will support an environment in which researchers can share their pre-publication data with one another. See Figure 6-1 and the associated discussion in Section 6.2 for an illustration of the use of a CW. Table 6-3 lists requirements for CWs.

Table 6-3. Collaborative Workspace Requirements

ID	Requirements	Priority
CWR_1	The system shall enable two or more users working on two or more different research projects to share specific experiment groups from their respective research projects.	2
CWR_2	The system shall enable two or more users working on two or more different research projects to share specific experiments from their respective research projects.	3
CWR_3	Only a user who is designated as a principal investigator (PI) on a DAIT contract/grant shall be authorized to establish a CW.	2
CWR_4	The system shall require that the PI establishing a CW must first accept the terms and conditions for using the CW feature.	2
CWR_5	The system shall allow only the PI who establishes a CW to authorize other users access to the CW. (A user with access to the CW is referred to as a "collaborator.")	2
CWR_6	The system shall enable the PI who established the CW to organize the contents of the workspace according to experiment groups, experiments, primary data sets, and secondary data sets.	2
CWR_7	All the contents of the CW shall be accessible to all collaborators.	2
CWR_8	The system shall prevent a user who is not a collaborator from accessing the CW.	2

CWR_9	The collaborator shall have the ability to grant access via the CW to any experiment group contained in any PPW to which the collaborator has access (a collaborator can "deposit" an experiment group into a CW).	2
CWR_10	The collaborator shall have the ability to grant access via the CW to an experiment contained in any PPW to which the collaborator has access (a collaborator can "deposit" an experiment into a CW).	3
CWR_11	The collaborator shall be able to perform queries on one or more data sets that are accessible via the CW.	2
CWR_12	The collaborator shall be able to perform analyses on one or more data sets that are accessible via the CW.	2
CWR_13	The collaborator shall be able to perform queries on one or more data sets that are accessible via the CW in combination with one or more data sets in the public data warehouse.	3
CWR_14	The collaborator shall be able to perform analyses on one or more data sets that are accessible via the CW in combination with one or more data sets in the public data warehouse.	3
CWR_15	The collaborator shall be able to perform queries on one or more data sets that are accessible via the CW in combination with one or more data sets in a PPW to which the collaborator has access.	3
CWR_16	The collaborator shall be able to perform analyses on one or more data sets that are accessible via the CW in combination with one or more data sets in a PPW to which the collaborator has access.	3
CWR_17	The system shall permit the collaborator to save results from execution of an analysis or query operation against data in a CW to the collaborator's PPW.	2
CWR_18	When a result data set is saved after running an operation on data in a CW, that result data set shall be made available to all collaborators via the CW.	2
CWR_19	If the owner of a CW (i.e., the PI who established the CW) is scheduled to lose access to ImmPort, then the system shall give that user advance notice (2 weeks) the CW will be subject to deletion from the system 2 weeks after the user loses access to the system and all assignments of data to the CW will be eliminated.	4
CWR_20	If there is a personnel change at the PI-level for a research project, the system shall permit the transfer of CW ownership to the new PI. In this situation, the data stored in the CW will not be deleted if the original PI loses access to the system.	4

Table 6-3 summarizes the CW requirements.

7.0 COLLECTION, STORAGE, AND MANAGEMENT OF DATA ASSOCIATED WITH DAIT-FUNDED EXPERIMENTAL STUDIES

This section includes tables listing data requirements for NIAID/DAIT-funded programs, contracts, and grants, and the experimental studies performed under these contracts and grants. Where lists of values appear in these tables in association with a data field, the lists may not be all inclusive, but are provided to illustrate the type of data that a data field will contain.

Refer to Section 3.3 Product Priorities for a description of the requirement prioritization numbering scheme used in this section.

7.1 DAIT-FUNDED CONTRACTS/GRANTS AND PROGRAMS

This subsection addresses requirements for DAIT-funded contracts/grants, and programs.

Table 7-1. DAIT-Funded Contract/Grant Information

ID	Requirement	Priority
DCG_1	The system shall store information on DAIT-funded contracts/grants. See Table 7-2. DAIT-Funded Contract/Grant Information.	1
DCG_2	The system shall store information on DAIT programs. See Table 7-3. DAIT Program Information.	1
DCG_3	The system shall enable the user to associate one or more DAIT-funded contract/grant with a program.	1

Table 7-1 identifies DAIT-funded contract/grant information to be stored in ImmPort.

Table 7-2 identifies data requirements for DAIT-funded contracts/grants.

Table 7-2. DAIT-Funded Contract/Grant Information

Type of Information	ID	Data Field
Contract/Grant	DAC_1	*Title
Contract/Grant	DAC_2	*Short title
Contract/Grant	DAC_3	*Contract/grant number
Contract/Grant	DAC_4	*Abstract (brief description of funded activities)
Contract/Grant	DAC_5	*Keywords
Contract/Grant	DAC_6	*Start date
Contract/Grant	DAC_7	*End date
Contract/Grant	DAC_8	ImmPort Access End Date
Contract/Grant	DAC_9	*Principal Investigator
DAIT	DAC_10	*Project Officer
DAIT	DAC_11	Deputy Project Officer
DAIT	DAC_12	Program (if applicable)

Table 7-2 identifies information that ImmPort will store for each DAIT-funded project. An asterisk (**) indicates a required field.

Table 7-3 identifies data requirements for DAIT programs.

Table 7-3. DAIT Program Information

Type of Information	ID	Data Field
Program	DPI_1	*Title
Program	DPI_2	*Short title
Program	DPI_3	*Description or abstract (description of funded activities)
Program	DPI_4	*Start date
Program	DPI_5	*End date
DAIT	DPI_6	*Project Officer
DAIT	DPI_7	Deputy Project Officer
Funded Contract/Grants	DPI_8	*Contract/grant number(s)

Table 7-3 identifies information that ImmPort will store for each DAIT program. An asterisk (**) indicates a required field.

7.2 RESEARCH PROJECTS AND COMPONENT EXPERIMENT GROUPS

Information on experimental data will be captured at distinct levels. These levels, presented from general to specific, are the research project level, the experiment group level, and the experiment level. Although data elements may be captured at only one of these distinct levels, the system will roll up specific data captured at a lower level for the user's view at higher levels. This subsection describes the data requirements at the research project and experiment group levels. Subsections 7.3 and 7.4 describe data requirements at the experiment level.

Table 7-4. Research Projects and Component Experiment Groups

ID	Requirements	Priority
RPC_1	The system shall store information about a research project. See Table 7-5. Research Project Information.	1
RPC_2	The system shall enable the user to group individual experiments into an experiment group.	1
RPC_3	The system shall store information about an experiment group. See Table 7-6. Experiment Group Information.	1
RPC_4	Linking an experiment to an experiment group shall be optional.	1
RPC_5	An experiment shall be able to belong to more than one experiment group at the same time.	1
RPC_6	The user shall be able to designate the variable types for the experiment group's constant variables.	1
RPC_7	The user shall be able to designate the variable types for the experiment group's conditional/manipulated variables.	1
RPC_8	The user shall be able to designate the variable types for the experiment group's responding variables.	1
RPC_9	The system shall capture the data and statistical approaches used for sampling size determination used in the selection of subjects for a research project or an experiment group within a research project .	2

Table 7-4 identifies requirements related to research projects and component experiment groups.

Table 7-5 identifies data requirements for research projects.

Table 7-5. Research Project Information

Type of Information	ID	Data Field
Research Project	RPI_1	*Project title
Research Project	RPI_2	*Goal (short text description)
Research Project	RPI_3	*Description (Abstract type text defining goals and approach)
Research Project	RPI_4	*Keywords
Personnel	RPI_5	*Principal Investigator
Personnel (roll-up)	RPI_6	Collaborator(s)

Table 7-5 identifies overall research study information that the ImmPort system will store for each research project. An asterisk (**) indicates a required field.

Table 7-6. identifies data requirements for experiment groups.

Table 7-6. Experiment Group Information

Data Type	ID	Data Item
Experiment Group	EGI_1	*Goal (short text description)
Experiment Group	EGI_2	*Description (Abstract type text defining research approach)
Experiment Group	EGI_3	*Keywords
Experiment Group Constant Variables	EGI_4	<ul style="list-style-type: none"> ■ Variables Type (the following are options) ■ Sample ■ Subject ■ Reagent ■ Dose ■ Time ■ Temperature ■ Radiation ■ Other (free text)
Experiment Group Conditional/Manipulated Variables	EGI_5	<ul style="list-style-type: none"> ■ Variables Type (the following are options) ■ Sample ■ Subject ■ Reagent ■ Dose ■ Time ■ Temperature ■ Radiation ■ Other (free text)
Experiment Group Responding Variables	EGI_6	Responding variable (3 fields to populate) <ul style="list-style-type: none"> ■ Concept (free text, e.g., Tolerance, Immune response, Proliferation) ■ Analyte (free text, e.g., TNF alpha protein, IFN gamma mRNA, SNP rs12345) ■ Measurement type (free text, e.g., Absorbance, Fluorescence intensity, CPM)
Experiment group design (roll-up of individual experiments in the group)	EGI_7	Experimental technique (platform) types (one entry for each technique [or platform] used by an experiment in the group)
Experiments included (roll-up)	EGI_8	Experiment ID (one entry for each experiment in the group)

Table 7-6 identifies information that the ImmPort system will store for each experiment group within a research project. An experiment group is a series of experiments that are grouped together by common experimental features including experimental technique/platform and experimental protocol. An asterisk (**) indicates a required field.

7.3 SUBJECT AND SAMPLE INFORMATION

In this subsection, software requirements on subject and sample information are defined and expanded on with corresponding tables identifying data requirements. Tables 7-7 through 7-18 represent a compilation of subject data that the system will need to store.

Human subject information includes data on the following:

- Demographics
- Other subjects in the same study who are biologically related
- Personal medical history

- Family medical history
- Laboratory test results
- Clinical findings
- Treatment/therapeutics
- Composite diagnostics evaluation information

A limited subset of the data for human subjects is collected for nonhuman subjects (see Table 7-16).

Sample information includes data on the following:

- Biological sample data
- Biological/medical tests and assays

7.3.1 Subject Information

Table 7-7 identifies requirements related to subject information.

Table 7-7. Subject Information Requirements

ID	Requirements	Priority
SSI_1	The system shall be able to store demographic data for each study participant.	1
SSI_2	The system shall be able to store data on biologically related study participants.	1
SSI_3	The system shall be able to store personal medical history data for each study participant.	1
SSI_4	The system shall be able to store family medical history data for each study participant.	1
SSI_5	If pedigree information is provided in the form of the medical history of family members of the study participant, and these family members are not part of the study, then the system shall be able to store data for each medical history event recorded for each relative of a study participant.	1
SSI_6	The system shall be able to store laboratory results data for each study participant.	1
SSI_7	The system shall be able to store clinical findings data for each study participant.	1
SSI_8	The system shall be able to store treatment/therapeutics data for each study participant. Both therapeutic drugs taken previously or currently being taken and past surgical procedures are included in these categories.	1
SSI_9	The system shall be able to store diagnostics evaluation data for each study participant.	2
SSI_10	The user shall be able to query for all research projects in which a subject has participated, assuming the same subject ID is used for the subject across research projects.	2
SSI_11	The system shall capture inclusion/exclusion criteria used to select subjects for experiments under a research project or an experiment group within a research project.	2
SSI_12	The system shall be able to store data on non-human subjects.	1
SSI_13	The system shall store subject data in a manner that will facilitate identification of changes and trends in the data over time.	1

Table 7-7 identifies subject information that ImmPort will store.

Table 7-8 identifies data requirements for demographic information.

Table 7-8. Demographic Information

Type of Information	ID	Data Field
Participant (Subject)	DEI_1	*Unique identifier
Contributor of data	DEI_2	Date entered
Contributor of data	DEI_3	Contributor name
Demographic	DEI_4	Date of enrollment in study

Demographic	DEI_5	Birth year (or birth year range)
Demographic	DEI_7	Year (or year range) of death
Demographic	DEI_8	Gender
Demographic	DEI_9	Ethnicity
Demographic	DEI_10	Race
Demographic	DEI_11	Age at enrollment (calculated)
Demographic	DEI_12	Residential location (5 fields to populate) <ul style="list-style-type: none"> ■ City or town ■ County ■ State or province ■ Country ■ Region (e.g., West Africa, North America)
Other	DEI_13	Other data to be provided by Investigator
Other	DEI_14	Marital status

Table 7-8 identifies demographic information that ImmPort shall store for each study participant. An asterisk (*) indicates a required field.

Table 7-9 identifies data requirements for information on other study participants related to the current participant.

Table 7-9. Related Subject Information

Type of Information	ID	Data Field
Participant (Subject)	RSI_1	*Unique identifier
Contributor of data	RSI_2	Date entered
Contributor of data	RSI_3	Contributor name
Contributor of data	RSI_4	*Unique identifier (Required if related subject information is provided)
Contributor of data	RSI_5	<ul style="list-style-type: none"> ■ *Related Subject Relationship (the following are options) ■ Mother ■ Father ■ Son ■ Daughter ■ Brother ■ Sister ■ Maternal Grandmother ■ Maternal Grandfather ■ Paternal Grandmother ■ Paternal Grandfather ■ Maternal Aunt ■ Maternal Uncle ■ Paternal Aunt ■ Paternal Uncle ■ Maternal Female Cousin ■ Maternal Male Cousin ■ Paternal Female Cousin ■ Paternal Male Cousin ■ Niece ■ Nephew

Table 7-9 identifies related participant or subject information that ImmPort shall store for each study participant. An asterisk (*) indicates a required field when related subject data is provided.

The personal history information may be of any of the information types identified in the “Record type” data field in Table 7-10. For each record type, the information identified in the second column with that record type in the first column will be stored by ImmPort. Though it is possible for there to be multiple occurrences of each record type, the “Other diagnosis” and “Immunization” record types, which record chronic and infectious disease history and immunization history respectively, will normally each have multiple occurrences. If a participant is part of multiple studies and the same unique identifier is used, many personal history records, with different values in the date entered field, may exist.

Table 7-10. Personal Medical History Information

Type of Information	ID	Data Field
Participant (Subject)	PMH_1	*Unique Identifier
Contributor of data	PMH_2	Date entered
Contributor of data	PMH_3	Contributor name
Information Type	PMH_4	<ul style="list-style-type: none"> ■ *Record Type (the following are options) ■ Entry Diagnosis ■ Other Diagnosis ■ Pregnancy ■ Breast Feeding ■ Menstrual history ■ Immunization ■ Tobacco use ■ Alcohol use ■ Illicit drug use ■ Employment ■ Education ■ Income ■ Other
Entry Diagnosis and Other Diagnosis	PMH_5	Diagnostic term
Entry Diagnosis and Other Diagnosis	PMH_6	ICD-9 code
Entry Diagnosis and Other Diagnosis	PMH_7	Year of onset
Entry Diagnosis and Other Diagnosis	PMH_8	Verification (the following are options) Subject only Medical record
Pregnancy	PMH_9	Pregnancy outcome
Breast feeding	PMH_10	Breast feeding duration
Menstrual History	PMH_11	Age at menarche
Menstrual History	PMH_12	Age at menopause
Immunization History	PMH_13	Vaccine Type (The following list is not all inclusive) <ul style="list-style-type: none"> ■ Dryvax vaccinia vaccine for Smallpox ■ Anthrax Vaccine Adsorbed (AVA) ■ Typherix vaccine -- Salmonella Typhi ■ Bivalent oral killed Cholera vaccine -- Vibrio Cholera
Immunization History	PMH_14	Date
Immunization History	PMH_15	Vaccination route (the following are options for the values) Oral Intramuscular

		Skin
Immunization History	PMH_16	Dose number (in series of immunizations)
Tobacco Use	PMH_17	<ul style="list-style-type: none"> ■ Tobacco amount (the following are options) ■ Never ■ <10 packs per year ■ 10-30 packs per year ■ >30 packs per year
Alcohol use	PMH_18	<ul style="list-style-type: none"> ■ Alcohol amount (the following are options) ■ none ■ <1std drink per day ■ 2-5 std drinks per day ■ >5 std drinks per day
Illicit drug use	PMH_19	<ul style="list-style-type: none"> ■ Drug (the following are options) ■ None ■ Marijuana ■ Cocaine ■ Heroin ■ Methamphetamine ■ Other
Employment	PMH_20	<ul style="list-style-type: none"> ■ Job Class (the following are options) ■ Administrative ■ Technical ■ Manual labor ■ Professional ■ Unemployed
Employment	PMH_21	Number of years
Employment	PMH_22	Year last employed
Education	PMH_23	Number of years
Income	PMH_24	<ul style="list-style-type: none"> ■ Income group (the following are options) ■ <10000 per year ■ 10,000-19,999 per year ■ 20,000-29,999 per year ■ 30,000-39,999 per year ■ 40,000-49,999 per year ■ 50,000-74,999 per year ■ >75,000 per year
Other Personal History Data	PMH_25	Other data type
Other Personal History Data	PMH_26	Other data value

Table 7-10 identifies personal medical history information that ImmPort shall store for each study participant. An asterisk (*) indicates a required field when personal medical history data is provided.

Table 7-11. Family Medical History

Type of Information	ID	Data Field
Participant (Subject)	FMH_1	*Unique identifier
Contributor of data	FMH_2	Date entered
Contributor of data	FMH_3	Contributor name
Relationship	FMH_4	<ul style="list-style-type: none"> ■ *Relationship (the following are options) ■ Mother ■ Father

		<ul style="list-style-type: none"> ■ Son ■ Daughter ■ Brother ■ Sister ■ Maternal Grandmother ■ Maternal Grandfather ■ Paternal Grandmother ■ Paternal Grandfather ■ Maternal Aunt ■ Maternal Uncle ■ Paternal Aunt ■ Paternal Uncle ■ Maternal Female Cousin ■ Maternal Male Cousin ■ Paternal Female Cousin ■ Paternal Male Cousin ■ Niece ■ Nephew
Deceased	FMH_5	*Yes or no
Age at death	FMH_6	Age at death
Medical history event	FMH_7	*Diagnostic Term
Medical history event	FMH_8	*ICD-9 code
Medical history event	FMH_9	Year of onset
Medical history event	FMH_10	Verification (the following are options) Subject only Medical record

Table 7-11 identifies family medical history information that ImmPort shall store for each study participant. An asterisk (**) indicates a required field when family medical history data is provided.

Table 7-12. Laboratory Test Results

Type of Information	ID	Data Field
Source	LTR_1	*Unique identifier
Contributor of data	LTR_2	Date entered
Contributor of data	LTR_3	Contributor name
Laboratory Test	LTR_4	<ul style="list-style-type: none"> ■ *Test Name (the following list is not all inclusive) ■ Urinary casts ■ Proteinuria ■ Hematuria ■ HLA genotype ■ Thrombocyte ■ Leukocyte ■ Lymphocyte ■ Anti-nuclear Ab ■ Anti-DNA Ab ■ Anti-Ro/SSA Ab ■ Anti-La/SSB Ab ■ Anti-U1RNP Ab ■ Anti-SM Ab ■ Anti-Scl-70/Topoisomerase Ab ■ Anti-centromere Ab

		<ul style="list-style-type: none"> ■ Anti-Jo-1/Mi-2/SNP Ab ■ Complement C3 ■ Complement C4 ■ Creatinine ■ BUN ■ AST ■ ALT ■ C-reactive protein ■ Uric acid ■ Hepatitis B core IgM ■ Hepatitis B surface Ag ■ Hepatitis B surface Ab ■ Hepatitis C ■ White blood cell count ■ PMN count ■ PMN percentage ■ Lymphocyte count ■ Lymphocyte percentage ■ Monocyte count ■ Monocyte percentage ■ Hemoglobin ■ Platelet count ■ Erythrocyte sedimentation rate
Laboratory Test	LTR_5	Assay description
Laboratory Test	LTR_6	*Date of Test
Laboratory Test	LTR_7	*Value
Laboratory Test	LTR_8	Normal lower limit or value
Laboratory Test	LTR_9	Normal upper limit

Table 7-12 identifies laboratory test result information that ImmPort shall store for each study participant. An asterisk (“*”) indicates a required field when laboratory test result data is submitted.

Table 7-13. Clinical Exam Findings

Type of Information	ID	Data Field
Participant (Subject)	CLF_1	*Unique identifier
Contributor of data	CLF_2	Date entered
Contributor of data	CLF_3	Contributor name
Physical finding	CLF_4	<ul style="list-style-type: none"> ■ *Description (the following list is not all inclusive) ■ Malar rash ■ Photosensitive rash ■ Subacute Cutaneous LE ■ Discoid LE ■ Oral or genital ulcers ■ Livedo reticularis ■ Non-erosive peripheral synovitis ■ Vasculitis ■ Raynaud's Phenomenon ■ Serositis ■ Glomerulonephritis ■ Seizures ■ Psychosis

		<ul style="list-style-type: none"> ■ Organic Brain Syndrome ■ Visual impairment ■ Cranial neuropathy ■ Lupus headache ■ CVA ■ Arthritis ■ Myositis ■ Alopecia ■ Pleurisy ■ Fever ■ Morning Stiffness ■ Limited joints ■ Swollen joint ■ Erosive peripheral synovitis ■ Psoriasiform rash ■ Keratoconjunctivitis Sicca ■ Nail Pitting ■ Rheumatoid nodules ■ Sacroillitis ■ Tender joint count
Physical finding	CLF_5	*Presence (the following are options) Positive Negative
Physical finding	CLF_6	Value (if quantitative)
Physical finding	CLF_7	Normal value or lower range value
Physical finding	CLF_8	Normal upper range value
Physical finding	CLF_9	*Physical finding date
Physical finding	CLF_10	Comments

Table 7-13 identifies clinical findings information that ImmPort shall store for each study participant. An asterisk (**) indicates a required field when clinical findings data is submitted.

Table 7-14. Treatment/Therapeutics Information

Type of Information	ID	Data Field
Participant (Subject)	TTI_1	*Unique identifier
Contributor of data	TTI_2	Date entered
Contributor of data	TTI_3	Contributor name
Visit date	TTI_4	Visit date
Treatment/therapeutic	TTI_5	*Name
Treatment/therapeutic	TTI_6	Class
Treatment/therapeutic	TTI_7	Start date
Treatment/therapeutic	TTI_8	Stop date
Treatment/therapeutic	TTI_9	Dose
Treatment/therapeutic	TTI_10	Frequency
Treatment/therapeutic	TTI_11	Route
Treatment/therapeutic	TTI_12	Adverse event (the following are options) <ul style="list-style-type: none"> ■ Yes ■ No
Treatment/therapeutic	TTI_13	Adverse event type
Treatment/therapeutic	TTI_14	Adverse event date

Table 7-14 identifies treatment and therapeutics information that ImmPort shall store for each study participant. An asterisk (*) indicates a required field when treatment/therapeutics information is provided.

Table 7-15. Composite Diagnostic Evaluation Information

Type of Information	ID	Data Field
Participant (Subject)	CDE_1	*Unique identifier
Contributor of data	CDE_2	Date entered
Contributor of data	CDE_3	Contributor name
Composite diagnostic evaluations	CDE_4	*Name of Test
Composite diagnostic evaluations	CDE_5	*Date of Evaluation
Composite diagnostic evaluations	CDE_6	*Value
Composite diagnostic evaluations	CDE_7	Normal value or lower range
Composite diagnostic evaluations	CDE_8	Normal upper range

Table 7-15 identifies composite diagnostic evaluation information that ImmPort shall store for each study participant. An asterisk (*) indicates a required field when composite diagnostic evaluation data is submitted.

Table 7-16. Non-Human Subject Information

Type of Information	ID	Data Field
Contributor of data	NSI_2	Date entered
Contributor of data	NSI_3	Contributor name
Subject	NSI_1	*Unique identifier
Subject	NSI_4	*Species
Subject	NSI_5	*Strain
Subject	NSI_6	Age
Subject	NSI_7	Gender
Subject	NSI_8	Special characteristics of strain
Subject	NSI_9	<ul style="list-style-type: none"> ■ *Location of colony (4 fields to populate) ■ Institution ■ City or town ■ State or province ■ Country
Special Conditions	NSI_10	Infectious disease
Special Conditions	NSI_11	Chronic disease

Table 7-16 identifies information that ImmPort shall store for each non-human subject used by the project. An asterisk (*) indicates a required field.

7.3.2 Biological Sample Information

Each research experiment consists of a biochemical assay or test performed on one or more biological samples, each obtained from a biological source. The biological source may be a person, an animal such as a mouse or rat, or some other source like microorganisms or tissue cultures.

A primary biological sample collected from a subject (human and non-human) can be used for an experiment. A secondary sample is prepared by processing either a primary biological sample or another secondary biological sample to produce a biological sample that differs from the starting material. Examples are preparation of peripheral blood mononuclear cells (PBMCs) from blood or preparation of DNA from PBMCs.

Table 7-17. Biological Sample Information Requirements

ID	Requirements	Priority
SSI_14	The system shall be able to store data on biological samples.	1
SSI_15	The system shall be able to store data on sources of biological samples.	1
SSI_16	The system shall be able to link data on biological samples with their corresponding study subjects (human or nonhuman) whose information is also stored in the system.	1

Table 7-17 identifies biological sample information that ImmPort will store.

Table 7-18. Biological Sample Information

Type of Information	ID	Data Field
Biological Sample Identification	BSI_1	*Unique identifier
Contributor of data	BSI_2	Date entered
Contributor of data	BSI_3	Contributor name
Biological Sample Identification	BSI_4	*Sample Type (the following list is not all inclusive) <ul style="list-style-type: none"> ■ blood ■ CSF ■ urine ■ tissue ■ serum ■ plasma ■ peripheral blood mononuclear cells (PBMC) ■ sorted CD19+ cells ■ sorted CD4+ cells ■ sorted CD33+ cells ■ DNA ■ RNA ■ polyA+ RNA ■ total protein ■ soluble protein ■ cytoplasmic protein ■ membrane protein ■ phospholipids
Biological Sample Identification	BSI_5	*Source description/name (e.g. DNA extracted from blood in ACD from an anonymous patient with Type I diabetes)
Biological Sample Identification	BSI_6	<ul style="list-style-type: none"> ■ Source unique identifier ■ Source or subject unique identifier (primary sample) or ■ Sample unique identifier (secondary sample)
Biological Sample Identification	BSI_7	*Species
Biological Sample Identification	BSI_8	Date collected
Biological Sample Identification	BSI_9	Amount (weight or volume of sample in unit of measure)
Biological Sample Identification	BSI_10	Unit of measure (menu)
Biological Sample Identification	BSI_11	Number of aliquots
Biological Sample Identification	BSI_12	Storage location (5 fields to populate) <ul style="list-style-type: none"> ■ Location code ■ Institution ■ City or town ■ State or province ■ Country
Biological Sample Identification	BSI_13	Storage temperature
Biological Sample Identification	BSI_14	Other storage information (free text)

Biological Sample Processing	BSI_15	Processing procedure type
Biological Sample Processing	BSI_16	Processing procedure description
Biological Sample Description	BSI_17	*Sample description/name (free text)
Biological Sample Description	BSI_18	*Relevant sample phenotype

Table 7-18 identifies data that ImmPort shall store for each biological sample collected. An asterisk (**) indicates a required field.

7.4 INDIVIDUAL EXPERIMENTS

Each experiment is assigned to a research project and may be part of an experiment group.

Individual experiments are defined by the data types identified in Table 7-19 are generic in nature and can be used to store information for any type of experiment within ImmPort. The generic experiment components and data types specified in Table 7-19 are used in Section 7.5 to define required data types for individual biochemical assays.

The information specified in Table 7-19, along with subject and sample data linked to the unique identifier of the sample used for the experiment, comprise the metadata describing an experiment. The primary experimental result data file is the actual experimental result. Reports are generated by analysis of one or more primary or secondary experimental data files from one or more experiments and may be associated with an experiment, an experiment group, or research project. The ImmPort system will require that all experiments be associated with the grants or contracts under which they were funded.

The system will store the listed information about an experiment.

Table 7-19. Experiments

Experiment Component	ID	Data Types with Definitions
Experimental metadata	EXP_1	*Experiment title/Description (e.g. Evaluation of a series of SNPs in an African-American population)
Experimental metadata	EXP_2	*Experiment unique identifier (system generated)
Experimental metadata	EXP_3	*Contract/Grant Number [under which this experiment was performed]
Experimental metadata	EXP_4	Hypothesis
Experimental metadata	EXP_5	Rationale
Experimental metadata	EXP_6	Biological sample unique identifier [one entry for each sample used in experiment]
Experimental metadata	EXP_7	*Experimental technique and platform information (3 fields to populate) Experimental technique Experimental platform Experimental subplatform
Experimental metadata	EXP_8	Experimental Protocol (text file selected from a pull down menu of protocols that have been submitted to the PI's work space; or the submitter can choose to submit a new protocol by browsing for the files from their work stations)

Experimental metadata	EXP_9	<p>*Constant variables [one entry for each constant variable]</p> <ul style="list-style-type: none"> ■ Variable type (examples) <ul style="list-style-type: none"> ■ Sample ■ Subject ■ Reagent ■ Dose ■ Time ■ Temperature ■ Radiation ■ Other ■ Variable value (free text, corresponding to above example types) <ul style="list-style-type: none"> ■ Sample1232 (sample ID) ■ Patient xyz (patient ID) ■ IL4, cyclosporine (value for reagent) ■ 10mm, 23 mg/ml (value for dose) ■ 30 sec, 3 mo (value of time) ■ 37 degree C (value for temperature) ■ 200 rad (value for radiation) ■ Stress (value for other) ■ Variable unit of measure
Experimental metadata	EXP_10	<p>*Conditional/manipulated variables [one entry for each conditional/manipulated variable]</p> <ul style="list-style-type: none"> ■ Variable type (examples) <ul style="list-style-type: none"> ■ Sample ■ Subject ■ Reagent ■ Dose ■ Time ■ Temperature ■ Radiation ■ Other ■ Variable value (free text, corresponding to above example types) <ul style="list-style-type: none"> ■ Sample123 (sample ID) ■ Patient xyz (patient ID) ■ IL4, cyclosporine (value for reagent) ■ 10mm, 23 mg/ml (value for dose) ■ 30 sec, 3 mo (value of time) ■ 37 degree C (value for temperature) ■ 200 rad (value for radiation) ■ Stress (value for other) ■ Variable unit of measure
Experimental metadata	EXP_11	<p>*Responding variables (measured variables, result) [one entry for each responding variable]</p> <ul style="list-style-type: none"> ■ Concept (e.g., tolerance, immune response, proliferation) ■ Analyte (e.g., TNF alpha protein, IFN gamma mRNA, SNP rs12345) ■ Measurement type (e.g., absorbance, fluorescence intensity, CPM) ■ Measurement value (free text) and measurement value units of measure (when applicable)
Experimental metadata	EXP_12	<p>*Species (e.g. <i>Homo sapiens</i>)</p>

Personnel	EXP_13	*Submitter (auto generated from login)
Personnel	EXP_14	*PI (auto generated from the contract/grant number and research project)
Personnel	EXP_15	Experimenter(s) (Semi-colon delimited names of all the researchers involved, e.g. Smith, Sarah; Doe, John Q.)
Key results	EXP_16	Result value (one entry for each result value)
Upload experimental metadata file	EXP_17	<ul style="list-style-type: none"> ■ File format definition [Column definition (Column 1, Column 2, etc; the columns should include all variables and responding variables/results) ■ Unique header identifier (e.g. Sample_ID, ProbeSet_ID) ■ Data type (e.g. text, date, integer, float, etc.) ■ Definition (e.g. the unique Affymetrix identifier for a particular set of oligonucleotide probes)
Upload experimental metadata file	EXP_18	Upload file name [containing experimental metadata]
Result file (or report file)	EXP_19	File Name
Result file (or report file)	EXP_20	File type
Result file (or report file)	EXP_21	Data types included
Result file (or report file)	EXP_22	Data processing software used (e.g. MAS)
Result file (or report file)	EXP_23	Data processing software version (e.g. 5.0)
Result file (or report file)	EXP_24	Data analysis algorithm (e.g. Affymetrix image processing algorithm)
Result file (or report file)	EXP_25	Input result file name (If result file is not primary result file, include name of file that was processed by the data analysis algorithm to produce the current (secondary) result file), [One entry for each input result file processed to generate the current result file]

Table 7-19 contains a generic representation of data needed to define an experiment that must be stored by the ImmPort system. An asterisk (“*”) indicates a required field.

7.5 BIOCHEMICAL ASSAY DATA

This section identifies the experiment types, experimental techniques, and biochemical assays used by the Population Genetics Program, which are representative of all DAIT-funded immunology research. The ImmPort system will capture biochemical assay data obtained from various types of laboratory experiments performed for DAIT-funded research projects. Although all experiment types are defined by the generic data types specified in the previous section, each generic data type takes on a specific value for each experiment type.

Table 7-20. Biochemical Assay Data

ID	Requirement	Priority
BAD_1	The system shall support storage of experimental data associated with use of the Illumina Bead Array technique for SNP analysis and genotyping.	1
BAD_2	The system shall support storage of experimental data associated with the GeneChip-based genotyping (Affymetrix SNP chip) technique for SNP analysis and genotyping.	1
BAD_3	Expanded to BAD_3.1 and BAD_3.2.	1
BAD_3.1	The system shall support storage of experimental data associated with the microarray technique using Genpax for measuring mRNA expression.	2
BAD_3.2	The system shall support storage of experimental data associated with the microarray technique using Affymetrix GeneChip for measuring mRNA expression.	1
BAD_4	The system shall support storage of experimental data associated with the use of the microsatellite analysis technique for SNP analysis and genotyping.	2
BAD_5	The system shall support storage of experimental data associated with the use of HLA typing and other genotyping techniques for SNP analysis and genotyping.	2
BAD_6	The system shall support storage of experimental data associated with the use of the sequencing technique for SNP analysis and genotyping.	2
BAD_7	The system shall support storage of experimental data associated with the use of the flow cytometry technique for measuring protein expression.	2
BAD_8	The system shall support storage of experimental data associated with use of the ELISA technique for measuring immune responses.	2
BAD_9	The system shall support storage of experimental data associated with use of the ELISPOT technique for measuring immune response.	3
BAD_10	The system shall support storage of experimental data associated with use of the mass spectrometry technique for measuring protein expression.	3
BAD_11	The system shall support storage of experimental data associated with use of the Luminex Bead Assay for measuring protein function.	3
BAD_12	The system shall support the storage of experimental data associated with use of the cytometric bead immunoassay for measuring protein function.	3
BAD_13	The system shall support the storage of experimental data associated with use of the 2D-GEL technique for measuring protein expression.	4
BAD_14	The system shall support the storage of experimental data associated with use of the ESI-LC technique for measuring protein expression.	4
BAD_15	The system shall support the storage of experimental data associated with use of the PCR technique for measuring mRNA expression.	4
BAD_16	The system shall support the storage of experimental data associated with use of the CHIP technique for measuring DNA binding activity.	4
BAD_17	The system shall support the storage of experimental data associated with use of the EMSA technique for measuring DNA binding activity.	4
BAD_18	The system shall support the storage of experimental data associated with functional validation techniques using siRNA	4
BAD_19	The system shall support the storage of experimental data associated with use of transgenic or null mutant mice.	4
BAD_20	The system shall support the storage of experimental data associated with use of the Western Blot technique for quantifying the function of proteins and genes.	4

Table 7-20 lists biochemical assays classified as a combination of experiment type and experimental techniques, with priority, which will be supported by ImmPort.

7.5.1 Experimental Data for Genotyping Assays

The ImmPort System will support storage of experimental data associated with use of genotyping assays to perform high throughput genotyping of SNPs associated with candidate genes (see Table 7-21).

Table 7-21. Experimental Data for Genotyping Assays

ID	Experiment Component	Data Types with Definitions
EDG_1	Experimental metadata	*Experiment title/Description (e.g. Evaluation of a series of SNPs in an African-American population)
EDG_2	Experimental metadata	*Experiment unique identifier (system generated)
EDG_3	Experimental metadata	<ul style="list-style-type: none"> ▪ *Experimental platform information ▪ Illumina Bead array ▪ Array chip name/ID ▪ Affymetrix SNP chip ▪ Affymetrix chip name/ID
EDG_4	Experimental metadata	*Contract/Grant Number (under which this experiment was performed)
EDG_5	Experimental metadata	Hypothesis
EDG_6	Experimental metadata	Rationale
EDG_7	Experimental metadata	Experimental Protocol (text file selected from a pull down menu of protocols that have been submitted to the PI's work space; or the submitter can choose to submit a new protocol by browsing for the files from their work stations)
EDG_8	Experimental metadata	Unique Sample Identifier(s) (one Sample ID for each sample in the experiment, Please refer to Table 7-18 for details on sample data) (Mandatory only for Affymetrix Chip)
EDG_9	Experimental metadata	*Conditional/manipulated variables [one entry set for each conditional variable] (e.g. Source of genomic DNA) <ul style="list-style-type: none"> ■ Variable type ■ Variable value ■ Variable value unit of measure
EDG_10	Experimental metadata	*Constant variables [one entry set for each constant variable] (e.g. optical fiber bundle probe, Allele specific oligonucleotide (ASO), Locus specific oligonucleotide (LSO)) <ul style="list-style-type: none"> ■ Variable type ■ Variable value ■ Variable value unit of measure
EDG_11	Experimental metadata	*Responding variables (measured variables, result) [one entry for each responding variable] (e.g. Genotype) <ul style="list-style-type: none"> ■ Concept ■ Analyte ■ Measurement type ■ Measurement value (free text) ■ Measurement value units of measure (when applicable)
EDG_12	Experimental metadata	*Species (e.g. <i>Homo sapiens</i> , variable)
EDG_13	Personnel	*Submitter (auto generated from login)
EDG_14	Personnel	*PI (auto generated from the system)
EDG_15	Personnel	*Experimenter(s) (Semi-colon delimited names of all the researchers involved, e.g. Smith, Sarah; Doe, John Q.)

EDG_16	File format definition	<ul style="list-style-type: none"> ■ *Column definition (Column 1, Column 2, etc; the columns should include the conditional/manipulated variables and responding variables/results) ■ For Affymetrix SNP chip platforms (for each Sample ID) <ul style="list-style-type: none"> ■ *SNP_ID (Affy's internal name) ■ dbSNP RS ID (The dbSNP ID) ■ *Chromosome (Chromosome number) ■ *Physical location (Exact position on the chromosome) ■ TSC ID (the SNP consortium ID) ■ *Genotype call (Allele1_call, Allele2_call) ■ Confidence ■ Additional column(s) defined by user ■ Data type for the additional column(s) (e.g. text, date, numerical, etc.) ■ For Illumina platform <ul style="list-style-type: none"> ■ *Sample ID ■ *SNP name (Internal ID) ■ dbSNP name (dbSNP name) ■ *Chromosome (Chromosome number) ■ *Coordinate (Exact location of SNP on the chromosome) ■ Source (Source of data) ■ Species ■ *Genotype (allele 1 ,allele 2) ■ Confidence score (Quality of data) ■ Genome build ■ Additional column(s) defined by user ■ Data type for the additional column(s) (e.g. text, date, numerical, etc.) ■ For other genotyping platforms <ul style="list-style-type: none"> ■ *Sample ID ■ *SNP ID (Internal ID) ■ *Genotype (Allele A ,Allele B) ■ *Chromosome (Chromosome number) ■ *Physical location (Exact location of SNP on the chromosome) ■ Additional column(s) ■ Data type for the additional column(s) (e.g. text, date, numerical, etc.)
EDG_17	File format definition	*Upload file name
EDG_18	Data pre-processing	*Software used (e.g. GenCall and GenTrain)
EDG_19	Data pre-processing	*Version used (e.g. 1.0)
EDG_20	Data pre-processing	Analysis algorithm (e.g. Proprietary algorithm of the GenCall and GenTrain software)

Table 7-21 lists experimental data from Illumina Bead Array assay experiments for genotyping that must be stored by the ImmPort system.

7.5.2 Microarray for Gene Expression Analysis

The ImmPort System shall support storage of experimental data associated with use of the microarray assay for gene expression analysis and to evaluate unknown genes or genes with unknown function (used by the deCODE Genetics and McMaster Population Genetics Projects).

Table 7-22 lists data associated with use of the microarray assay.

Table 7-22. Experimental Data for Microarray Assay for Gene Expression

ID	Experiment Component	Data Types with Definitions
EDM_1	Experimental metadata	*Experiment title/Description (e.g. Characterization of B cell responses to 32 ligands)
EDM_2	Experimental metadata	*Experiment unique identifier (system generated)
EDM_3	Experimental metadata	*Contract/Grant Number (under which this experiment was performed)
EDM_4	Experimental metadata	Hypothesis
EDM_5	Experimental metadata	Rationale (e.g. To characterize differences in B cell responses to various stimuli)
EDM_6	Experimental metadata	Experimental Protocol (text file selected from a pull down menu of protocols that have been submitted to the PI's work space; or the submitter can choose to submit a new protocol by browsing for the files from their work stations)
EDM_7	Experimental metadata	*Experimental platform <ul style="list-style-type: none"> ■ Affymetrix array ■ Affymetrix chip name/ID ■ cDNA spotted array ■ Array chip name/ID ■ Custom array ■ Array chip name/ID
EDM_8	Experimental metadata	*Conditional/manipulated variables [one entry set for each conditional variable] (e.g. Sample source of RNA) <ul style="list-style-type: none"> ■ Variable type ■ Variable value ■ Variable value unit of measure
EDM_9	Experimental metadata	*Constant variables [one entry set for each constant variable] (e.g. Affymetrix U133A chip, Hybridization reagents) <ul style="list-style-type: none"> ■ Variable type ■ Variable value ■ Variable value unit of measure
EDM_10	Experimental metadata	*Responding variables (measured variables, result) [one entry for each responding variable] (e.g. Fluorescence intensity values) <ul style="list-style-type: none"> ■ Concept ■ Analyte ■ Measurement type ■ Measurement value (free text) ■ Measurement value units of measure (when applicable)
EDM_11	Experimental metadata	*Species (e.g. <i>Mus musculus</i> , variable)
EDM_12	Experimental metadata	*Unique Sample Identifier(s) (one Sample ID for each sample in the experiment, Please refer to table 5-15 for details on sample data)
EDM_13	Personnel	*Submitter (auto generated from login)
EDM_14	Personnel	*PI (auto generated from the system)

EDM_15	Personnel	*Experimenter(s) (Semi-colon delimited names of all the researchers involved, e.g. Smith, Sarah; Doe, John Q.)
EDM_16	File format definition	<p>*Column definition (Column 1, Column 2, etc; the columns should include the conditional/manipulated variables and responding variables/results for each Sample ID)</p> <p>For Affymetrix platforms:</p> <ul style="list-style-type: none"> ■ *ID_Ref (ProbeSet ID of Affymetrix chip) ■ *Value ("signal"-a measure of the abundance of a transcript) ■ Abs Call (the call in an absolute analysis that indicates if the transcript was present (P), Absent (A), Marginal (M), or No Call (NC)) ■ Detection_P_Value (P value that indicates the significance level of the detection call) ■ Additional column(s) ■ Data type for the additional column(s) (e.g. text, date, numerical, etc.) <p>For cDNA spotted array platforms:</p> <ul style="list-style-type: none"> ■ *Channel1_Mean (channel 1 mean raw intensity) ■ *Channel2_Mean (channel 2 mean raw intensity) ■ *Channel1_BG_Mean (channel 1 mean background level) ■ *Channel2_BG_Mean (channel 2 mean background level) ■ Channel1_Median (channel 1 median raw intensity) ■ Channel2_Median (channel 2 median raw intensity) ■ Channel1_BG_Median (channel 1 median background level) ■ Channel2_BG_Median (channel 2 median background level) ■ *Channel1_SD (channel 1 standard deviation) ■ *Channel2_SD (channel 2 standard deviation) ■ *Channel1_BG_SD (channel 1 background standard deviation) ■ *Channel2_BG_SD (channel 2 background standard deviation) ■ *Channel1_Processed (channel 1 processed signal, after dye normalization and background subtraction, used to compute the log ratio) ■ *Channel2_Processed (channel 2 processed signal, after dye normalization and background subtraction, used to compute the log ratio) ■ Ratio (raw ratio, base 10) ■ *Value (processed ratio, base 2, after basic filtering) ■ Additional column(s) ■ Data type for the additional column(s) (e.g. text, date, numerical, etc.) <p>For custom platforms</p> <ul style="list-style-type: none"> ■ *Probe ID ■ *Value ■ Additional column(s) ■ Data type for the additional column(s) (e.g. text, date, numerical, etc.)
EDM_17	File format definition	*Upload file
EDM_18	Data pre-processing	*Software used (e.g. GCOS)
EDM_19	Data pre-processing	*Version used (e.g. 51.2)
EDM_20	Data pre-processing	Analysis algorithm

Table 7-22 lists experimental data from Microarray experiments for gene expression that must be stored by the ImmPort system.

7.6 CAPTURE RESULTS DATA FROM THE STUDY OF GENETIC ASSOCIATIONS

The ImmPort system will be capable of storing results from experimental studies performed by the projects making up the Population Genetics Program. These result types, along with the project that will produce each type of experimental result, are shown in Table 7-23.

Table 7-23. Experimental Results Types from Population Genetics Program Projects

ID	Population Genetics Project	Requirement for Storage of Experimental Results	Priority
CRD_1	McMaster	The system shall capture the genotyping data of selected genes from SNP analysis.	1
CRD_2	McMaster	The system shall capture the data of genetic association between informative SNPs and clinical data including phenotypes.	4
CRD_3	McMaster	The system shall capture the statistical analyses to estimate clinically important changes over time in functional and immunological variables.	4
CRD_4	Mayo Clinic	The system shall capture the genotyping data of selected genes from SNP analysis.	1
CRD_5	Mayo Clinic	The system shall capture the data from variance analysis to compare HLA cell surface expression between the two subgroups defined by having the lowest or highest measurements of immune response.	4
CRD_6	Mayo Clinic	The system shall capture the data from correlation analysis to assess the associations between the measures of immune response and the measures of ex vivo secreted cytokine protein levels.	4
CRD_7	deCODE Genetics	The system shall capture the genotyping data from SNP analysis.	1
CRD_8	deCODE Genetics	The system shall capture the linkage analysis and case-control haplotype association of all LD blocks.	3
CRD_9	deCODE Genetics	The system shall perform sex-specific linkage analysis to look for host genes that might have a greater impact on sex.	4
CRD_10	UAB	The system shall capture the genotyping data of selected genes from SNP analysis.	1
CRD_11	UAB	The system shall capture the data of genetic association between informative SNPs and antibody responses and adverse reactions to anthrax vaccine in upwards of 1000 genotyped AVA recipients in the AVA000 trial.	4
CRD_12	UAB	The system shall capture data that characterize the functional significance of those variants in genes related to B cell antibody production and regulation.	4
CRD_13	RTI	The system shall capture the genotyping data of selected genes from SNP analysis.	1
CRD_14	RTI	The system shall capture the statistical data of categorization of immune response, non-, poor, and good responder groups.	4
CRD_15	RTI	The system shall capture the data of SNP-haplotype frequencies estimated from the software of HAPLOPOP.	4
CRD_16	RTI	The system shall capture the data of association analysis between SNP and SNP haplotypes and vaccine response.	4
CRD_17	University of Washington	The system shall capture the genotyping data of selected genes from SNP analysis.	1
CRD_18	University of Washington	The system shall identify genetic risk associations for clinical myocarditis using polymorphism analysis approach: Resequencing of the candidate genes Whole genome analysis High-speed chromatin profiling to identify distal regulatory sequences in candidate risk genes	4
CRD_19	University of Washington	The ImmPort system shall capture the statistical analysis results of the genotype-phenotype associations (chi-squared tests).	4

Table 7-23 lists experimental result types from Population Genetics Program projects that must be stored by the ImmPort system.

8.0 REFERENCE DATA

The ImmPort system shall store and maintain data obtained from public databases operated by the NIH and other organizations that is of particular importance to DAIT-funded research scientists. The primary initial focus of these data will be the genes associated with the field of immunology and all information associated with those genes. ImmPort shall store these reference data in an accessible form. The ImmPort system shall provide data from the following sources associated with Version 1.0 of the ImmPort Gene List of immunology-related genes.

Refer to Section 3.3 Product Priorities for a description of the requirement prioritization numbering scheme used in this section.

Table 8-1. Reference Data Categories

ID	Requirements	Priority
REF_1	The system shall store basic gene information.	1
REF_2	The system shall store basic protein information.	1
REF_3	The system shall store taxonomy information.	1
REF_4	The system shall store genomic sequence assembled chromosome data.	1
REF_5	The system shall store genomic sequence chromosome data.	1
REF_6	The system shall store gene expression information.	1
REF_7	The system shall store microsatellite polymorphism data.	1
REF_8	The system shall store SNP polymorphism data.	1
REF_9	The system shall store dbMHC polymorphism data.	1
REF_10	The system shall store network gene-gene interaction data.	1
REF_11	The system shall store network protein-protein interaction data.	1
REF_12	The system shall store pathways data.	1
REF_13	The system shall store AfCS Molecule Pages data.	1

Table 8-1 lists the categories of data that must be stored by the ImmPort system.

Table 8-2. General Functional Requirements for Reference Data

ID	Requirements	Priority
REF_14	The user shall be able to access at a minimum the 2 latest build versions for the human genome.	1
REF_15	The user shall be able to access at a minimum the 2 latest build versions for the mouse genome.	1
REF_16	The user shall be able to access at a minimum the 2 latest build versions for the rat genome.	1
REF_17	The user shall be able to access at a minimum the 2 latest build versions for the drosophila genome.	1
REF_18	The user shall be able to access at a minimum the latest build version for the chicken genome.	1
REF_19	When a genome build is taken offline, the build shall be archived.	1
REF_20	The user shall be able to retrieve the version history of public reference data.	1
REF_21	The user shall be able to view a record of obsolete data in the public data warehouse.	1
REF_22	The system shall maintain a list of immunology-specific genes.	1
REF_23	The user shall be able to view the build number of a genome sequence file.	2
REF_24	The user shall be able to view the corresponding NCBI build number of a UCSC gene structure file	1
REF_25	The system shall make the reference data current on a monthly basis.	1

Table 8-2 lists the general functional requirements that must be stored by the ImmPort system.

8.1 BASIC GENE DATA

REF_1. The system shall store basic gene information (see Table 8-3). This information also includes gene structure data.

Table 8-3. Basic Gene Information

ID	Table Name	Description	Links	Priority
BGI_1	Gene_info	This file contains basic gene information including Tax_id, GeneID, Symbol, Synonyms, Chromosome, Map location, Description etc.	ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_info	1
BGI_2	Gene2go	This file reports the GO terms that have been associated with Genes in Entrez Gene.	ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2go	1
BGI_3	Gene2pubmed	This file contains the gene ID and its association with unique identifier(s) in PubMed for a citation	ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2pubmed	1
BGI_4	HomoloGene	HomoloGene is a system for automated detection of homologs among the annotated genes of several completely sequenced eukaryotic genomes.	ftp://ftp.ncbi.nih.gov/pub/HomoloGene/ under the "Current Build", download the data of "homologene.data"	1
BGI_5	Gene2sts	This file contains the gene ID and its association with the unique identifier(s) given to a primer pair by UniSTS	ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2sts	1
BGI_6	Gene2refseq	This file is a comprehensive report of the RefSeq accessions that are related to a GeneID.	ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2refseq	1
BGI_7	Gene2Unigene	This file contains the gene ID and its association with the UniGene cluster(s).	ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2unigene	1
BGI_8	Sequence table	This file contains FASTA format sequence for a RefSeq accession number.	ftp://ftp.ncbi.nih.gov/refseq/H_sapiens/mRNA_Prot/	1
BGI_9	Mim2gene	The file contains the gene ID and its association with OMIM ID.	ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/mim2gene/	1
BGI_10	Mim_title	This file contains descriptions for phenotype/disease associated with OMIM IDs	ftp://ftp.ncbi.nih.gov/repository/OMIM	1
BGI_11	Gene_history	This file contains information about Entrez Gene IDs that are no longer current.	ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/mim2gene	1
BGI_12	Gene Structure Information from UCSC: KnownGene Table	The UCSC (University of California, Santa Cruz) Genome Browser provides a rapid and reliable display of any requested portion of genomes at any scale, together with dozens of aligned annotation tracks (known genes, predicted genes, ESTs, mRNAs, CpG islands, assembly gaps and coverage, chromosomal bands, mouse homologies, and more. We will only utilize the gene structure information from this database (e.g. exon start and end, cds start and end positions).	http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/	1
BGI_13	Gene Structure Information from UCSC: kgXref		http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/	1
BGI_14	Gene Structure Information from UCSC: KnowntoLocusLink	This table links RefSeq accession to Locus Link (EntrezGene) IDs	http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/	1
BGI_15	Gene Structure Information from UCSC: refgene	This table belongs to the refseq genes track. It contains the gene structure information on refseq genes.	http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/	1
BGI_16	Gene Structure Information from UCSC: reflink	This table belongs to the refseq genes track. Links between refseq accession number and locuslink ID can be obtained using this table.	http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/	1

Table 8-3 lists the basic gene information that must be stored by the ImmPort system.

Table 8-4. Basic Gene Information for Species

ID	Requirements	Priority
BGI_17	The system shall store basic gene information for the human species.	1
BGI_18	The system shall store basic gene information for the mouse species.	1
BGI_19	The system shall store basic gene information for the rat species.	1
BGI_20	The system shall store basic gene information for the drosophila species.	1
BGI_21	The system shall store basic gene information for the chicken species.	1

Table 8-4 lists the basic gene information for species that must be stored by the ImmPort system.

8.2 BASIC PROTEIN DATA

REF_2. The system shall store basic protein information (see Table 8-5).

Table 8-5. Basic Protein Data

ID	Table Name	Description	Links	Priority
BPI_1	UniProt	UniProt (Universal Protein Resource) is the world's most comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.	http://www.pir.uniprot.org/database/download.shtml Download the data from UniPRotKB	1
BPI_2	PIR (iProClass)	iProClass is one part of PIR database. It should provide the mapping of UniProt accession number to Entrez gene ID. It is a central point for exploration of protein information, provides summary descriptions of protein family, function and structure for PIR-PSD, Swiss-Prot, and TrEMBL sequences, with links to over 90 biological databases.	ftp://ftp.pir.georgetown.edu/pir_databases/iproclass/ iproclass.xml.gz iproclass.xml.dtd	1

Table 8-5 lists the sources of basic protein data for the ImmPort system.

8.3 TAXONOMY DATA

REF_3. The system shall store taxonomy data (see Table 8-6).

Table 8-6. Taxonomy Data

ID	Table Name	Description	Links	Priority
TAD_1	name.dmp	The file contains the mapping of taxonomy ID to name of organism.	ftp://ftp.ncbi.nih.gov/pub/taxonomy/TaxID The "scientific name" will be used for the names of organism	1

Table 8-6 lists the sources of basic protein data for the ImmPort system.

8.4 GENOMIC SEQUENCE DATA

8.4.1 Assembled Chromosomes

REF_4. The system shall store genomic sequence assembled chromosomes data (see Table 8-7).

Table 8-7. Assembled Chromosomes

ID	Table Name	Description	Links	Priority
ASC_1	chr_NC_gi	The chr_NC_gi file provides the accession and gi for the reference sequence (RefSeq)	ftp://ftp.ncbi.nih.gov/genomes/Organism/Assembled_chromosomes/chr_NC_gi	1

		chromosome records representing the reference genome assembly. columns: 1. chromosome 2. accession.version 3. gi	Note: Replace organism above by H_sapiens, M_musculus, R_norvegicus, Gallus_gallus, Drosophila_melanogaster for human, mouse, rat, chicken and fruit fly genomes respectively The data Chr_NC.gi is not available for fruit fly	
ASC_2	Sequence data (NC_sequence)	Sequence in FASTA format for each chromosome of the reference assembly. Each file is named according to the abbreviation for the species and the chromosome label.	ftp://ftp.ncbi.nih.gov/genomes/Organism/Assembled_chromosomes/XX_chr*.fa Note: The NC_sequence data is not available for chicken and fruit fly	1

Table 8-7 lists the sources of NCBI assembled chromosome data for the ImmPort system.

8.4.2 Chromosome Data (Contigs)

REF_5. The system shall store genomic sequence chromosome data (see Table 8-8).

Table 8-8. Chromosome data (Contigs)

ID	Table Name	Description	Links	Priority
CHD_1	Contig data	The seq_contig.md file provides information on the order and orientation of the contigs along the chromosome.	ftp://ftp.ncbi.nih.gov/genomes/organism/seq_contig.md Note: The seq_contig.md is not available for fruit fly	2
CHD_2	Sequence data (contig sequence)	The files in the chromosome directories provide concatenated sequence data for contigs that have been assembled from individual GenBank records. The contigs in the chromosome FTP directories are the same ones that are presented on the Entrez Map Viewer.	ftp://ftp.ncbi.nih.gov/genomes/organism/Chr_*/XX_ref_chr*.fa Note: for drosophila download the FASTA Nucleic Acid file	2

Table 8-8 lists the sources of chromosome data for the ImmPort system.

8.5 GENE EXPRESSION DATA

REF_6. The system shall store gene expression information (see Table 8-9).

Table 8-9. Gene Expression Data

ID	Table Name	Description	Links	Priority
GED_1	GEO expression data (download data for given GSE serials)	GEO (Gene Expression Omnibus) is "a high-throughput gene expression / molecular abundance data repository, as well as a curated, online resource for gene expression data browsing, query and retrieval"	ftp://ftp.ncbi.nih.gov/pub/geo/data/geo/by_series/	1

GED_2	Affymetrix platform table	The detail information of Affymetrix probe information, which used for mapping gene ID/gene name/gene symbol to probe	ftp://ftp.ncbi.nih.gov/pub/geo/data/geo/by_platform/ Note: The platform ID can be found in the expression profile of the serial ID.	1
-------	---------------------------	---	--	---

Table 8-9 lists the sources of gene expression data for the ImmPort system.

8.6 POLYMORPHISM DATA

The ImmPort system will include polymorphism data.

8.6.1 Microsatellite Polymorphism Data

REF_7. The system shall store microsatellite polymorphism data (see Table 8-10).

Table 8-10. Microsatellite Polymorphism Data

ID	Table Name	Description	Links	Priority
PMD_1	CEPH	Centre d'Etude du Polymorphisme Humain (CEPH) maintains a database of genotypes for genetic markers that have been typed on the CEPH reference family resource for linkage mapping (Citation: Genomics 6: 575-577, 1990; Science 265: 2049-2054, 1994). ceph_db/Ver_10/asc/ contains text files that represent all the concatenated markers and their genotypes for each chromosome.	ftp://ftp.cephb.fr/ceph_genotype_db/ceph_db/Ver_10/asc/	1
PMD_2	dbMHC	This database gives descriptive information for some of the known short tandem repeats within the MHC. It is integration of microsatellite characteristics in the MHC region. (Citation: Gourraud P.A., Mano S., Barnetche T., Carrington M., Inoko H. & Cambon-Thomsen A. (2004) Integration of microsatellite characteristics in the MHC region: a literature and sequence based analysis. Tissue Antigens, 64, 543. PubMed ID 15496197.)	http://www.ncbi.nlm.nih.gov/mhc/xslcgi.fcgi?cmd=msearch&user_id=0&probe_id=0&source_id=0&locus_id=0&locus_group=&proto_id=0&banner=&kit_id=0&du mmy=0	1

Table 8-10 lists the sources of microsatellite polymorphism data for the ImmPort system.

8.6.2 SNP Polymorphism Data

REF_8. The system shall store SNP polymorphism data (see Table 8-11).

Table 8-11. SNP Polymorphism Data

ID	Table Name	Description	Links	Priority
PSD_1	dbSNP: flat	The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms. Each dbSNP entry includes the sequence context of the polymorphism (i.e., the surrounding sequence), the occurrence frequency of the polymorphism (by population or individual), and the experimental method(s), protocols, and conditions used to assay the variation. More information: http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.ch apter.1143	ftp://ftp.ncbi.nih.gov/snp/	1

PSD_2	dbSNP : fasta	<p>The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms. Each dbSNP entry includes the sequence context of the polymorphism (i.e., the surrounding sequence), the occurrence frequency of the polymorphism (by population or individual), and the experimental method(s), protocols, and conditions used to assay the variation.</p> <p>More information: http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.1143</p> <p>Please refer to SNP data document for details on fields required to be downloaded</p> <p>This particular table contains information on fasta format sequence along with build and other details for a SNP</p>	ftp://ftp.ncbi.nih.gov/snp/human/rs_fasta	1
PSD_3	dbSNP: genotype	<p>The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms. Each dbSNP entry includes the sequence context of the polymorphism (i.e., the surrounding sequence), the occurrence frequency of the polymorphism (by population or individual), and the experimental method(s), protocols, and conditions used to assay the variation.</p> <p>More information: http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.1143</p> <p>Please refer to SNP data document for details on fields required to be downloaded</p> <p>This particular table contains information of genotypes on individuals in a particular population (XML file format)</p>	ftp://ftp.ncbi.nih.gov/snp/human/genotype	1
PSD_4	HapMap	<p>The International HapMap Project is a multi-country effort to identify and catalog genetic similarities and differences in human beings. This directory contains data files with genotypes submitted by HapMap genotyping centers to the DCC. Data for the following populations are available:</p> <p>CEU: CEPH (Utah residents with ancestry from northern and western Europe)</p> <p>CHB: Han Chinese in Beijing, China</p> <p>JPT: Japanese in Tokyo, Japan</p> <p>YRI: Yoruba in Ibadan, Nigeria</p> <p>More information: http://www.hapmap.org/</p>	http://www.hapmap.org/genotypes/latest/	1

Table 8-11 lists the sources of SNP polymorphism data for the ImmPort system.

Table 8-12. SNP Polymorphism Data for Species

ID	Requirements	Priority
PSD_5	The system shall store SNP polymorphism information for the human species.	1
PSD_6	The system shall store SNP polymorphism information for the mouse species.	1
PSD_7	The system shall store SNP polymorphism information for the rat species.	1
PSD_8	The system shall store SNP polymorphism information for the drosophila species.	1
PSD_9	The system shall store SNP polymorphism information for the chicken species.	1

Table 8-12 lists the sources of SNP polymorphism data for the ImmPort system.

8.6.3 dbMHC Polymorphism Data

REF_9. The system shall store dbMHC polymorphism data (see Table 8-13).

Table 8-13. dbMHC Polymorphism Data

ID	Table Name	Description	Links	Priority
PDD_1	Allele frequency tables (3 tables) (Includes class I/classII and haplotype frequency data)	The goal of the Diversity/Anthropology Component is to determine HLA class I and class II allele and haplotype frequencies in various human populations. Studies of allelic diversity in different populations can shed light on the evolution of HLA polymorphism as well as on the evolution and migration of human populations.	http://www.ncbi.nlm.nih.gov/mhc/ihwg.cgi?cmd=page&user_id=0&probe_id=0&source_id=0&locus_id=0&locus_group=&proto_id=0&banner=&kit_id=0&page=AnthroMain	1
PDD_2	Anthropology data table	The goal of the Diversity/Anthropology Component is to determine HLA class I and class II allele and haplotype frequencies in various human populations. Studies of allelic diversity in different populations can shed light on the evolution of HLA polymorphism as well as on the evolution and migration of human populations.	http://www.ncbi.nlm.nih.gov/mhc/ihwg.cgi?cmd=page&user_id=0&probe_id=0&source_id=0&locus_id=0&locus_group=&proto_id=0&banner=&kit_id=0&page=AnthroMain	1
PDD_3	HCT table	The Hematopoietic Cell Transplantation (HCT) component of the International Histocompatibility Working Group (IHWG) has generated and collected HLA genotype and clinical outcome information on hematopoietic cell transplants performed worldwide. The IHWG HCT studies seek to determine whether complete allele matching for HLA-A, B, C DRB1, DQB1, and DPB1 is necessary to optimize transplantation. More information: http://www.ncbi.nlm.nih.gov/projects/mhc/ihwg.cgi?cmd=page&page=HCTov	http://www.ncbi.nlm.nih.gov/mhc/ihwg.cgi?cmd=page&user_id=0&probe_id=0&source_id=0&locus_id=0&locus_group=&proto_id=0&banner=&kit_id=0&page=HCTintro	1
PDD_4	Allele nucleotide sequence table (nuc_fasta and prot_fasta files)	The IMGT/HLA Sequence Database allows you to retrieve sequence information upon a specific HLA allele as named in the WHO Nomenclature Committee Reports.		1

Table 8-13 lists the sources of dbMHC polymorphism data for the ImmPort system.

8.7 PROTEIN/GENE NETWORK DATA

8.7.1 Pathways Data

REF_10. The system shall store pathways data (see Table 8-14).

Table 8-14. Pathways Data

ID	Table Name	Description	Links	Priority
MPD_1	KEGG (Kyoto Encyclopedia of Genes and Genomes)	KEGG (Kyoto Encyclopedia of Genes and Genomes) integrates current knowledge on molecular interaction networks in biological processes (PATHWAY database), the information about the universe of genes and proteins (GENES/SSDB/KO databases), and the information about the universe of chemical compounds and reactions (COMPOUND/GLYCAN/REACTION databases).	ftp://ftp.genome.ad.jp/pub/kegg/linkdb/genes/ genes_ko.list (Gene2KO table) genes_ncbi-geneid.list (Gene2EntrezID table) genes_pathway.list (Gene2Pathway table) genes_uniprot.list (Gene2UniProt table)	1

MPD_2	BioCyc	"The BioCyc collection of Pathway/Genome Databases (DBs) provides electronic reference sources on the pathways and genomes of different organisms."	http://www.biocyc.com The data have been sent before. Original data source is from pathways.col	1
-------	--------	---	--	---

Table 8-14 lists the sources of metabolic pathways data for the ImmPort system.

8.7.2 Interaction Data

REF_11. The system shall store gene-gene interaction data (see Table 8-15).

Table 8-15. Gene-Gene Interaction

ID	Table Name	Description	Links	Priority
GGI_1	B cell regulatory network	The interactions of gene of human B-cell	The data is in the file of "TheMatrixIII_15%_clean.adj.txt", please check the ImmPort FTP site The information of platform (Affymetrix HG95Av2) can be downloaded from GEO with GPL91. The information in the platform provides us mapping from affymetrix probe IDs to Entrez Gene IDs.	1
GGI_2	Affymetrix Platform	The file contains the mapping from affymetrix probe IDs to Entrez Gene IDs. The Affymetrix platform used in the B cell regulatory network is HG95Av2.	The platform can be downloaded from GEO with GPL91. ftp://ftp.ncbi.nih.gov/pub/geo/data/geo/by_platform/GPL91_family.sof.t.gz	1

Table 8-15 lists the sources of gene-gene interaction data for the ImmPort system.

REF_12. The system shall store protein-protein interaction data (see Table 8-16).

Table 8-16. Protein-Protein Interactions

ID	Table Name	Description	Links	Priority
PPI_1	BIND	<p>BIND (Biomolecular Interaction Network Database) is “a collection of records documenting molecular interactions”. It is an interaction database with three classifications for molecular associations: molecules that associate with each other to form interactions, molecular complexes that are formed from one or more interaction(s) and pathways that are defined by a specific sequence of two or more interactions.</p>	<p>ftp://ftp.blueprint.org/pub/BIND/data/bindflatfiles/bindindex/yyyymdd.refs.txt The dtd can be downloaded from http://bind.ca/Action?pg=22000&path=spec</p> <p>The data of data set for human are:</p> <p>ftp://ftp.blueprint.org/pub/BIND/data/bindflatfiles/bindindex/yyyymdd.ints.txt (Interaction Table)</p> <p>ftp://ftp.blueprint.org/pub/BIND/data/bindflatfiles/bindindex/yyyymdd.refs.txt (Evidence Table)</p> <p>ftp://ftp.blueprint.org/pub/BIND/data/bindflatfiles/bindindex/yyyymdd.labels.txt (Labels Table)</p> <p>ftp://ftp.blueprint.org/pub/BIND/data/bindflatfiles/bindindex/yyyymdd.complex2subunit.txt (Complex2Subunits Table)</p> <p>Other species data such as Gallus gallus (chicken, tax_id=9031) and Drosophila melanogaster (fruit fly, tax_id=7227) are also available</p>	1
PPI_2	MIPS	<p>MIPS (Munich Information center for Protein Sequences) is “a collection of manually curated high-quality PPI (protein-protein interaction) data collected from the scientific literature by expert curators”</p>	<p>http://mips.gsf.de/proj/ppi/allppis.xml (interaction table)</p>	2
PPI_3	DIP	<p>The DIPTM database catalogs experimentally determined interactions between proteins. It combines information from a variety of sources to create a single, consistent set of protein-protein interactions.</p>	<p>http://dip.doe-mbi.ucla.edu/dip/Download.cgi</p> <p>The data for interaction table are from: DIP Hsapi20050126.txt (human) DIP Mmusc20050126.txt (mouse) The Fruit fly data is also available</p> <p>The data for Evidence table are from: XML file</p>	2

PPI_4	Reactome	Reactome is a curated resource of core pathways and reactions in human biology. The data stored in ImmPort is only the data of human protein-protein interaction.	http://www.reactome.org/download/index.html Download the data from "Human protein-protein interaction pairs in tab-delimited format."	1
PPI_5	AfCS Y2H	AfCS Y2H (Alliance for Cellular Signaling Yeast 2-Hybrid screen) is "used to list proteins that have given interactions using the yeast two-hybrid system (AfCS large-scale yeast two-hybrid screens)	The data of interaction of AfCS Yeast2Hybrid is from the file of "04042005_bait_preyn_interactions.txt" README_Y2H.txt	1

Table 8-16 lists the sources of protein-protein interaction data for the ImmPort system.

Table 8-17 presents the requirements for protein-protein interaction data for species.

Table 8-17. Protein-Protein Interactions for Species

ID	Requirements	Priority
PPI_6	The system shall store protein-protein data for the human species.	1
PPI_7	The system shall store protein-protein data for the mouse species.	1

Table 8-17 lists protein-protein interaction data for species for the ImmPort system.

8.8 AFCS MOLECULE PAGES

The AfCS Molecule Pages database focuses on the mouse species. (see Table 8-18).

REF_13. The system shall store AfCS molecule data.

Table 8-18. AfCS Molecule Pages Data

ID	Table Name	Description	Links	Priority
STP_1	Tab-delimited file relating AfCS IDs to protein database records	The AfCS Molecule Pages is a "database of key facts about proteins involved in cellular signaling". Each page is dedicated to one protein with a unique AfCS ID. This table links AfCS IDs to protein accession numbers.	ftp://ftp.afcs.org/pub/mpdata/afcs2prot.dat	1
STP_2	Tab-delimited file relating AfCS IDs to LocusLink (EntrezGene) IDs	This table links AfCS IDs to Locus Link (EntrezGene) IDs.	ftp://ftp.afcs.org/pub/mpdata/afcs2ll.dat	1
STP_3	Hyperlink to: Information contained within AfCS molecule pages	Some proteins in the AfCS Molecule Pages contain additional information provided by an expert author that may include pathway information, protein function information, etc. As of now, the author-entered information is not in a downloadable format, therefore we will simply provide the hyperlink to the AfCS Molecule Page associated with the queried gene/protein.	http://www.signaling-gateway.org/molecule/	1

Table 8-18 lists the sources of signal transduction pathways data for species for the ImmPort system.

8.9 IMMPORT GENE LIST

REF_22. The system shall maintain a list of immunology-specific genes.

Table 8-19. Publicly Available Data for Candidate Genes Comprising the ImmPort Gene List

ID	Data Type	Links to Other Reference Data
IGL_1	Entrez Gene ID	Primary key
IGL_2	Taxonomy ID	Foreign Key linked to Taxonomy data
IGL_3	Entrez Gene symbol	
IGL_4	Locus tag	
IGL_5	Gene Name (Description)	
IGL_6	Gene Synonyms	
IGL_7	Gene Type	
IGL_8	Nomenclature Symbol	
IGL_9	Nomenclature Full Name	
IGL_10	Nomenclature Status	
IGL_11	OMIM ID	Hyperlink to NCBI OMIM
IGL_12	UniGene ID	
IGL_13	RNA Nucleotide Accession Number (mRNA RefSeq Accession Number)	Foreign Key linked to RefSeq mRNA data for hyperlink to NCBI mRNA RefSeq
IGL_14	RNA nucleotide GI	
IGL_15	Protein Accession Number (RefSeq Protein)	Foreign Key linked to RefSeq protein data for hyperlink to NCBI Protein RefSeq
IGL_16	Protein GI	
IGL_17	Gene Ontology	Hyperlink to Gene Ontology
IGL_18	Homologene ID	
IGL_19	Homologene Protein Accession	
IGL_20	Homologene Protein GI	
IGL_21	Position: {Chromosomal Map location}	
IGL_22	LD Select Based Tag SNPs	

Table 8-19 lists the data items to be captured for the ImmPort gene list.

9.0 LOADING EXPERIMENTAL AND RELATED DATA INTO IMMPORT

The ImmPort system will make available a PPW for each DAIT-funded research project. Investigators associated with a research contract/grant can load, store, and analyze unpublished experimental data with the data shielded from view by other ImmPort system users. These investigators can also directly load data for submission to the public data warehouse. To be able to load data into the PPW or directly to the public data warehouse, ImmPort will address the data loading requirements specified in Tables 9-1 and 9-2.

Note that the system will not provide data cleansing. It is the responsibility of the research projects to cleanse their data prior to submission for loading into the ImmPort system.

Refer to Section 3.3 Product Priorities for a description of the requirement prioritization numbering scheme used in this section.

Table 9-1. Experimental Data Load Requirements

ID	Requirements	Priority
LER_1	The system shall provide a data load capability through which the user can submit experimental data comprising multiple files. (The number of files depends on the experimental technique used and the file format that the user chooses.)	1
LER_1a	The system shall provide a data load capability through which the user can submit subject and sample data (defined in Section 7.3).	1
LER_1b	The system shall provide a data load capability through which the user can submit experimental platform data (defined in TBD).	1
LER_1c	The system shall provide a data load capability through which the user can submit series data (defined in TBD).	1
LER_1d	The system shall provide a data load capability through which the user can submit experimental results data (defined in Section 7.5. Biochemical Assay Data, EDG_16).	1
LER_2	[Changed to LER_1a.]	
LER_3	The system shall capture metadata on the experimental data load.	2
LER_4	The system shall provide interactive forms with upload capability through which a user can submit experimental data and the associated metadata.	3
LER_5	The system shall ensure the data submission complies with established minimum information about the experiments in Section 7.	1
LER_6	The system shall enable users to submit metadata associated with experimental results data.	1
LER_7	The system shall provide error checking of data submitted for loading into the ImmPort system to ensure the data meet minimum technical requirements for storage in the system. The system will not perform error checking on the scientific validity or reliability of the data.	1
LER_8	The system shall provide an error message if user input is not consistent with the system standard.	1
LER_9	The system shall reject data that do not meet the technical standard.	1
LER_10	The system shall notify the user if the user needs to correct the file for resubmission.	1
LER_11	The system shall notify the user	
LER_11	The system shall notify the user upon successful loading of submitted data.	1
LER_12	When a data set is updated in the public data warehouse, the system shall assign a version number to each version of the data set. The data set version preceding an updated data set version will not be deleted from the public data warehouse.	2
LER_12a	The system shall retain all previous data set versions.	2

LER_13	The system shall capture relationships between data sets, e.g., those relationships between primary data sets and secondary data sets and those between metadata and primary and/or secondary data sets.	1
LER_14	The system shall request metadata on the experimental design (with minimum mandatory metadata) to be able to deposit the data in the correct area.	1
LER_15	The system shall enable the user to define the headers of one or more tabular files. (It is anticipated that each record in the file will correspond to an individual sample and columns will correspond to the metadata associated with each sample).	1
LER_16	The system shall enable the user to upload one or more tabular files that contain information regarding each individual sample evaluated in the experiment.	1
LER_17	The system shall require the user to submit experimental data using predefined column headers for each experimental platform/technique in ImmPort v.1.0.	1
LER_18	The system shall enable the user to load data into a PPW.	1
LER_18a	The system shall display experimental data (meeting the required technical standards) with an aggregate zipped file size of TBD in the PPW within 1 business day of the user loading them into the system for storage.	2
LER_19	The system shall enable the user to submit data directly to the public data warehouse without having to first load data in a PPW.	1
LER_19a	The system shall display experimental data (meeting the required technical standards) with an aggregate zipped file size of TBD in the data warehouse within 1 business day of the user loading into the system for publication in the data warehouse.	2

Table 9-1 lists requirements for loading of experimental data into the ImmPort system.

Table 9-2 prioritizes the implementation of experimental data load capabilities by the ImmPort system for experimental techniques and procedures identified as requiring support at the time this document was published. Data load capabilities for experimental techniques will be implemented in priority order. A wizard interface load capability will be implemented after the 2-year base period.

Table 9-2. Support for Data Load Capabilities – Prioritization

ID	Requirement	Priority
DLC_1	The system shall support batch loading of experimental data associated with use of the Illumina Bead Array technique for SNP analysis and genotyping.	1
DLC_1a	The system shall require the user to batch load experimental data in the HapMap XML format.	1
DLC_2	The system shall support batch loading of experimental data associated with use of the Affymetrix SNP Chip genotyping technique for SNP analysis and genotyping.	1
DLC_3	[Expanded to DLC_3.1 and DLC_3.2.]	
DLC_3.1	The system shall support batch loading of experimental data associated with the microarray technique using GenePix for measuring mRNA expression.	2
DLC_3.1a	The system shall support storage of the GenePix results file (.GPR) consisting of expression values.	2
DLC_3.1b	The system shall support storage of the GenePix print array layout file (.GAL).	2
DLC_3.1c	The system shall support storage of the GenePix alignment file (.GPS). The GPS file stores the positions and sizes of the grid of blocks and features (spots) that correspond to the array, as well as the settings for acquisition, analysis, and display. During analysis, the grid of the blocks and features is aligned with the array prior to extracting data.	2
DLC_3.1d	The system shall support storage of the GenePix raw image file (.tiff).	2
DLC_3.2	The system shall permit the user to batch load experimental data associated with the microarray technique using Affymetrix GeneChip for measuring mRNA expression.	1
DLC_3.2a	The system shall permit the user to batch load experimental data in the GEO SOFT format.	1
DLC_3.2b	The system shall permit the user to batch load experimental data in a .txt file resulting from conversion of the Affymetrix .chp file (i.e., the processed Affymetrix GeneChip file).	1
DLC_3.2c	The system shall support storage of the Affymetrix GeneChip processed data file (.chp).	1
DLC_3.3d	The system shall support storage of the Affymetrix GeneChip intensity values file (.cel).	
DLC_3.2e	The system shall support storage of the Affymetrix GeneChip raw image file (.dat).	

DLC_3.2f	The system shall support storage of the Affymetrix GeneChip chip and quality control summary file (.rpt).	
DLC_3.2g	The system shall support storage of the Affymetrix GeneChip sample and chip information file (.exp).	
DLC_4	The system shall support batch loading of experimental data associated with the use of the microsatellite analysis technique for SNP analysis and genotyping.	2
DLC_5	The system shall support batch loading of experimental data associated with the use of HLA typing and other genotyping techniques for SNP analysis and genotyping.	2
DLC_6	The system shall support batch loading of experimental data associated with the use of the sequencing technique for SNP analysis and genotyping.	2
DLC_7	The system shall support batch loading of experimental data associated with the use of the flow cytometry technique for measuring protein expression.	2
DLC_8	The system shall support batch loading of experimental data associated with use of the ELISA technique for measuring immune responses.	2
DLC_9	The system shall support batch loading of experimental data associated with use of the ELISPOT technique for measuring immune response.	3
DLC_10	The system shall support batch loading of experimental data associated with use of the mass spectrometry technique for measuring protein expression.	3
DLC_11	The system shall support batch loading of experimental data associated with use of the Luminex Bead Assay for measuring protein function.	3
DLC_12	The system shall support batch loading of experimental data associated with use of the cytometric bead immunoassay for measuring protein function.	3
DLC_13	The system shall support batch loading of experimental data associated with use of the 2D-GEL technique for measuring protein expression.	4
DLC_14	The system shall support batch loading of experimental data associated with use of the ESI-LC technique for measuring protein expression.	4
DLC_15	The system shall support batch loading of experimental data associated with use of the PCR technique for measuring mRNA expression.	4
DLC_16	The system shall support batch loading of experimental data associated with use of the CHIP technique for measuring DNA binding activity.	4
DLC_17	The system shall support batch loading of experimental data associated with use of the EMSA technique for measuring DNA binding activity.	4
DLC_18	The system shall support batch loading of experimental data associated with functional validation techniques using siRNA	4
DLC_19	The system shall support batch loading of experimental data associated with use of transgenic or null mutant mice.	4
DLC_20	The system shall support batch loading of experimental data associated with use of the Western Blot technique for quantifying the function of proteins and genes.	4

Table 9-2 lists biochemical assays classified as a combination of experiment type and experimental technique/platform, with the priority with which each will be receive either batch or interactive data load support by the ImmPort system.

10.0 DATA MANIPULATION

10.1 DATA QUERY AND EXTRACTION

The ImmPort system will have querying (retrieval) capabilities that include the capability for a user to save the data retrieved from the ImmPort database within ImmPort or to download the data to a local system for analysis. Table 10-1 summarizes data query and extraction requirements and assigns a priority of 1 or 2 to each.

Refer to Section 3.3 Product Priorities for a description of the requirement prioritization numbering scheme used in this section.

Table 10-1. Data Query and Extraction Requirements with Priority

ID	Requirements	Priority
DQE_1	The system shall provide a “menu-driven” capability to assist the user in construction of a query against data in a PPW to which the user is authorized access.	1
DQE_2	The system shall provide a “menu-driven” capability to assist the user in construction of a query against data in a CW to which the user is authorized access.	2
DQE_3	The system shall provide a “menu-driven” capability to assist the user in construction of a query against data in the public data warehouse.	1
DQE_4	The system shall provide a “menu-driven” capability to assist the user in construction of a query against data from a combination of any of the areas (PPW, CW, and public data warehouse) to which the user is authorized access.	3
DQE_5	The system shall provide a capability for the authorized user to construct and submit an SQL query against data in a PPW to which the user is authorized access.	2
DQE_6	The system shall provide a capability for the authorized user to construct and submit an SQL query against data in a CW to which the user is authorized access.	2
DQE_7	The system shall provide a capability for the authorized user to construct and submit an SQL query against data in the public data warehouse.	2
DQE_8	The system shall provide a capability for the authorized user to construct and submit an SQL query against data from a combination of any of the areas (PPW, CW, the public data warehouse) to which the user is authorized access.	2
DQE_9	The system shall provide a capability for the authorized user to submit a standard query providing limited modifiable selection and sort parameters against data in a PPW to which the user is authorized access.	1
DQE_10	The system shall provide a capability for the authorized user to submit a standard query providing limited modifiable selection and sort parameters against data in a CW to which the user is authorized access.	2
DQE_11	The system shall provide a capability for the authorized user to submit a standard query providing limited modifiable selection and sort parameters against data in the public data warehouse.	1
DQE_12	The system shall provide a capability for the authorized user to submit a standard query providing limited modifiable selection and sort parameters against data from a combination of any of the areas (PPW, CW, the public data warehouse) to which the user is authorized access.	2
DQE_13	The system shall enable the user to perform a query on reference data originating from external public sources and stored in ImmPort.	1
DQE_14	The system shall provide a capability for the user to extract and download a data set comprising the result set of a query.	1
DQE_15	The system shall enable the user to save data resulting from a query to a PPW to which the user is authorized access.	1
DQE_16	The system shall provide a limited level of query history and results for the user’s current session.	1
DQE_17	The system shall enable the user to reuse a previously constructed query recorded in the current session’s query history.	1
DQE_18	The system shall enable the user to modify and reuse a previously constructed query recorded in the current session’s query history.	1
DQE_19	The system shall enable the user to save a query for future use beyond the current session.	1
DQE_20	The system shall enable the user to reuse a previously constructed query saved from a previous session.	1

ID	Requirements	Priority
DQE_21	The system shall enable the user to modify a previously constructed query saved from a previous session (e.g., changing the logical parameters of the query).	1
DQE_22	The system shall provide a capability to limit the ability of a non-proficient SQL user to submit an SQL query.	4
DQE_23	The system shall notify the user if the estimated time for completing a query exceeds a threshold specified by DAIT based on impact to system performance.	2
DQE_24	The system shall have a capability to queue a query for off-hour execution if the estimated time for completing the query exceeds a threshold specified by DAIT based on impact to system performance.	2
DQE_25	The system shall notify the user when a query is queued for off-hour execution.	2
DQE_26	The system shall notify the user when a query is complete.	2
DQE_27	The system shall instruct the user on how to retrieve query results.	1
DQE_28	The system shall provide the capability for the user to query and display experimental data in the public data warehouse by research contract/grant.	2
DQE_29	The system shall provide the capability for the user to query and display data in the public data warehouse or a PPW by experimental platform or technique.	2
DQE_30	The system shall provide the capability for the user to query and display experimental data in the public data warehouse by phenotypic characteristic.	2
DQE_31	The system shall provide the capability for the user to query and display data in the public data warehouse or a PPW by biochemical assay type.	2
DQE_32	The system shall provide ontology-aided query capabilities.	2
DQE_33	The system shall return results for a query on individual phenotypic information if the total number of phenotype records is equal to or greater than a minimum number defined by NIAID.	2
DQE_34	The system shall enable the user to query the experimental results data values.	2
DQE_35	The system shall provide the capability to query the metadata submitted with experimental data in the PPW, CW, or data warehouse.	1
DQE_36	The system shall provide the ability to construct a single query using a combination of mutually exclusive Boolean operators (e.g., a single query statement that uses both the "AND" and "OR" operators).	2
DQE_37	The system shall enable the user to search for keywords describing other research projects in ImmPort.	1
DQE_38	The system shall enable the user to search for keywords describing experiment groups in PPWs with which he is associated.	1
DQE_39	The system shall enable the user to search for specified constant variable types associated with experiment groups in PPWs with which he is associated.	1
DQE_40	The system shall enable the user to search for specified conditional/manipulated variable types associated with experiment groups belonging to research projects with which he is associated.	1
DQE_41	The system shall enable the user to search for specified responding variable types associated with experiment groups belonging to research projects with which he is associated.	3
DQE_42	The system shall enable the user to determine at the time of registration whether he would like his contact information to be made available to other users.	2
DQE_43	The system shall enable the user to change settings on whether his/her contact information is available to other users.	2
DQE_44	The system shall make contract/grant information searchable.	2
DQE_45	The system shall enable the user to query the experimental results data values.	2
DQE_46	The system shall enable the user to query for phenotypes that have been genotyped with SNPs near a particular gene.	3
DQE_47	The system shall enable the user to run a query to determine if a given gene has been genotyped by another ImmPort research project.	3
DQE_48	The system shall enable the user to search genotype data in particular population(s) from experimental SNP data.	3
DQE_49	The system shall enable the user to search experimental microarray data.	2

ID	Requirements	Priority
DQE_50	The system shall enable the user to search subject and sample information when searching experimental microarray data	2
DQE_51	The system shall enable the user to submit microarray data to the Biomind analytics software.	1
DQE_52	The system shall enable the user to download data from the data public warehouse.	1
DQE_53	The system shall enable the user to export data resulting from a query to a text file.	1
DQE_54	When returning query results, the system shall identify the original source of a data set.	1
DQE_55	The system shall present the user with the most current information (as determined by accession version numbers of records) for a given genome build. (Note that when the system supports multiple versions of the genomes for one species, the newest version of an mRNA for each genome build may be different.)	2
DQE_55a	If a user submits a query with input specifying an older accession version number, the system shall inform the user that the record has been updated.	2

Table 10-1 lists data query requirements for the ImmPort system.

10.2 REFERENCE DATA QUERY REQUIREMENTS

The system will provide query capabilities for basic gene data.

Table 10-2. Basic Gene Data Query Requirements

ID	Requirements	Priority
BGD_1	The system shall enable the user to search basic gene data information by Entrez Gene ID.	1
BGD_2	The system shall enable the user to search basic gene data information by chromosome region.	1
BGD_3	The system shall enable the user to search basic gene data information by gene name.	1
BGD_4	The system shall enable the user to search basic gene data information by gene symbol.	1
BGD_5	The system shall enable the user to search basic gene data information by NCBI mRNA accession number.	1
BGD_6	The system shall enable the user to search basic gene data information by NCBI protein accession number.	1
BGD_7	The system shall enable the user to view basic gene information for each record that is returned.	1
BGD_8	The system shall enable the user to view transcript information for each record that is returned.	1
BGD_9	The system shall enable the user to view protein information for each record that is returned.	1
BGD_10	The system shall enable the user to view homolog information for each record that is returned.	1
BGD_11	The system shall enable the user to view gene expression information for each record that is returned.	1
BGD_12	The system shall enable the user to view publication information for each record that is returned.	1

Table 10-2 lists basic gene data query requirements for the ImmPort system.

The system will provide query capabilities for basic protein data.

Table 10-3. Basic Protein Data Query Requirements

ID	Requirements	Priority
BPD_1	The system shall enable the user to search basic protein data information by Entrez Gene ID.	1

ID	Requirements	Priority
BPD_2	The system shall enable the user to search basic protein data information by gene name.	1
BPD_3	The system shall enable the user to search basic protein data information by gene symbol.	1
BPD_4	The system shall enable the user to search basic protein data information by UniProt accession number.	1
BPD_5	The system shall enable the user to search basic protein data information by protein name.	1
BPD_6	The system shall enable the user to search basic protein data information by peptide accession number.	1
BPD_7	The system shall enable the user to view basic protein information for each record that is returned.	1
BPD_8	The system shall enable the user to view reference information for each record that is returned.	1
BPD_9	The system shall enable the user to view features information for each record that is returned.	1
BPD_10	The system shall enable the user to view protein sequence for each record that is returned.	1
BPD_11	The system shall enable the user to view functions information for each record that is returned.	1
BPD_12	The system shall enable the user to view pathways information for each record that is returned.	1

Table 10-3 lists basic protein data query requirements for the ImmPort system.

The system will provide query capabilities for pathway data.

Table 10-4. Pathway Data Query and Extraction Requirements

ID	Requirements	Priority
PDQ_1	The system shall enable the user to search pathway data information by Entrez Gene ID.	1
PDQ_2	The system shall enable the user to search basic protein data information by gene symbol.	1
PDQ_3	The system shall enable the user to search basic protein data information by gene name.	1
PDQ_4	The system shall enable the user to search basic protein data information by pathway name.	1
PDQ_5	The system shall enable the user to search basic protein data information by UniProt protein accession.	1
PDQ_6	The system shall enable the user to search basic protein data information by RefSeq protein accession.	1

Table 10-4 lists pathway data query requirements for the ImmPort system.

The system will provide query capabilities for protein network data.

Table 10-5. Protein Network Data Query and Extraction Requirements

ID	Requirements	Priority
PND_1	The system shall enable the user to search protein network data information by Entrez Gene ID.	1
PND_2	The system shall enable the user to search protein network data information by gene symbol.	1
PND_3	The system shall enable the user to search protein network data information by gene name.	1
PND_4	The system shall enable the user to search protein network data information by UniProt accession number.	1

ID	Requirements	Priority
PND_5	The system shall enable the user to search protein network data information by protein name.	1
PND_6	The system shall enable the user to search protein network data information by RefSeq protein accession.	1
PND_7	The system shall require the user to limit the query to one organism.	1
PND_8	The system shall return the interaction ID	

Table 10-5 lists basic protein network data query requirements for the ImmPort system.

The system will provide query capabilities for SNP polymorphism data.

Table 10-6. SNP Polymorphism Data Query Requirements

ID	Requirements	Priority
PDQ_1	The system shall enable the user to search SNP polymorphism data information by Entrez Gene ID.	1
PDQ_2	The system shall enable the user to search SNP polymorphism data information by gene symbol.	1
PDQ_3	The system shall enable the user to search SNP polymorphism data information by gene name.	1
PDQ_4	The system shall enable the user to search SNP polymorphism data information by SNP ID name.	1
PDQ_5	The system shall enable the user to search SNP polymorphism data information by NCBI mRNA accession.	1
PDQ_6	The system shall enable the user to search SNP polymorphism data information by chromosome region.	1
PDQ_7	The system shall enable the user to limit the SNP polymorphism data query by SNP function.	1
PDQ_7	The system shall enable the user to limit the SNP polymorphism data query by population.	2
PDQ_8	The system shall enable the user to limit the SNP polymorphism data query by source.	2

Table 10-6 lists SNP polymorphism data query requirements for the ImmPort system.

10.3 EXPERIMENTAL DATA QUERY REQUIREMENTS

The system will provide query capabilities for experimental SNP polymorphism data.

Table 10-7. SNP Polymorphism Data Query Requirements

ID	Requirements	Priority
PDQ_1	The system shall enable the user to search SNP polymorphism data information by Entrez Gene ID.	1

Table 10-7 lists SNP polymorphism data query requirements for the ImmPort system.

The system will provide query capabilities for experimental microarray data.

Table 10-8. Microarray Data Query Requirements

ID	Requirements	Priority
PDQ_1	The system shall enable the user to search SNP polymorphism data information by Entrez Gene ID.	1

Table 10-8 lists SNP polymorphism data query requirements for the ImmPort system.

10.4 DATA ANALYSIS

ImmPort system users will be able to perform simple and complex analyses on the data stored in ImmPort.

Table 10-9. Data Analysis Requirements with Priority

ID	Requirements	Priority
ANA_1	The system shall allow the user to construct analytical pipelines consisting of multiple step analyses of specified data, each step using the output of the previous step.	2
ANA_2	The system shall record analytical pipelines and the related results that are generated throughout the pipeline steps.	2
ANA_3	The system shall enable the user to return to any step within the analytical pipeline, modify analytical parameters, and re-run the analysis steps following the selected step.	2
ANA_4	The system shall only retain a record of analytical pipeline steps during the current user session unless the user explicitly saves the pipeline before logging off.	2
ANA_5	The system shall retain the result sets from each step of an analytical pipeline during a user session. Result sets will not be stored following the end of a session unless the user explicitly saves them before logging off.	2
ANA_6	The system shall maintain a library of common and popular open source analytical software.	2
ANA_6a	The system shall provide access to sequence alignment tools.	2
ANA_6b	The system shall provide access to protein structure visualization tools.	3
ANA_6c	The system shall provide access to gene expression analysis tools.	2
ANA_6d	The system shall provide access to haplotype analysis tools.	3
ANA_6e	The system shall provide access to phylogenetics analysis tools.	3
ANA_6f	The system shall provide access to epitope analysis tools.	3
ANA_6g	The system shall provide access to gene prediction tools (e.g., GRAIL, GeneScan).	3
ANA_6h	The system shall provide access to exon and open reading frame prediction tools (e.g., GLIMMER).	3
ANA_7	The system shall support data reformatting to meet the data format requirements for submission to analysis tools integrated within the system.	3
ANA_9	The system shall provide search and analysis capabilities using a natural language interface.	4
ANA_10	The system shall enable the user to interactively access data through views of the data tailored to specific research and analysis requirements.	2
ANA_11	The system shall provide one or more data marts with gene centric data view.	2
ANA_12	The system shall provide one or more data marts with a phenotypic data view.	2
ANA_13	The system shall allow users to tailor the output display.	3
ANA_14	The system shall deliver tools and methods that allow non-experts to use information.	2
ANA_15	The system shall enable the user to perform calculations and modeling to create data ranges or manipulate data across dimensions (i.e., organized attributes that relate to the user's perception of the data), through hierarchical levels of data and/or across descriptive identifiers of the data. Examples of dimensions are "time" or "location."	3
ANA_16	The user shall be able to perform trend analysis over sequential time periods.	3
ANA_17	The system shall enable the user to interactively determine the data dimensions displayed and their orientations in the viewing area.	3
ANA_18	The system shall support use of statistical software packages such as SAS, SPLUS, and R-Package to perform analyses such as Univariate Analysis (ANOVA) Multiple Analysis of Variance (MANOVA) Principle Component Analysis (PCA)	3
ANA_19	The system shall provide tools to compare and contrast data analysis results across phenotypes.	3

ANA_20	<p>The system shall support the algorithms for the analysis of the minimally processed data from microarray experiments, including:</p> <ul style="list-style-type: none"> ■ Normalization of microarray data ■ Differential expression analyses <ul style="list-style-type: none"> • Sorting genes by absolute difference, relative difference and p-values of differences ■ Statistical filtering ■ Clustering analysis <ul style="list-style-type: none"> • Hierarchical clustering • K- Clustering ■ Scatter plots ■ Determination of probability of gene ontology co-clustering by chance <ul style="list-style-type: none"> • CLASSIFY 	1
ANA_21	<p>The system shall support algorithms used for disease association analysis of the genetic data, including:</p> <ul style="list-style-type: none"> ■ Testing of all SNPs for disease association <ul style="list-style-type: none"> • Chi squared test • Significance of the results 	3
ANA_22	<p>The system shall support algorithm/software used for the selection of informative SNPs (tagSNPs) from a dense network of genotyping, including:</p> <ul style="list-style-type: none"> ■ Algorithm based on haplotype structures <ul style="list-style-type: none"> • SNPTagger • HapBlock ■ Algorithm based on patterns of linkage disequilibrium between common SNPs <ul style="list-style-type: none"> • Ldselect 	2
ANA_23	<p>The ImmPort system shall include a local genome and/or table browser to view the associated data as well download and filter the data of interest.</p>	1
ANA_24	<p>The system shall enable the user to compare data analysis results among different populations.</p>	2
ANA_25	<p>The system shall enable the user to compare data analysis results across different experimental techniques/platforms.</p>	2

Table 10-9 lists data analysis requirements for the ImmPort system, with a priority for each.

11.0 ONTOLOGY

An ontology is a formal explicit description of concepts in a domain of discourse (classes), the properties of each concept describing various attributes of that concept (properties or attributes), and the inheritable role relationships between concepts. Classes are the focus of an ontology and they describe concepts in the domain. A class can have subclasses that represent concepts that are more specific than the superclass based on a true “is a” relationship. An ontology, together with a set of individual instances of classes, constitutes a knowledge base. Developing an ontology includes defining classes in the ontology, arranging the classes in a taxonomic (subclass–superclass) hierarchy. The ontology defines relationships and describes allowed values for these relationships, filling in the values for relationships. A knowledge base can then be created by defining individual instances of these classes and by filling in specific relationship value information and relationship restrictions.

An ontology is developed to share a common understanding of the structure of information among people and software agents, to enable reuse of domain knowledge, to make domain assumptions explicit, to separate domain knowledge from the operational knowledge, as well as to analyze domain knowledge. Sharing a common understanding of the structure of information among people or software agents is one of the more common goals in developing an ontology. Computer agents can extract and aggregate information if the same underlying ontology of the terms is used to answer user queries or as input data to other applications. Table 11-1 provides a summary of the ImmPort system ontology-related requirements.

Version 1.0 of the ImmPort Ontology will be implemented as part of ImmPort Version 1.0, along with a web-based browser to view the contents of the ImmPort Ontology.

Refer to Section 3.3 Product Priorities for a description of the requirement prioritization numbering scheme used in this section.

11.1 ONTOLOGIES AND CONTROLLED VOCABULARIES

This ontology will be based on the National Cancer Institute (NCI) Thesaurus. During the 2-year base POP, the development focus will be on extending the NCI Thesaurus in the research domains of human immunology and genetics related to the human immune system.

Table 11-1. Ontologies and Controlled Vocabularies Requirements

ID	Requirement	Priority
OCV_1	The ImmPort Ontology shall provide an immunology focus.	1
OCV_2	The ontology shall provide a standardized vocabulary. (A “standardized” vocabulary indicates that a preferred specification is designated to refer to a concept. Other specifications similar to the preferred specification are designated as being synonyms.)	1
OCV_3	The ontology shall provide a thesaurus for its concepts through use of synonyms.	1
OCV_4	The user shall be able to select terminology from the ontology when constructing a data query in the public data warehouse.	2
OCV_5	The system shall expand the query entered by the user using synonyms in the ontology to bridge the gap (i.e., find relationships) between the user’s choice of terminology and the terminology used in the system.	2
OCV_6	The ontology shall be available to specify data types during the data batch load process.	2
OCV_7	The user shall be able to use the results from an ontology search to search the public data warehouse.	2
OCV_8	The user shall be able to browse the ontology over the Web	1
OCV_9	The user interface shall provide the ability to view the tree hierarchy in the ontology structure.	1
OCV_10	The user interface shall have live links to facilitate navigation between classes through relationships.	1
OCV_11	The user shall be able to search the ontology by class name using exact spelling.	1
OCV_12	The user shall be able to search the ontology by class name using partial spelling (with or without	3

	wild card)	
OCV_13	The user shall be able to search the ontology by property and corresponding value.	1
OCV_14	The user shall be able to search the ontology using up to 2 properties and a corresponding value for each property. (For example, a user may want to search a chromosome location and a disease at the same time.)	2
OCV_15	The user shall be able to search the ontology by attribute and corresponding value.	1
OCV_16	The ontology browser shall load within 1 minute of the user's request.	1
OCV_17	The ontology browser shall refresh within 12 seconds of the user's request.	1
OCV_18	The user interface shall provide the ability to locate a concept in the tree hierarchy within the ontology.	2
OCV_19	The user interface shall provide links out to concepts in other ontologies.	4

Table 11-1 lists ontologies and controlled vocabularies requirements for the ImmPort system.

11.2 AGGREGATION AND STANDARDIZATION OF EXTERNAL DATA

Based on the data elements from the experimental data models and the array-specific portion of the Microarray Gene Expression Data (MGED) ontology, the ImmPort Ontology will provide conceptual data elements for experiments using research techniques used by DAIT-funded research groups, beginning with the Population Genetics program. The ImmPort Ontology will provide terms that can be used to populate menus used on data submission forms for various types of experimental data. The mapped ontologies and vocabularies will be deployed to facilitate the integration of external and archived data into the ImmPort system research database.

Table 11-2. Aggregation and Standardization of External Data Requirements

ID	Requirement	Priority
ASE_1	The ontology shall provide specific conceptual data elements for the Illumina Bead Array technique for SNP analysis and genotyping	1
ASE_2	The ontology shall provide specific conceptual data elements for the GeneChip-based genotyping (Affymetrix SNP chip) technique for SNP analysis and genotyping	1
ASE_3	The ontology shall provide specific conceptual data elements for the microarray assay technique using GenePix to measure mRNA expression.	1
ASE_3.1	The ontology shall provide specific conceptual data elements for the microarray assay technique using GenePix to measure mRNA expression.	1
ASE_3.2	The ontology shall provide specific conceptual data elements for the microarray assay technique using Affymetrix GeneChip to measure mRNA expression.	1
ASE_4	The ontology shall provide specific conceptual data elements for microsatellite analysis for SNP analysis and genotyping.	2
ASE_5	The ontology shall provide specific conceptual data elements for HLA typing and other genotyping technologies for SNP analysis and genotyping.	2
ASE_6	The ontology shall provide specific conceptual data elements for the sequencing technique for SNP analysis and genotyping.	2
ASE_7	The ontology shall provide specific conceptual data elements for the flow cytometry technique to measure protein expression.	2
ASE_8	The ontology shall provide specific conceptual data elements for the ELISA technique to measure immune response.	2
ASE_9	The ontology shall provide specific conceptual data elements for the ELISPOT technique to measure immune response.	3
ASE_10	The ontology shall provide specific conceptual data elements for mass spectrometry for measuring protein expression.	3
ASE_11	The ontology shall provide specific conceptual data elements for the Luminex Bead assay to measure protein function.	3
ASE_12	The ontology shall provide specific conceptual data elements for the Cytometric Bead Immunoassay to measure protein expression.	3

ASE_13	The ontology shall provide specific conceptual data elements for the 2D-GEL technique to measure protein expression.	4
ASE_14	The ontology shall provide specific conceptual data elements for the ESI-LC to measure protein expression.	4
ASE_15	The ontology shall provide specific conceptual data elements for the PCR technique to measure mRNA expression.	4
ASE_16	The ontology shall provide specific conceptual data elements for the CHIP technique to measure DNA binding activity.	4
ASE_17	The ontology shall provide specific conceptual data elements for the EMSA technique to measure DNA binding activity.	4
ASE_18	The ontology shall provide specific conceptual data elements for functional validation using siRNA.	4
ASE_19	The ontology shall provide specific conceptual data elements for conducting experiments using transgenic or null mutant mice.	4
ASE_20	The ontology shall provide specific conceptual data elements for the Western Blot technique to quantify the function of proteins and genes.	4

Table 11-2 lists ontologies and controlled vocabularies requirements for the ImmPort system.

12.0 SEMANTIC DATA MAPPING

Semantic data mapping will be implemented during the second year of the BISC Base POP in an interim release preceding ImmPort Version 2.0.

Semantic data mapping will support the capability to link concepts in the semantic layer (the ontology) to corresponding data items in the data warehouse, as well as to external repositories and data stores. The semantic model provides a conceptual context to the data and establishes standards for data representation by specifying data types, data formats, cardinality restrictions and allowed values. These two elements of semantic data mapping—conceptual context and data standards—will support and enhance many of ImmPort’s tasks of data access, gathering and loading.

The ImmPort system will perform large-scale data integration from a large number of remote, disparate data stores operated by the DAIT-funded research community to a centralized data warehouse and topical data marts hosted by ImmPort. The system will aggregate the data from the individual research groups and store it in the ImmPort system’s data warehouse in a single unified format. When research results are submitted by researchers and projects to the data warehouse, data elements that do not comply with the standards established in the semantic model and corresponding data warehouse will be converted and transformed to that format. The ImmPort system will also enable users to access external data sources to supplement internal research data with information generated by other organizations and stored in external databases.

Semantic data mapping will play a central role as mediator between the disparate local data stores used by the DAIT-funded research projects and the data warehouse. Semantic mapping will be used to locate and transform data from the experimental techniques/platforms and formats used by the individual projects to one uniform data model backed by the ImmPort ontology.

The mapping of databases and other data stores to one semantic model enables users of the ImmPort system to perform semantically enhanced data search and retrieval. The user, relieved from the need to know the design details of the data asset, creates a conceptual query expressed in terms of concepts in the semantic model. Through a process of inference and transformation the conceptual query is translated to database queries using the appropriate database table and field names, making the table joins where necessary and ultimately generating a query statement expressed in the syntax and terms of the particular target data asset. The semantic model and a semantic inference engine enhance user’s search and query capabilities through inference navigation that enable the discovery of related concepts and values that are not explicit in the data.

The semantic data mapping will support the ImmPort tasks listed in Table 12-1 in the priority order indicated. Refer to Section 3.3 Product Priorities for a description of the requirement prioritization numbering scheme used in this section.

Table 12-1. Semantic Mapping Capabilities and Features of the ImmPort System

ID	Capability/Feature	Requirement Description	Priority
SDM_1	Data Loading	The ImmPort system shall support generation of queries and the corresponding transformations as necessary to extract information from a disparate data source to be loaded into ImmPort’s central data warehouse.	3
SDM_2	Semantic Information Search, Discovery Retrieval	The ImmPort system shall enable browsing and navigation of the ontology identifying concepts through their semantic relationships and locate corresponding data items in the various available data stores. Semantic discovery function identifies all data items related to a concept via association, reference or inheritance.	2

ID	Capability/Feature	Requirement Description	Priority
SDM_3	Metadata Repository	The ImmPort system shall support metadata repository services, holding definitions and details of data in internal and external data stores used by the community of ImmPort users or provided by them. It will provide cataloging, searching and reporting capabilities concerning data items in the repository as well as maintain validity and integrity of the repository.	3
SDM_4	Semantic Information Model Maintenance	The ImmPort system shall maintain an information model that corresponds to the ontology, including semantic constructs and restrictions as needed to maintain the link between the semantic layer and the database layer and support semantic navigation. The system will support import and reverse engineering capabilities as needed to load data assets of various types and enrich the semantic information model.	2
SDM_5	Business Rule Repository	The ImmPort system shall provide an authoring environment for business rules and business rules language as necessary to express transformation, conversions and other constructs needed to define concepts and assist in the format unification process.	3
SDM_6	Query Generation	The ImmPort system shall provide semantic query facility where users can create a query by selecting concepts and properties from the semantic model and express criteria to form a conceptual query that will be translated to queries in the native query language of the target database platform.	2
SDM_7	Query Management	The ImmPort system shall support storage retrieval and deployment of publicly and privately stored queries. Via the semantic mapping, the system will analyze and alert system administrators and users when a query is invalidated due to changes in the ontology or in a data schema.	3

Table 12-1 summarizes the ImmPort system semantic mapping capabilities and features.

13.0 USER ASSISTANCE REQUIREMENTS

13.1 ASSISTANCE

The ImmPort system will provide accessible forms of assistance to the end user. Technical assistance will also be available. User assistance requirements are listed in Table 13-1. These requirements will initially be implemented with ImmPort Version 1.0 and will be enhanced on a continuing basis during the course of the BISC Phase II contract.

Refer to Section 3.3 Product Priorities for a description of the requirement prioritization numbering scheme used in this section.

Table 13-1. User Assistance Requirements

ID	Requirements	Priority
UAR_1	The system shall provide a training guide for users. This guide will be available in formats that can be displayed in a Web browser or downloaded for printing.	2
UAR_1a	The system training guide shall address how to use functionality related to reference data queries.	1
UAR_1b	The system training guide shall address how to use functionality related to managing users.	2
UAR_1c	The system training guide shall address how to use functionality related to managing contracts/grants.	2
UAR_1d	The system training guide shall address how to use functionality related to managing programs.	2
UAR_1e	The system training guide shall address how to use functionality related to experimental data queries.	1
UAR_1f	The system training guide shall address how to use functionality related to managing private data areas.	1
UAR_1g	The system training guide shall address how to use functionality related to loading experimental data.	1
UAR_2	The system shall provide tutorials on how to use its features in formats that can be displayed in a Web browser or downloaded for printing.	3
UAR_3	The system shall enable the user to access online help.	1
UAR_4	The system shall provide the user with the capability to request technical assistance.	1

Table 13-1 lists the ImmPort system user assistance requirements.

13.2 TECHNICAL ASSISTANCE

The BISC Team shall provide technical assistance in use of the ImmPort system, including loading of experimental results into the ImmPort system and the use of tools provided by the ImmPort system for analysis of experimental data.

The BISC Team will provide staff knowledgeable in the use of the ImmPort system with bioinformatics expertise to assist users.

14.0 OTHER REQUIREMENTS

14.1 APPLICABLE STANDARDS

Table 14-1 identifies standards that apply to ImmPort. The ImmPort system shall meet all requirements and regulations set forth by the agencies, organizations, and statutes listed in Table 14-1 in Version 1.0.

Table 14-1. Applicable Standards

ID	Requirements
1.	National Institute of Health (NIH)
2	National Institute of Allergy and Infectious Diseases (NIAID)
3	Division of Allergy, Immunology, and Transplantation (DAIT)
4	The Privacy Act of 1974
5	The Freedom of Information Act (FOIA)
6	Electronic Freedom of Information Act Amendments of 1996
7	Health Insurance Portability and Accountability Act of 1996 (HIPAA) Requirements and Regulations
8	Capability Maturity Model Integration (CMMI) Level 3 Standards
9	Northrop Grumman IT, BISC Database Style Guide
10	Americans with Disabilities Act (ADA)

Table 14-1 lists agencies, organizations, and statutes that have promulgated standards that apply to the ImmPort system.

14.2 DATA SHARING FUNCTIONS

The ImmPort system will include an extensive data warehouse containing an integration of experimental data supplied by DAIT-funded investigators and data extracted for a variety of public databases. The ImmPort system will be freely accessible to all investigators funded through the DAIT as a public resource. Since all DAIT contracts and grants will stipulate that research results are to be submitted to the BISC Phase II contractor for storage in the ImmPort system, this system will become the repository of record for DAIT-funded research over the next five years. The BISC data-sharing plan is therefore an extension of the National Institutes of Health (NIH) data sharing policy and the data sharing plans of the individual investigators submitting data to the ImmPort system. Thus, submission of data to the ImmPort system meets the data sharing requirement for DAIT-funded contracts and grants.

The NIH policy on data sharing, found at the following NIH website:

“http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm”, states:

“There are many reasons to share data from NIH-supported studies. Sharing data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new data sets when data from multiple sources are combined.”

“The NIH policy on data sharing applies:

- To the sharing of final research data for research purposes.
- To basic research, clinical studies, surveys, and other types of research supported by NIH. It applies to research that involves human subjects and laboratory research that does not involve human subjects. It is especially important to share unique data that cannot be readily replicated.
- To applicants seeking \$500,000 or more in direct costs in any year of the proposed project period through grants, cooperative agreements, or contracts.

- To research applications submitted beginning October 1, 2003.”

“Final research data are recorded factual material commonly accepted in the scientific community as necessary to document, support, and validate research findings. This does not mean summary statistics or tables; rather, it means the data on which summary statistics and tables are based. For most studies, final research data will be a computerized data set. For example, the final research data for a clinical study would include the computerized data set upon which the accepted publication was based, not the underlying pathology reports and other clinical source documents. For some but not all scientific areas, the final data set might include both raw data and derived variables, which would be described in the documentation associated with the data set.”

“In NIH's view, all data should be considered for data sharing. Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data. To facilitate data sharing, investigators submitting a research application requesting \$500,000 or more of direct costs in any single year to NIH on or after October 1, 2003 are expected to include a plan for sharing final research data for research purposes, or state why data sharing is not possible.”

The NIH data sharing policy encourages the timeliness of data sharing, stating that “recognizing that the value of data often depends on their timeliness, data sharing should occur in a timely fashion. NIH expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final data set. The specific time will be influenced by the nature of the data collected. Data from small studies can be analyzed and submitted for publication relatively quickly. If data from large epidemiologic or longitudinal studies are collected over several discrete time periods or waves, it is reasonable to expect that the data would be released in waves as data become available or main findings from waves of the data are published. NIH recognizes that the investigators who collected the data have a legitimate interest in benefiting from their investment of time and effort. NIH continues to expect that the initial investigators may benefit from first and continuing use but not from prolonged exclusive use.”

To encourage early submission of experimental data to the ImmPort system and optimal use of the ImmPort system's data analysis capabilities, the system will provide at least two levels of data access. A public data warehouse will be open to all DAIT-funded investigators. The ImmPort system will also make available to each research team a PPW. Access to data deposited into a research project's PPW will be limited to the members of that project's research team and to other ImmPort system users authorized access by that research team. Submission of data to a PPW will allow a research team to integrate their preliminary experimental data with the public data housed in the main public data warehouse and to utilize the data analysis tools supported in the ImmPort system. Since access to PPWs is restricted to the submitting research team, submission of data to a PPW does not constitute “data sharing”. Transfer of data from the PPW to the public data warehouse should occur no later than the acceptance for publication of the original experimental findings resulting from the data set. Transfer of data to the main public data warehouse will meet the requirements stipulated in the NIH Data Sharing Policy and should be considered to be public disclosure for intellectual property purposes.

To further protect the intellectual property of DAIT-funded research scientists, all ImmPort system users will be asked to agree to a data sharing policy that will be worded along the lines of the following sample provided within the NIH data sharing policy:

“User registration is required in order to access or download files. As part of the registration process, users must agree to the conditions of use governing access to the public release data, including restrictions against attempting to identify study participants, destruction of the data after analyses are completed,

reporting responsibilities, restrictions on redistribution of the data to third parties, and proper acknowledgement of the data resource. Registered users will receive user support, as well as information related to errors in the data, future releases, workshops, and publication lists. The information provided to users will not be used for commercial purposes, and will not be redistributed to third parties.”

14.3 ELECTRONIC NOTEBOOK

The electronic notebook feature will be developed during the option POP of the BISC II contract.

APPENDIX A

Table A1. Experimental Techniques To Be Supported by the ImmPort System

No.	Experimental Techniques/Platforms	Population Genetics Project Group Using the Platform					
		RTI	UAB	U Wash	Mayo	deCODE	McMaster
1	ELISPOT				✓		
2	ELISA	✓	✓	✓	✓		✓
3	Sequencing		✓		✓		
4	HLA typing and other genotyping technologies	✓	✓		✓		
5	Illumina Bead Array		✓				✓
6	Microsatellite markers					✓	
7	GeneChip-based genotyping (Affymetrix SNP chip)			✓	✓		
8	Microarray (Affymetrix cDNA array, Genpax array)					✓	✓
9	Flow Cytometry (FACS)	✓	✓	✓	✓		✓
10	2D-GE	✓					
11	ESI-LC	✓					
12	Mass Spectrometry	✓					
13	PCR (not for SNP analysis)					✓	✓
14	Luminex Bead Assay	✓				✓	✓
15	Cytometric Bead Immunoassay			✓			✓
16	Western Blot			✓			✓
17	EMSA		✓	✓			✓
18	CHIP			✓			✓
19	siRNA					✓	✓
20	Transgenic or Null Mutant Mice					✓	✓

APPENDIX B PATIENT PHENOTYPE DATA

Table B1. DEMOGRAPHICS

Demographic Data Item	Sample Values
Year of birth	1969
Age at Sample	43
Gender	Male
	Female
Ethnicity	Hispanic/Latino
	Not Hispanic
Race	Native American
	African American
	Asian
	Pacific Islander
	White
	No Race
Entry Diagnosis	Rheumatoid Arthritis
	Systemic Lupus Erythematosus
	Multiple Sclerosis
	Inflammatory Bowel Disease
Residential Location	City
	State or Province
	Country

Table B1 lists demographic data items to be collected.

Table B2. FAMILY RISK FACTOR HISTORY

Relationship	RA	SLE	MS	IBD	DM	Stroke	Heart	Cancer	Smoking
Mother									
Father									
Sister-1									
Sister-2									
Sister-3									
Brother-1									
Brother-2									
Brother-3									
Daughter-1									
Daughter-2									
Daughter-3									
Son-1									
Son-2									
Son-3									
Maternal Grandmother									
Maternal Grandfather									
Paternal Grandmother									
Paternal Grandfather									
Maternal Aunt-1									
Maternal Aunt-2									
Maternal Aunt-3									
Maternal Uncle-1									
Maternal Uncle-2									
Maternal Uncle-3									
Paternal Aunt-1									
Paternal Aunt-2									
Paternal Aunt-3									
Paternal Uncle-1									
Paternal Uncle-2									
Paternal Uncle-3									
Maternal Female Cousin-1									
Maternal Female Cousin-2									
Maternal Female Cousin-3									
Maternal Male Cousin-1									
Maternal Male Cousin-2									
Maternal Male Cousin-3									
Paternal Female Cousin-1									
Paternal Female Cousin-2									
Paternal Female Cousin-3									
Paternal Male Cousin-1									
Paternal Male Cousin-2									
Paternal Male Cousin-3									
Niece-1									

Table B2 lists family history data items to be collected.

Table B3. THERAPEUTICS

Medication	Class	Subject Dose	RA	SLE	MS	IBD
Aspirin		mg/day	X	X	X	
NSAID	NSAID	Type, mg/day	X	X	X	
Coxib	COX-2	Type, mg/day	X	X	X	
Prednisone	Corticosteroid	mg/day	X	X	X	
Methylprednisolone	Corticosteroid	mg/day			X	
Plaquenil	DMARD	mg/day	X	X		
Methotrexate	DMARD	mg/day	X	X	X	
Leflunomide	DMARD	mg/day	X	X		
Sulfasalazine	DMARD	mg/day	X	X		
Mycophenolate	DMARD	mg/day	X	X		
Azathioprine	DMARD	mg/day	X	X	X	
Cyclophosphamide	DMARD	mg/day		X	X	
Cyclosporine	DMARD	mg/day			X	
Statin	CV	type	X	X		
ACE-1	CV	type	X	X		
CCB	CV	type	X	X		
Infliximab	Biologic		X			
Etanercept	Biologic		X			
Adalimumab	Biologic		X			
Rituximab	Biologic		X	X		
Anakinra	Biologic		X			
Other DMARD/Biologic						
Estrogens	Hormone				X	
Thalidomide	DMARD				X	
Plasmapheresis					X	
Glatiramer Acetate	DMARD				X	
Interferon Beta 1a -	Biologic				X	
Interferon Beta 1a	Biologic				X	
Interferon Beta 1b	Biologic				X	
IVIG	Biologic				X	
Mitoxantrone	DMARD				X	
Other						

Table B3 lists therapeutics data items to be collected.

Table B4. DISEASE ACTIVITY & HEMATOLOGY (at time of sample)

Test	Specimen	Descriptors			Autoimmune Diseases			
		Method	Normal Value	Subject Value	RA	SLE	MS	IBD
ESR	Serum	Modified Westergren	M: ≤20mm/h F: ≤30mm/h		X	X	X	X
CRP	Serum	Nephelometry	<.8mg/dL		X	X	X	X
Complement C3	Serum	Nephelometry	75-161 mg/dL			X		
Complement C4	Serum	Nephelometry	16-47 mg/dL			X		
Anti-DNA titer:	Serum	Immunoassay	<30 IU m/L			X		
HAQ Score		Interview			X			
SLEDAI		H&P; Labs				X		
DAS 28		H&P; Labs			X			
Ambulation Index		H&P					X	
Activity Index		H&P					X	X
Hematology								
WBC	Blood	Automated Hematology Analyzer	7-30 mg/dL		X	X		
PMN Percentage	Blood	Automated Hematology Analyzer			X	X	X	X
Abs Neutrophil Count	Blood	Calculation			X	X	X	X
Lymphocyte Percentage	Blood	Automated Hematology Analyzer			X	X	X	X
Abs Lymphocyte Count	Blood	Calculation			X	X	X	X
Monocyte Percentage	Blood	Automated Hematology Analyzer			X	X	X	X
Abs Monocyte Count		Calculation			X	X	X	X
Hemoglobin	Blood	Automated Hematology Analyzer	M:13.8-17.2 g/dL F:12.0-15.6 g/dL		X	X	X	X
Hematocrit	Blood	Automated Hematology Analyzer	M: 41-50% F: 35-46%		X	X	X	X
Platelet Count	Blood	Automated Hematology Analyzer	130-400x103/μL		X	X	X	X

Table B4 lists disease hematology (at time of sample) data items to be collected.

Table B5. DISEASE ACTIVITY - NEPHROLOGY/LFTs (at time of sample)

Test	Descriptors				Autoimmune Diseases			
	Specimen	Method	Normal Value	Subject Value	RA	SLE	MS	IBD
Serum Creatinine	Serum	Colorimetry	M: 0.7-1.3mg/dL F:0.6-1.1mg/dL	mg/dL	X	X		
BUN	Serum	Colorimetry	7-30 mg/dL		X	X		
Dipstick Proteinuria*			Absent	Negative 1+/30mg/dL 2+/100 mg/dL 3+/300mg/dL 4+/2000mg/dL	X	X		
Urinary WBC*		Microscopy	Absent	0-2, 3-5, 6-10, 11-20, 21-40		X		
Urinary RBC*		Microscopy	Absent	0-2, 3-5, 6-10, 11-20, 21-40		X		
Urinary Casts*		Microscopy	Absent	0-2, 3-5, 6-10, 11-20, 21-40		X		
24 Hour Protein						X		
Creatinine Clearance	Serum, Urine		M:82-125 mL/Min F:75-115 mL/Min			X		
AST	Serum	Enzymatic Colormetry	<42 U/L		X	X		
ALT	Serum	Enzymatic Colormetry	<48U/L		X	X		
Albumin	Serum	Enzymatic Colormetry	3.5-4.0 g/dL		X	X		

Table B5 lists nephrology/LFT disease activity (at time of sample) to be collected.

*Classification ranges given in "subject value" column

Table B6. HEMATOLOGIC FEATURES

Test	Definition	Present in past	Present Now	Never Had	Unknown	RA	SLE	MS	IBD
Lymphopenia	<1500/mm ³						X	X	
Leukopenia	<4000/mm ³						X	X	
Thrombocytopenia	<100,000/mm ³						X		
Hemolytic Anemia							X	X	
Normocytic anemia						X	X	X	
Iron Deficiency Anemia						X	X	X	
Lymphadenopathy							X		

Table B6 lists hematologic features to be collected.

Table B7. RENAL FEATURES

Test	Present in past	Present Now	Never Had	Unknown	RA	SLE	MS	IBD
Hx Hematuria						X		
Hx Proteinuria						X		
Type IIa GN						X		
Type IIb GN						X		
Type III GN						X		
Type IV GN						X		
Type V GN						X		
Dialysis						X		
Renal Transplant						X		
Non-Lupus CKD					X	X		

Table B7 lists renal features to be collected.

Table B8. DERMATOLOGY FEATURES

Test	Present in past	Present Now	Never Had	Unknown	RA	SLE	MS	IBD
Malar Rash						X		
Photosensitive Rash						X		
Subacute Cutaneous LE						X		
Discoid LE						X		
Oral/Genital Ulcers						X		
Non-Scarring Alopecia						X		
Psoriasis					X	X		
Livedo Reticularis						X		
Nail Pitting					X			

Table B8 lists dermatology features to be considered.

Table B9. RHEUMATOLOGY FEATURES

Test	Present in past	Present Now	Never Had	Unknown	RA	SLE	MS	IBD
Keratoconjunctivitis Sicca						X		
Rheumatoid Nodules					X			
Erosive Peripheral Arthritis					X			
Non-erosive Peripheral Arthritis					X	X		
Raynaud's Phenomenom						X		
Sacroiliitis					X			
Serositis						X		
Vasculitis						X		
Polymyositis						X		
Dermatomyositis						X		
Fibromyalgia					X	X		

Table B9 lists rheumatology features to be collected.

Table B10. CONCOMITANT ILLNESSES

Test	Present in Past	Present Now	Never Had	Unknown	RA	SLE	MS	IBD
Smoking					X	X		
Hypertension					X	X		
CAD/MI					X	X		
Stroke/TIA					X	X	X	
Type I Diabetes					X	X		
Type II Diabetes					X	X		
Cancer					X	X		
DVT					X	X		
PE					X	X		
Spontaneous Abortion						X		
Thyroid disease						X		

Table B10 lists concomitant illnesses to be collected.

Table B11. CNS FEATURES

Test	Present in Past	Present Now	Never Had	Unknown	RA	SLE	MS	IBD
Lupus HA/Aseptic Meningitis						X		
Depression					X	X	X	
Seizure						X	X	
Cognitive Dysfunction						X	X	
Cranial Neuropathy						X		
Trigeminal Neuralgia							X	
Psychosis						X		
Tuberculosis					X	X		
MRI Showing:							X	
Gadolinium Enhancement							X	
Infratentorial Lesion							X	
Juxtacortical Lesion							X	
Periventricular/Ovoid							X	
Sagittal (Dawson fingers)							X	
Cerebrospinal Fluid							X	
Elevated IgG Index							X	
Oligoclonal Bands							X	
Low Glucose							X	
High Protein							X	
Leukocytes							X	
Erythrocytes							X	

Table B11 lists CNS features to be collected.

Table B12. ID FEATURES

Test	Ever had	Present	Absent	Unknown	RA	SLE	MS	IBD
HIV 1 Infection							X	
HIV 2 Infection							X	
HTLV I Infection							X	
HTLV II Infection							X	
Vaccination							X	
Diphtheria							X	
Hepatitis B							X	
Influenza							X	
Measles							X	
Meningococcal Disease							X	
Lyme Disease							X	
Mumps							X	
Pertussis							X	
Pneumococcal Polysaccharide							X	
Polio							X	
Rubella							X	
Tetanus							X	
Varicella							X	

Table B12 lists ID features to be collected.

Table B13. SEROLOGY & LABS

Test	Titer/ Units	Present in Past	Present Now	Never Had	Unknown	RA	SLE	MS	IBD
Antinuclear Antibody						X	X	X	
ANA titer: or, ANA units	1:								
ANA pattern: diffuse						X	X	X	
peripheral									
speckled									
Anti-DNA							X		
Anti-Ro							X		
Anti-La							X		
Anti-Sm							X		
Anti-RNP							X		
Anti-CCP						X	X		
Rheumatoid Factor						X	X		
Scl-70							X		
Anti-Centromere							X		
Anti-Jo1/Mi2/Srp						X	X		
Anti-Smooth Muscle						X	X		
Anti-LKM						X	X		
False Pos RPR							X		
Lupus Anticoagulant							X		
APLA IgG							X		
APLA IgM							X		
APLA IgA							X		
ACA IgG							X		
ACA IgM							X		
ACA IgA							X		
B2GPI IgG							X		
B2GPI IgM							X		
B2GPI IgA							X		
Hypocomplementemia							X		
HLA-B27						X	X		
Elevated TSH								X	
Low Vitamin B12								X	
Hep C Antibody							X		
Hep B Core Ab							X		
Hep B sAntigen							X		
Hep B sAntibody							X		
Abnormal AST/ALT							X		

Table B13 lists data items on serology and labs to be collected.

Table B14. DAS28

Test	Right		Left		RA	SLE	MS	IBD
	Swollen	Tender	Swollen	Tender	X			
Shoulder					X			
Elbow					X			
Wrist					X			
MCP1					X			
MCP2					X			
MCP3					X			
MCP4					X			
MCP5					X			
PIP1					X			
PIP2					X			
PIP3					X			
PIP4					X			
PIP5					X			
Knee					X			
CRP value					X			
VAS					X			

Table B14 lists DAS28 data items to be collected.

Table B15. SLEDAI (SLE patients only)

SLEDAI-2K evaluation	Value	Definition
Seizure	8	Recent onset; exclude metabolic, infectious or drug causes
Psychosis	8	Hallucinations, incoherence, marked illogical thinking, bizarre, disorganized or catatonic behavior. Exclude uremia and drug causes
Organic brain syndrome	8	Altered mental function with impaired orientation, memory or other intellectual function with rapid onset and fluctuating clinical features. Inability to sustain attention to the environment plus at least 2 of the following: perceptual disturbance, incoherent speech, insomnia or daytime drowsiness, or increased or decreased psychomotor activity. Exclude metabolic, infectious or drug causes.
Visual disturbances	8	Retinal changes of SLE. Include cytoid bodies, retinal hemorrhages, serious exudate of hemorrhages of the choroid, or optic neuritis. Exclude hypertension, infection or drug causes.
Cranial nerve disorder	8	New onset of sensory or motor neuropathy involving cranial nerve
Lupus headache	8	Severe, persistent headache; may be migranous, but must be responsive to narcotic analgesia
CVA	8	New onset of cerebrovascular accident(s). Exclude arteriosclerosis.
Vasculitis	8	Ulceration, gangrene, tender finger nodules, periungual infarction, splinter hemorrhages, or biopsy or angiogram proof of vasculitis.
Arthritis	4	> 2 joints with pain and signs of inflammation
Myositis	4	Proximal muscle aching /weakness , associated with elevated creatine phosphokinase/aldolase or electromyogram changes or a biopsy showing myositis
Urinary casts	4	Heme-granular or red blood cell casts
Hematuria	4	>5 red blood cells/high power field. Exclude infection.
Proteinuria	4	>0.5 grams /24 hours
Pyuria	4	>5 white blood cells/high power field. Exclude infection.
New Rash	2	New onset or recurrence of inflammatory type rash.
Alopecia	2	New onset or recurrence of abnormal, patchy or diffuse hair loss
Mucosal Ulcers	2	New onset or recurrence of oral or nasal ulcerations
Pleurisy	2	Pleuritic chest pain with pleural rub or effusion, or pleural thickening.
Pericarditis	2	Pericardial pain with at least 1 of the following: rub, effusion or electrocardiogram confirmation.
Low Complement	2	Decrease in CH50, C3, or C4 below lower limit for testing lab.
Increased DNA binding	2	>25% binding by Farr assay or above normal range for testing laboratory.
Fever	1	>38° C. Exclude infection
Thrombocytopenia	1	<100,000 platelets/mm3
Leukopenia	1	<3,000 White blood cell/mm3. Exclude drug causes.

Table B15 lists SLEDAI data items to be collected.

APPENDIX C USE CASES

This section includes *essential use cases* (i.e., simplified, implementation-free descriptions of a user’s interactions with the system) that have been developed to further identify and clarify system requirements. The use cases have been packaged according to system functionality, addressing needs of users from the DAIT community and the DAIT-funded research community. Use cases dealing with system functionality of relevance primarily for the ImmPort development team (e.g., managing system performance reports) have not been developed. Additional use cases will be developed as needed, either for a future version of this SSRD or in support of design activities.

1. MANAGE USERS (MU)

These use cases describe common activities involved in managing access of users to the ImmPort system, starting from the registration process (required of an individual requesting system access) and ending with the deactivation of a user (revoking the user’s access to the ImmPort system).

Table C1. Manage Users

Use Case ID	Description
MU-1	<i>Register User</i> describes the flow of events for a user (other than PO) to register in ImmPort.
MU-2	<i>Approve or Reject Registration</i> describes the flow of events for approving or rejecting a user registration request.
MU-3	<i>Create User</i> describes the flow of events for creating a new user.
MU-4	<i>Query User</i> describes how a user searches for another user in the system.
MU-5	<i>View or Update User</i> describes the flow of events for viewing or updating a user’s personal information.
MU-6	<i>Delete User</i> describes how an authorized user deactivates a user, revoking the user’s access to ImmPort.

Table C1 lists use cases under *Manage Users*.

2. LOG IN/OFF (LG)

The log in/log off procedures describe identification, authentication, and sign-off. A use case for situations in which a user forgets his/her login information is included.

Table C2. Log In/Off

Use Case ID	Description
LG-1	<i>User Login</i> describes the login identification and authentication process.
LG-2	<i>Retrieve Login Information</i> describes how a user obtains his/her ID and authenticator to regain access to ImmPort.
LG-3	<i>Log Off</i> describes the sign-off process.

Table C2 lists use cases under *Log In/Off*.

3. MANAGE CONTRACT/GRANT (CG)

Users from the DAIT-funded research community (see Section 12.4 for an explanation of this community) must be associated with at least one DAIT contract/grant. These use cases describe activities needed to manage information on contracts/grants.

Table C3. Manage Contract/Grant

Use Case ID	Description
CG-1	<i>Create Contract/Grant</i> describes how an authorized user enters information for a contract/grant not yet in the system.
CG-2	<i>Delete Contract/Grant</i> describes how an authorized user deletes a contract/grant in the system.
CG-3	<i>Modify Contract/Grant</i> describes how an authorized user changes information about a contract/grant already entered in the system.
CG-4	<i>Search for Contract/Grant</i> describes how an authorized user searches for a specific contract/grant in the system.

Table C3 lists use cases under *Manage Contract/Grant*.

4. MANAGE PROGRAMS (MP)

Grants/contracts may be grouped with other grants/contracts for administrative purposes into programs (e.g., Population Genetics, HLA), as determined by NIAID/DAIT. These use cases describe activities needed to manage information on programs.

Table C4. Manage Programs

Use Case ID	Description
MP-1	<i>Create Program</i> describes how an authorized user enters new information for a program in the system.
MP-2	<i>Delete Program</i> describes how an authorized user deletes a program in the system.
MP-3	<i>Modify Program</i> describes how an authorized user changes information about a program already stored in the system.
MP-4	<i>Search for Program</i> describes how an authorized user searches for a specific program in the system.

Table C4 lists use cases under *Manage Programs*.

5. USE PUBLIC DATA (UP)

As described in the system requirements, the scientific data stored in ImmPort can be classified as public, available to all ImmPort users, or private, available only to selected ImmPort users. This use case describes user interactions with public data.

Table C5. Use Public Data

Use Case ID	Description
UP-1	<i>View Public Data</i> describes the flow of events for a user to view data stored in the public data warehouse.
UP-2	<i>Query Public Data</i> describes the flow of events for a user to view data stored in the public data warehouse.
UP-3	<i>Download Public Data</i> describes the flow of events for a user to download data stored in the public data warehouse.
UP-4	<i>View Ontology</i> describes the flow of events for a user to view the ImmPort ontology.
UP-5	<i>Query Ontology</i> describes the flow of events for a user to query the ImmPort ontology.
UP-6	<i>Use Ontology To Query Public Data</i> describes the flow of events for using ontological terms in formulating a query on data in the public data warehouse.

Table C5 lists use cases under *Use Public Data*.

6. MANAGE PRIVATE DATA (PD)

Users will be able to selectively share their pre-publication data within ImmPort. These use cases describe common user activities devoted to managing private data, including movement of the data (e.g., loading data into a PPW, publishing the data to the public data warehouse) and controlling user access.

Table C6. Manage Private Data

Use Case ID	Description
PD-1	This use case describes the workflow of adding a user to the list of ImmPort users allowed to access a research project.
PD-2	This use case describes the workflow of revoking the research project access from some external user.
PD-3	<i>Publish Data to Public Data Area</i> describes the process of transferring private data to the public data warehouse, where it is accessible to all ImmPort users.

Table C6 lists use cases under *Manage Private Data*.

7. USE PRIVATE DATA (UD)

These use cases describe how users may use private data.

Table C7. Use Private Data

Use Case ID	Description
UD-1	<i>View Private Data</i> describes the flow of events for a user to view data within a PPW.
UD-2	<i>Query Private Data</i> describes the flow of events for a user to query data within a PPW.

Table C7 lists use cases under *Use Private Data*.

Use Case ID	MU-1	
Use Case Name	Register User	
Created By	Rene Pineda	Date: 03/10/2005
Updated By	Mark Dela Cruz	Date: 05/16/2005
Actors	Any individual seeking access to ImmPort	
Description	The <i>Register User</i> describes the flow of events for the actor to register in ImmPort.	
Preconditions	There are no preconditions.	
Postconditions	The user registration request is saved. Notification is sent to the BISC Project Officer and Functional Administrator for approval.	
Assumptions	Any individual can apply to register.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The actor selects to register. 2. The system prompts for the information as specified in the security system access requirements section of the SSRD. 3. The system displays its data sharing policy and requires the user to accept the terms and conditions. 4. The system prompts the actor to select a secret question (e.g., "What is your favorite pet name?") and provide an answer. This information will be used in case the user forgets his password (see <i>LG-3 Retrieve Login Information</i>). 5. The system validates the required fields are not null. 6. The system creates a registration event using the specified information, including a system-generated registration request date. 7. The system sends a notification to the BISC Project Officer and Functional Administrator for review and approval. 	
Alternate Course	<p>Cancel (after step 2)</p> <ol style="list-style-type: none"> 1. The actor selects to cancel the operation. 2. The registration request is not saved. <p>Actor may already be a user in the system (i.e., has system access capabilities) (after step 2)</p> <ol style="list-style-type: none"> 1. The system determines that the actor may already be a user in the system. 2. The system asks the actor to confirm whether he/she has an account under a given user name. 3. The actor confirms that he/she is a user <ol style="list-style-type: none"> a. The actor confirms that he/she is already a user b. <<Invoke>> <i>LG-1 Log In User</i>. 4. The actor indicates he/she is not a user. <ol style="list-style-type: none"> a. The actor indicates he/she is not a user. b. Return to step 3. 	
Exceptions	<p>Invalid data (after step 4)</p> <ol style="list-style-type: none"> 1. The first name, last name, and/or email address are null. 2. The system marks the corresponding fields as invalid, while allowing the actor to correct the errors. 	

Included Use Cases	LG-1 Log In User
Extended Use Cases	None
Priority	1
Frequency of Use	TBD
Business Rules	TBD

Use Case ID	MU-2	
Use Case Name	Approve or Reject Registration	
Created By	Rene Pineda	Date: 03/10/2005
Updated By	Mark Dela Cruz	Date: 06/29/2005
Actors	BISC Project Officer, Functional Administrator	
Description	The <i>Approve or Reject Registration</i> describes the flow of events for approving or rejecting user registration request.	
Preconditions	<ol style="list-style-type: none"> 1. An individual has completed registration. See <i>MU-1 Register User</i>. 2. The user is authenticated as DAIT BISC Project Officer or Functional Administrator. 	
Postconditions	<ol style="list-style-type: none"> 1. The registration request is approved or rejected. 2. The system informs the individual of the approval/rejection decision. 	
Assumptions	None.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The actor indicates desire to review a registration request. 2. The system validates the user is authorized. 3. The validated actor selects to approve. <ol style="list-style-type: none"> a. The registration request is approved. b. <<Invoke>> <i>MU-3 Create User</i>. 4. The validated actor selects to reject. <ol style="list-style-type: none"> a. The registration request is marked for rejection. 5. The system sends a notification to the individual who submitted the registration request and the PI that he/she designated in the registration information. 	
	Cancel (after step 2) <ol style="list-style-type: none"> 1. The actor selects to cancel the operation. 2. The registration request remains marked as pending approval. 	
Exceptions	None	
Included Use Cases	MU-3 Create User	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	MU-3	
Use Case Name	Create User	
Created By	Rene Pineda	Date: 03/10/2005
Updated By	Mark Dela Cruz	Date: 04/04/2005
Actors	BISC Project Officer, Functional Administrator	
Description	The <i>Create User</i> describes the flow of events for creating a new ImmPort user.	
Preconditions	<ol style="list-style-type: none"> 1. The user has never been created. 2. The corresponding registration request has been approved (see <i>MU-2 Approve or Reject User Registration</i>). 3. The grant/contract that the user will be associated with must exist and be active, provided the account is for a user belonging to the DAIT-funded research community. 4. The user is authenticated as DAIT BISC Project Officer or Functional Administrator. 	
Postconditions	The user is created.	
Assumptions	“The system prompts...” (See step 2 in the Normal Course) means the system asks for the information specified, in the case of manual user creation. However, when this use case gets instantiated from the approved registration process (see <i>MU-2 Approve or Reject Registration</i>), the information is submitted automatically.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The actor selects to create a new user account. 2. The system prompts for the information as specified in the security system access requirements section of the SSRD. 3. The system validates the required fields are not null. 4. The new user account is created using the supplied data, including a system-generated creation date. 	
Alternate Course	Cancel <ol style="list-style-type: none"> 1. The actor selects to cancel the operation. 2. The user account is not created. 	
Exceptions	Invalid data <ol style="list-style-type: none"> 1. Either the first name, last name, and/or email address are null 2. The system marks the corresponding fields as invalid, while allowing the actor to correct the errors. 	
Included Use Cases	None	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	MU-4	
Use Case Name	Query User	
Created By	Wenjie Hua	Date: 03/14/2005
Updated By	Mark Dela Cruz	Date: 04/04/2005
Actors	Any user	
Description	The <i>Query User</i> describes how the actor queries for information on users.	
Preconditions	The actor is logged on.	
Postconditions	A list of users with the primary information is displayed.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The actor indicates desire to query users. 2. The system provides the actor with choices for query criteria. 3. The actor enters query criteria data. <ol style="list-style-type: none"> a. Last Name b. First Name c. NIAID Contract/Grant Number 4. The system validates the input data. One of the following field is required: <ol style="list-style-type: none"> a. Last Name b. NIAID Contract/Grant Number 5. The system runs the query. 6. The system displays the query result. <ol style="list-style-type: none"> a. Last Name b. First Name c. Middle Initial d. E-mail 	
Alternate Course	<p>Cancel (after step 2)</p> <ol style="list-style-type: none"> 1. The actor selects to cancel the query. 2. The system aborts query; no results displayed. 3. The system returns the actor to the main query user page. <p>No data found (after step 6)</p> <ol style="list-style-type: none"> 1. The system displays a message that the query has yielded no search results. 2. The system lists no results. 3. The actor has the option to conduct a new search. 	
Exceptions	<p>Invalid search criteria</p> <ol style="list-style-type: none"> 1. The system displays error message. 2. The actor has the option to conduct a new search. <p>Query/data error</p> <ol style="list-style-type: none"> 1. The system displays error message. 2. The actor has the option to conduct a new search. 	
Included Use Cases	TBD	
Extended Use Cases	TBD	
Priority	1	
Frequency of Use	TBD	

Business Rules	TBD
----------------	-----

Use Case ID	MU-5	
Use Case Name	View or Update User	
Created By	Rene Pineda	Date: 03/10/2005
Updated By	Mark Dela Cruz	Date: 04/04/2005
Actors	Any user	
Description	The <i>View or Update User</i> describes the flow of events for viewing or updating a user's own personal information.	
Preconditions	The user has been previously created.	
Postconditions	The user information is updated, provided the user selected to update.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The user selects to view or update the user information. 2. The system validates the actor performing the request is authorized. 3. The system allows the user to view or update the following information: <ol style="list-style-type: none"> a. Last Name b. First Name c. Middle Initial d. Organization e. Phone Number f. Email Address <p>All other fields as specified in the security system access requirements section of the SSRD are displayed in read-only format.</p> <p>If the actor selects to update,</p> <ol style="list-style-type: none"> a. The system validates the first name, last name, and email address are not null b. The account is updated using the supplied information 	
Alternate Course	<p>Cancel (after step 2)</p> <ol style="list-style-type: none"> 1. The actor selects to cancel the operation. 2. The user account is not updated. 	
Exceptions	<p>Invalid data</p> <ol style="list-style-type: none"> 1. Either the first name, last name, and/or email address are null 2. The system marks the corresponding fields as invalid, while allowing the actor to correct the errors. 	
Included Use Cases	None	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	MU-6	
Use Case Name	Deactivate Users	
Created By	Wenjie Hua	Date: 03/14/2005
Updated By	Mark Dela Cruz	Date: 06/29/2005
Actors	BISC Project Officer or Functional Administrator	
Description	The <i>Delete Users</i> describes how the actor deactivates the user(s) selected from a query result list.	
Preconditions	<ol style="list-style-type: none"> 1. The actor is logged on. 2. A list of users with the primary information are displayed for the actor to select from: <ol style="list-style-type: none"> a. Last Name b. First Name c. Middle Initial d. NIAID Contract/Grant Number 3. The user is authenticated as DAIT BISC Project Officer or Functional Administrator. 	
Postconditions	The user(s) that the actor selects is/are deactivated successfully.	
Assumptions	None.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The actor selects the user(s) from the query result list to deactivate. 2. The actor confirms desire to deactivate the selected user(s). 3. The system deactivates user(s). 	
Alternate Course	<ol style="list-style-type: none"> 1. The actor cancels the deactivate action (after step 1). 2. The selected user(s) are not deactivated. 	
Exceptions	<p>The actor is a PI, and the user to be deleted is associated with other research grants/contracts with other PIs</p> <ol style="list-style-type: none"> 1. The system disassociates the user from only those research contracts/grants associated with the PI actor. 2. The user is still associated with other research grants/contracts with other PIs. 	
Included Use Cases	TBD	
Extended Use Cases	TBD	
Priority	1	
Frequency of Use	TBD	
Business Rules	After ImmPort v.1.0, the Principal Investigator will have the ability to only invoke the <i>Deactivate Users</i> use case on staff who are associated to his/her research contract/grant. (See SUA_21).	

Use Case ID	LG-1	
Use Case Name	Log in User	
Created By	Patty Berger	Date: 03/24/2005
Updated By	Mark Dela Cruz	Date: 04/04/2005
Actors	Any user	
Description	The <i>Log In User</i> describes the flow of events for an actor when logging into the ImmPort system.	
Preconditions	The user is not signed into the ImmPort system.	
Postconditions	The user is authenticated and will access the ImmPort system.	
Assumptions	The user is registered and authorized in ImmPort.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The user selects to login to ImmPort. 2. The system displays the security alert banner. 3. The user selects ok to continue. 4. The system prompts for the username and password. 5. If user enters username/password. <ol style="list-style-type: none"> a. The system validates the username and password for not null and authenticity. 6. If the user selects forgot your password and/or username. <ol style="list-style-type: none"> a. <<Invoke>> <i>LG-3 Retrieve Login Information</i>. 	
Alternate Course	Cancel <ol style="list-style-type: none"> 1. The user selects to cancel the operation. 2. The system returns user to the login screen. 	
Exceptions	<ol style="list-style-type: none"> 1. An error message will be displayed if the username and/or password are incorrect. 2. An error message will be displayed if username and/or password are null. 	
Included Use Cases	LG-3 Retrieve Login Information	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	LG-2	
Use Case Name	Log off	
Created By	Patty Berger	Date: 03/17/2005
Updated By	Mark Dela Cruz	Date: 04/04/2005
Actors	Any user	
Description	The <i>Log off</i> describes the flow of events for when a user is logging off of ImmPort	
Preconditions	The user is logged into the ImmPort system.	
Postconditions	The user will be logged off of the ImmPort system.	
Assumptions	The user is registered and authorized in ImmPort.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The user selects to log off. 2. If the user has unsaved work <ol style="list-style-type: none"> a. The system prompts user to save unsaved work. b. The user saves work. 3. The system logs off the user. 4. The system confirms successful log off. 	
Alternate Course	Cancel (before step 3) <ol style="list-style-type: none"> 1. The user selects to cancel the operation. 2. The system does not log off the user. 	
Exceptions	None	
Included Use Cases	None	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	LG-3	
Use Case Name	Retrieve Login Information	
Created By	Patty Berger	Date: 03/17/2005
Updated By		Date: 04/04/2005
Actors	Any user	
Description	The <i>Retrieve Login Information</i> describes the flow of events for when a user has forgotten his/her username or password.	
Preconditions	The user is not signed into the ImmPort system.	
Postconditions	The user will receive a notification of username or password via email.	
Assumptions	The user is registered and authorized in ImmPort.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The user indicates he/she has forgotten his/her username or password. 2. If the user has forgotten his/her password <ol style="list-style-type: none"> a. The system prompts for the username b. The system prompts for a secret question (i.e., "What is your favorite pet name?") 3. If the user has forgotten the username, the system prompts for the following fields: <ol style="list-style-type: none"> a. First Name b. Last Name c. Middle Initial d. Contract/Grant Number e. Principal Investigator f. The system prompts for a secret question (i.e., "What is your favorite pet name?") 4. The user enters the required data. 5. The system validates the user's response and submits a notification to the user, which includes the correct username and password, provided a match is found. 	
Alternate Course	Cancel (before step 5) <ol style="list-style-type: none"> 1. The user selects to cancel the operation. 2. The system returns user to the login screen. 	
Exceptions	<ol style="list-style-type: none"> 1. An error message will be displayed if the username and/or password are not in database or incorrect. 2. An error message will be displayed if username and/or password are null. 	
Included Use Cases	None	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	CG-1	
Use Case Name	Create Contract/Grant	
Created By	Justin Chiu	Date: 3/10/2005
Updated By	Mark Dela Cruz	Date: 5/16/2005
Actors	DAIT BISC Project Officer, Functional Administrator, System Administrator	
Description	This use case describes the flow of events for the user to create a contract/grant.	
Preconditions	<ol style="list-style-type: none"> 1. The user is logged in ImmPort. 2. The user is authenticated as DAIT BISC Project Officer, Functional Administrator, or System Administrator. 	
Postconditions	A contract/grant is created.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. User selects to create a contract/grant. 2. The system displays a create contract/grant page, which contains the fields necessary to create a contract/grant. The following are some of the fields: <ol style="list-style-type: none"> a. Contract/Grant Number b. Contract/Grant Title c. Contract/Grant Abstract d. Principal Investigator e. Start Date f. End Date g. Extended Date h. Program Association i. Project Officer 3. User submits the entered information. 4. A new contract/grant is created in the system. All relevant data associated with the contract/grant is stored in database. The system displays a confirmation screen along with information for the newly created contract/grant. 	
Alternate Course	Cancel (after step 2) <ol style="list-style-type: none"> 1. User selects to cancel the operation. 2. The contract/grant is not created. 	
Exceptions	Invalid data <ol style="list-style-type: none"> 1. Data entered does not pass validation in the system. 2. User is prompted to check entered information before resubmitting. 	
Included Use Cases	None	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	A contract/grant does not have to be associated with a program.	

Use Case ID	CG-2	
Use Case Name	Deactivate Contract/Grant	
Created By	Justin Chiu	Date: 3/10/2005
Updated By	Mark Dela Cruz	Date: 5/17/2005
Actors	DAIT BISC Project Officer, Functional Administrator, or System Administrator	
Description	This use case describes the flow of events for users to deactivate a contract/grant.	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. The user is authenticated as DAIT BISC Project Officer, Functional Administrator, or System Administrator. 	
Postconditions	<ol style="list-style-type: none"> 1. A contract/grant is disabled. 2. All contract/grant related information will be subject to removal/archival. 3. If a DAIT-funded research user is a member of only this contract/grant, his/her access to ImmPort is disabled. 	
Assumptions	Contracts/grants are not deleted; rather, they are deactivated so contract-/grant-related information is kept for historical reference.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. Include <i>CG-4 Search for Contract/Grant</i>. 2. User asks to deactivate the contract/grant selected in the use case above. 3. The system prompts for confirmation of user's intent to deactivate the contract/grant. 4. The system requires the user to explain the reason for deactivation. 5. User confirms his/her intent to deactivate. 6. The system deactivates the contract/grant. The system displays a confirmation screen indicating the contract/grant has been deactivated. 	
Alternate Course	Cancel (after step 3) <ol style="list-style-type: none"> 1. User selects to cancel the operation. 2. The contract/grant is not deactivated. 	
Exceptions	None	
Included Use Cases	CG-4 Search for Contract/Grant	
Extended Use Cases	None	
Priority	2	
Frequency of Use	TBD	
Business Rules	An ImmPort user must be associated with at least one contract/grant in order to access ImmPort. A contract/grant can be reactivated.	

Use Case ID	CG-3	
Use Case Name	Modify Contract/Grant	
Created By	Justin Chiu	Date: 3/10/2005
Updated By	Justin Chiu	Date: 3/25/2005
Actors	DAIT BISC Project Officer. Functional Administrator, or System Administrator	
Description	This use case describes the flow of events for the actors to modify a contract/grant.	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. The user is authenticated as DAIT BISC Project Officer, Functional Administrator, or System Administrator. 	
Postconditions	A contract/grant is modified.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. Include use case <i>CG-4 Search for Contract/Grant</i>. 2. User asks to modify the contract/grant selected in the use case above. 3. The system displays corresponding fields in editable format. 4. User modifies contract/grant information (including member users) and submits the changes. 5. The system saves the updated information to database. The system displays a confirmation screen along with the updated information for the contract/grant. 	
Alternate Course	Cancel (after step 3) <ol style="list-style-type: none"> 1. User selects to cancel the operation. 2. The contract/grant is not modified. 	
Exceptions	Invalid data <ol style="list-style-type: none"> 1. Modified data does not pass validation in the system. 2. User is prompted to check entered information before resubmitting. 	
Included Use Cases	CG-4 Search for Contract/Grant	
Extended Use Cases	None	
Priority	2	
Frequency of Use	TBD	
Business Rules	None	

Use Case ID	CG-4	
Use Case Name	Search for Contract/Grant	
Created By	Justin Chiu	Date: 3/10/2005
Updated By	Justin Chiu	Date: 3/25/2005
Actors	DAIT BISC Project Officer, DAIT User, Functional Administrator, or System Administrator	
Description	This use case describes the flow of events for the actors to search and view contracts/grants.	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. The user is authenticated as DAIT BISC Manager, DAIT User, Functional Administrator, or System Administrator. 	
Postconditions	A search is performed and information about a contract/grant is displayed.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. User selects to search or view contracts/grants. 2. The system displays existing contracts/grants, along with search fields for the User to perform a search. 3. User enters search parameters into search fields and submits the search. 4. The system displays those contracts/grants that match the search parameters. 5. User asks to view a contract/grant from the list. 6. The system displays information about the contract/grant. The following are some of the fields: <ol style="list-style-type: none"> a. Contract/Grant Number b. Principal Investigator c. Start Date d. End Date e. Extended Date f. Program Association g. Member Users 	
Alternate Course	Browse contracts/grants (replaces steps 3 and 4) <ol style="list-style-type: none"> 1. User browses by navigating the list of existing contracts/grants. 2. The system displays existing contracts/grants. 	
Exceptions	None	
Included Use Cases	None	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	DAIT Users can only view contracts/grants that they oversee.	

Use Case ID	MP-1	
Use Case Name	Create Program	
Created By	Justin Chiu	Date: 3/10/2005
Updated By	Justin Chiu	Date: 3/24/2005
Actors	DAIT BISC Manager, Functional Administrator, or System Administrator	
Description	This use case describes the flow of events for the actors to create a program.	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. The user is authenticated as DAIT BISC Manager, Functional Admin, or System Administrator. 	
Postconditions	A program is created.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. User selects to create a program. 2. The system displays a create program page, which contains the fields necessary to create a program. The following are some of the fields: <ol style="list-style-type: none"> a. Program Name b. Program Officer c. Program Start Date d. Program End Date e. Program Extended Date f. Member Contracts/Grants 3. User submits the entered information. 4. The system validates the data entered 5. A new program is created in the system. All relevant data associated with the program is stored in database. 6. The system displays a confirmation screen along with information for the newly created program. 	
Alternate Course	Cancel action (after step 2) <ol style="list-style-type: none"> 1. User selects to cancel the operation. 2. The program is not created. 	
Exceptions	Invalid data <ol style="list-style-type: none"> 1. Data entered does not pass validation checks in system. 2. User is prompted to check entered information before resubmitting. 	
Included Use Cases	None	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	A program may exist without any member contracts/grants.	

Use Case ID	MP-2	
Use Case Name	Delete Program	
Created By	Justin Chiu	Date: 3/10/2005
Updated By	Justin Chiu	Date: 3/24/2005
Actors	DAIT BISC Manager, Functional Administrator, or System Administrator	
Description	This use case describes the flow of events for the actors to delete a program.	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. The user is authenticated as DAIT BISC Project Officer, Functional Administrator, or System Administrator. 	
Postconditions	<ol style="list-style-type: none"> 1. A program is disabled. 2. All program-related information will be subject to removal/archival. 3. Member contracts/grants of the program deleted will be updated to floating projects. 	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. Include <i>MP-4 Search for Program</i>. 2. User asks to delete the program selected in the use case above. 3. The system prompts for confirmation of User's intent to delete the program. 4. User confirms his/her intent to delete. 5. The system deletes the program. The system displays a confirmation screen indicating the program has been deleted. 	
Alternate Course	Cancel action (after step 3) <ol style="list-style-type: none"> 1. User selects to cancel the operation. 2. The program is not deleted. 	
Exceptions	None	
Included Use Cases	MP-4 Search for Program	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	Contracts/grants are allowed to be floating, i.e. they do not have to be members of any programs.	

Use Case ID	MP-3	
Use Case Name	Modify Program	
Created By	Justin Chiu	Date: 3/10/2005
Updated By	Justin Chiu	Date: 3/24/2005
Actors	DAIT BISC Manager, Functional Administrator, or System Administrator	
Description	This use case describes the flow of events for DAIT personnel to modify a program.	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. The user is authenticated as DAIT BISC Manager, Functional Administrator, or System Administrator. 	
Postconditions	A program is modified.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. Include <i>MP-4 Search for Program</i>. 2. User asks to modify the program selected in the use case above. 3. The system displays info about the program in editable format. 4. User modifies program information (including member projects) and submits the changes. 5. The system prompts the User to confirm the changes. 6. User confirms to save the changes. 7. The system saves the updated information to database. The system displays a confirmation screen along with the updated information for the program. 	
Alternate Course	<p>Cancel (after step 3)</p> <ol style="list-style-type: none"> 1. User selects to cancel the operation. 2. The program is not modified. <p>Cancel (after step 5)</p> <ol style="list-style-type: none"> 1. User selects to cancel the operation. 2. The program is not modified. 	
Exceptions	<p>Invalid data</p> <ol style="list-style-type: none"> 1. Modified data does not pass validation checks in the system. 2. User is prompted to check entered information before resubmitting. 	
Included Use Cases	MP-4 Search for Program	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	A program may exist without any member contracts/grants.	

Use Case ID	MP-4	
Use Case Name	Search for Program	
Created By	Justin Chiu	Date: 3/10/2005
Updated By	Justin Chiu	Date: 3/24/2005
Actors	DAIT BISC Project Officer, Functional Administrator, or System Administrator	
Description	This use case describes the flow of events for the actors to search and view programs.	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. The user is authenticated as DAIT BISC Manager, Functional Administrator, or System Administrator. 	
Postconditions	A search is performed and information about a program is displayed.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. User selects to search or view programs. 2. The system displays existing programs, along with search fields for the User to perform a search. 3. User enters search parameters into search fields and submits the search. 4. The system displays those programs that match the search parameters. 5. User asks to view a program from the list. 6. The system displays information about the program. The following are some of the fields: <ol style="list-style-type: none"> g. Program Name h. Program Officer i. Program Start Date j. Program End Date k. Program Extended Date l. Member Contracts/Grants 	
Alternate Course	Browse programs (replaces steps 3 and 4) <ol style="list-style-type: none"> 1. User browses by navigating the list of existing programs. 2. The system displays existing programs. 3. Return to step 5. 	
Exceptions	None	
Included Use Cases	None	
Extended Use Cases	None	
Priority	1	
Frequency of Use	TBD	
Business Rules	None	

Use Case ID	UP-1	
Use Case Name	View public data	
Created By	Vince Desborough	Date: 3/09/05
Updated By	Mark Dela Cruz	Date: 06/29/05
Actors	Any user	
Description	<p>The <i>View Public Data</i> use case describes the flow of events for a user to view publicly accessible data (which are stored in the public data warehouse). The public data warehouse will consist of both “online” and “near online” data storage. Therefore, access to these two types of data should happen automatically without the user having to enter extra commands. The user should be able to move forward or backward in the viewing state, at any time, without consequences.</p> <p>Viewing data will include viewing:</p> <ol style="list-style-type: none"> 1. Reference data (those data originating from external public databases) 2. Experimental data (published data submitted by DAIT-funded users) <p>Note: Ontology data in the data warehouse will be addressed in a separate use case.</p>	
Preconditions	<ol style="list-style-type: none"> 1. The user is a registered, valid ImmPort User. 2. The user is logged onto ImmPort system. 	
Postconditions	<p>User views the intended public data.</p> <p>User has ability to save viewable information to desktop (if desired).</p>	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. User selects to show public data sets (tables/views) in the data warehouse. 2. User selects the type of data (reference and/or experimental) that he/she wishes to view. 3. The system presents a view of the data set(s) along with all the data elements contained in the data set(s). 4. The user has the capability to browse data warehouse contents to select particular available data sets (subject area). 5. The system then displays the desired viewable subject area data set(s) and may offer capability to show specific data content at different levels to the user. (Example: particular years of data available, data elements, data element descriptions, etc.). 6. User has ability to save displayed viewable information to desktop if desired. 	
Alternate Course	TBD	
Exceptions	TBD	
Included Use Cases	TBD	
Extended Use Cases	TBD	
Priority	High	

Frequency of Use	TBD
Business Rules	TBD

Use Case ID	UP-2	
Use Case Name	Query public data	
Created By	Vince Desborough	Date: 9-Mar-2005
Updated By	Vince Desborough	Date: 28-Mar-2005
Actors	Any user	
Description	<p>Interaction with publicly accessible data. The ImmPort user has the capability to query the data warehouse. The data warehouse will contain both “online” and “near online” data storage. Therefore, access to these two types of data should happen automatically.</p> <p>For purposes of this use case, the user will not need to be concerned about whether data is “online” or “near online”. It will just take longer to satisfy their request if all or some of the data requested is “near online”.</p> <p>Querying data will include viewing data in the areas of:</p> <ol style="list-style-type: none"> 1. Reference data (from external public databases) 2. Experimental data <p>NOTE: Ontology data in the data warehouse will be addressed in a separate use case.</p>	
Preconditions	<ol style="list-style-type: none"> 1. User is a registered, valid ImmPort User. 2. User logged onto ImmPort system. 	
Postconditions	<p>User can cancel query or view of data at any time without consequences. The system should recognize the user’s command and terminate any ongoing sessions (either in batch or interactive) that might still be running.</p> <p>User is able to control/direct query output (results) when querying data. This includes output content (summary, detail, range of data), output format, and output destination. The output destination may be to the screen, the local PC, or a PPW (if he/she has write permission).</p> <p>User has ability to save a query selection (parameters), either to the PPW area (if he/she has permission) or to the local PC. This query would then be able to be recalled at a later date during a future ImmPort session and be able to be re-executed.</p>	
Assumptions	<p>All users will have access to the simple query capability of reference/experimental data (as exemplified in the Normal Course below).</p> <p>Only select users will have access to the complex query capability of reference/experimental data (as exemplified in the Alternate Course). The means of determining who has access to the complex query capability is yet to be determined.</p>	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. System displays data sets (tables/views) that user is allowed to access (view). 2. A high-level query tool(s) will be provided to the user. These 	

	<p>tool(s) will allow user to perform a system-guided query. The different query types, and type(s) of result information provided will be defined.</p> <ol style="list-style-type: none"> 3. User selects query options or parameters. This may include clicking on folder-selection, use of check boxes, radio buttons, pull-down lists, manual field data entry, etc. 4. User selects type of query output and the data format of the output. 5. User submits the query. 6. The system performs the query with intelligent query management/feedback to the user. Feedback may contain estimated time of completion and may provide query execution options. Some option examples: <ul style="list-style-type: none"> ▪ if query will take a long time, user may be permitted to submit query in batch mode, with some notification to use when it is complete. ▪ if query is estimated to take so long (3 minutes or so), then a second "submit" confirmation from user may be requested for just before interactive execution of query ▪ if query result data set is large, then user may be notified, and allowed to save results on server in PPW, rather than to the screen. 7. System displays results or a result data set, or saves it to place user has requested,, in the format they requested. 8. User has ability to store/save information that was queried.
<p>Alternate Course</p>	<p>Complex Query in Data Warehouse</p> <ol style="list-style-type: none"> 1. User is provided list of database extraction tool(s) to use for data query/analysis. 2. The user has the ability to select data set(s), join data set(s), and select particular data elements (columns) from the data set. The user has the ability to construct a data query built using structured query language (SQL) principles. These will most likely follow standard Oracle SQL. This GUI tool will support display, joining, and selection of data sets (Oracle table/views), and will support ImmPort defined "Power" user functions, allowing them to perform high-level complex queries if desired. 3. The user will have ability to review/edit, save, these queries (either simple or complex) in their current user session. 4. A GUI SQL interface will be provided to the user. 5. An actual SQL code window may or may not be available to the user. If available, this window would display the active SQL code generated by the tool, which is the code that would be passed to the data warehouse. 6. User selects and builds query and also selects the way in which to invoke this query. 7. User has control over the content, type, and format of output, and how this output is to be displayed and/or stored. 8. User submits query. 9. System performs query with intelligent query

	<p>management/feedback to user. Feedback may contain estimated time of completion, and may allow user query execution options. Some option examples:</p> <ul style="list-style-type: none"> ▪ if query will take a long time, user may be permitted to submit query in batch mode, with some notification to use when it is complete. ▪ if query is estimated to take so long (3 minutes or so), then a second "submit" confirmation from user may be requested for just before interactive execution of query ▪ if query result data set is large, then user may be notified, and allowed to save results on server in project workspace, rather than to the screen. <p>10. The system displays results or a result data set or saves it to the location the user has requested in the format he/she has specified.</p> <p>11. User has ability to store/save the result information that was queried. Result information (data set) may also be able to be used in subsequent complex queries.</p> <p>12. User has the ability to save the query for future reference in a future session, possibly available to other users if saved in the PPW the user is assigned to.</p>
Exceptions	TBD
Included Use Cases	TBD
Extended Use Cases	TBD
Priority	2
Frequency of Use	TBD
Business Rules	TBD

Use Case ID	UP-3	
Use Case Name	Download public data	
Created By	Vince Desborough	Date: 9-Mar-2005
Updated By	Vince Desborough	Date: 28-Mar-2005
Actors	Any user	
Description	<p>Download public data. The ImmPort user is to be able to download specific public data sets (available in ImmPort data warehouse) to their PC or to their PPW (if they have write permission).</p> <p>This download capability will be provided for specific ImmPort Reference and Experimental data sets.</p>	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. User logged onto ImmPort system 	
Postconditions	User is able to control/direct download destination, either to their local PC, or to PPW (if they have write permission).	
Assumptions	Only DAIT approved public data sets will be provided for download.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. System displays downloadable data sets available. 2. User selects data set to download. 3. System prompts user to select output destination (either PC or PPW (if he/she has write permission)). 4. User has option to download data set. 5. Data set is saved. 	
Alternate Course	<p>Cancel</p> <ol style="list-style-type: none"> 1. The user selects to cancel the operation at anytime (before step 5). 	
Exceptions	TBD	
Included Use Cases	TBD	
Extended Use Cases	TBD	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	UP-4	
Use Case Name	View Ontology data	
Created By	Vince Desborough	Date: 9-Mar-2005
Updated By	Vince Desborough	Date: 28-Mar-2005
Actors	Any user	
Description	View ontology data. The ImmPort can view ontology data in the ImmPort data warehouse, without the need to perform any queries. The viewable ontology tree will be easy and intuitive to the user.	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. User logged onto ImmPort system 	
Postconditions	<p>User should be able to cancel view of data at any time, without consequences.</p> <p>User has ability to save viewable ontology information to their local PC.</p>	
Assumptions	ImmPort system will consist of ontology data, Reference data, and Experimental data.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. System displays ontology in a hierarchical tree (can expand down, by clicking on a term/nomenclature). 2. User has ability to scroll down, up, and navigate within ontology. 3. User has ability to store/save selected ontology information shown if desired. 	
Alternate Course	TBD	
Exceptions	TBD	
Included Use Cases	TBD	
Extended Use Cases	TBD	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	UP-5	
Use Case Name	Query Ontology data	
Created By	Vince Desborough	Date: 09/03/05
Updated By	Mark Dela Cruz	Date: 5/17/05
Actors	Any user	
Description	User can query ontology data in the data warehouse. This query will be intuitive and system-guided.	
Preconditions	<ol style="list-style-type: none"> 1. The user is registered and authorized in ImmPort. 2. User logged onto ImmPort system 	
Postconditions	<p>User can cancel query or view of data at any time without consequences. The system should recognize the user's command and terminate any ongoing sessions (either in batch or interactive) that might still be running.</p> <p>User is able to control/direct query output (results) when querying data. This includes output content (summary, detail, range of data), output format, and output destination. The output destination may be to the screen, the local PC, or a PPW (if he/she has write permission).</p> <p>User has ability to save a query selection (parameters), either to the PPW area (if he/she has permission) or to the local PC. This query would then be able to be recalled at a later date during a future ImmPort session and be able to be re-executed.</p>	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. User selects to perform an ontology query. 2. System displays ontology tree (can expand down, by clicking on a term/nomenclature) 3. User types in a term to search for, using an intelligent text search window. They enter the term they want to look up. 4. System executes the search, and automatically takes the user to first occurrence of term, and identifies all other occurrences with ability to go to next occurrence. 5. User has ability to store/save information that was queried. 	
Alternate Course	TBD	
Exceptions	TBD	
Included Use Cases	UP-4 View Ontology data	
Extended Use Cases	TBD	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	UP-6	
Use Case Name	Data Sharing – Query Ontology data, then use that result to Query/lookup publicly accessible data in data warehouse (provide an "integrated ontology" capability)	
Created By	Vince Desborough	Date: 09/03/05
Updated By	Mark Dela Cruz	Date: 05/17/05
Actors	Any user	
Description	The ImmPort user will be able to query ontology data, and then use that ontology result to perform an integrated query/lookup of publicly accessible data in the data warehouse.	
Preconditions	User is a registered, valid ImmPort User. User logged onto ImmPort system.	
Postconditions	<p>User can cancel query or view of data at any time without consequences. The system should recognize the user's command and terminate any ongoing sessions (either in batch or interactive) that might still be running.</p> <p>User is able to control/direct query output (results) when querying data. This includes output content (summary, detail, range of data), output format, and output destination. The output destination may be to the screen, the local PC, or a PPW (if he/she has write permission).</p> <p>User has ability to save a query selection (parameters), either to the PPW area (if he/she has permission) or to the local PC. This query would then be able to be recalled at a later date during a future ImmPort session and be able to be re-executed.</p>	
Assumptions	The ontology data query and subsequent Reference and Experimental queries performed will be intuitive and system-guided.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. User selects to perform an ontology query. 2. System displays ontology tree (can expand down, by clicking on a term/nomenclature) 3. User types in intelligent text search window, the term he/she is looking up. 4. System automatically takes user to first occurrence of term and identifies all other occurrences with ability to get to next occurrence. 5. Ontology tool supports showing all aspects of ontology information for that concept, class, subclass, etc. 6. User has ability to use the ontology query result to search the data warehouse. 7. User selects type of data warehouse query output, and the data format of the output. 8. User submits the query. 9. The system performs the query on reference data with intelligent query management/feedback to user. Feedback may contain 	

	<p>estimated time of completion, and may allow user query execution options. Some option examples:</p> <ul style="list-style-type: none"> ▪ if query will take a long time, user may be permitted to submit query in batch mode, with some notification to use when it is complete. ▪ if query is estimated to take so long (3 min. or so), then a 2nd "submit" confirmation from user may be requested for just before interactive execution of query ▪ if query result data set is large, then user may be notified, and allowed to save results on server in project workspace, rather than to the screen. <p>10. The system displays results or a result data set or saves it to place user has requested in the format the user has specified. 11. User has ability to store/save information that was queried.</p>
Alternate Course	TBD
Exceptions	TBD
Included Use Cases	UP-4 View Ontology data UP-5 Query Ontology data
Extended Use Cases	TBD
Priority	2
Frequency of Use	TBD
Business Rules	TBD

Use Case ID	UP-7	
Use Case Name	Retrieve gene information by gene ID, gene symbol or gene name	
Created By	Jennifer Cai	Date: 4/25/2005
Updated By	Richard Scheuermann	Date: 4/26/2005
Actors	Any user	
Description	The use case of <i>retrieve gene information by gene ID, gene symbol or gene name</i> describes the flow of events for the actors to retrieve gene information by gene symbol	
Preconditions	1. The actor is registered and authorized in ImmPort System.	
Postconditions	The results from data retrieval are displayed either in html format and viewed through the browser or are supplied in a report format (.rtf or .txt) based on actor's needs.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The actor is logged on 2. The actor performs one of the following actions: <ol style="list-style-type: none"> a. The actor inputs a single gene ID/gene symbol/gene name and selects organism from pull down menu to retrieve the gene records b. The actor inputs multiple gene ID/gene symbols/gene names and selects organism from pull down menu to retrieve gene records 3. Based on the input, the ImmPort generates report of gene including the following information: <ul style="list-style-type: none"> • Gene ID (with hyperlink to NCBI) • Gene symbol • Gene synonyms • Gene name • Chromosome position • Hyperlink to genome browser view of gene • Unigene ID (with hyperlink to NCBI) • HomoloGene group ID (with hyperlink to homologous gene record from other species) • mRNA Refseq ID (with hyperlink to sequence record in FASTA format) • protein Refseq ID (with hyperlink to sequence record in FASTA format) • Organism taxonomic name • Hyperlink to polymorphism page (SNPs and microsatellites) • Hyperlink to gene expression page • Hyperlink to alternative transcripts page • Hyperlink to protein structure page (motifs, domains, 3D structure) 4. If the input is a gene symbol, ImmPort system first searches the official gene symbols in the database. If no record found with input gene symbol, the gene's synonyms are to be used to retrieve the gene 	

	information.
Alternate Course	Cancel action (after step 2) 4. User selects to cancel the operation. 5. No operation is to conducted
Exceptions	
Included Use Cases	None
Extended Use Cases	
Priority	1
Frequency of Use	TBD
Business Rules	
Comments	

Use Case ID	UP-8	
Use Case Name	Integration of results from SNP-based gene association study with ImmPort gene annotation and GEO database	
Created By	Jennifer Cai	Date: 4/25/2005
Updated By	Richard Scheuermann	Date: 4/27/2005
Actors	Any user	
Description	This use case describes the flow of events for the actors to integrate results from SNP-based gene association studies with gene annotation data derived from the ImmPort Gene List database and with gene expression data derived from the GEO public database while specifying a disease process or cell type.	
Preconditions	<ol style="list-style-type: none"> 1. The actor is registered and authorized in the ImmPort System. 2. The actor has selected a SNP for further analysis based on the results of a gene association study. 	
Postconditions	The results from data retrieval are displayed either in html format and viewed through the browser or are supplied in a report format (.rtf or .txt) based on actor's needs.	
Assumptions	In this use case, the actor is only interested in obtaining gene expression data from a subset of records with some relationship with a specific disease process or cell type.	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The actor is logged on 2. The actor inputs a dbSNP ID identified by a gene association study into the ImmPort system. 3. The ImmPort system generates a report with following information for the gene(s) located closest to the SNP: <ul style="list-style-type: none"> • Gene ID (with hyperlink to NCBI) • Gene symbol • Gene synonyms • Gene name • Chromosome position • Hyperlink to genome browser view of gene • Unigene ID (with hyperlink to NCBI) • HomoloGene group ID (with hyperlink to homologous gene record from other species) • mRNA Refseq ID (with hyperlink to sequence record in FASTA format) • protein Refseq ID (with hyperlink to sequence record in FASTA format) • Organism taxonomic name • Hyperlink to gene expression page 4. If the actor selects the hyperlink to gene expression page, the actor will be transferred to a gene expression page specific for the Gene ID, Gene symbol and Gene name, which are listed in the header of the page. 5. This page would contain two pull down menus for selection: <ul style="list-style-type: none"> • Disease process • Cell type 6. These pull down menu would be populated with terms derived from the ImmPort Ontology 7. The actor would select one or more disease processes and/or one or more cell types from the pull down lists, and submit. 8. The ImmPort system would generate a report with gene name, gene ID and gene symbol and a hyperlink to the GEO database (For example, for each IL12A, 	

	<p>http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=geo&term=IL12A)</p> <p>9. The actor clicks on the GEO URL and retrieves all the gene expression profiles of that gene</p>
Alternate Course	<p>Cancel action (after step 2)</p> <p>6. User selects to cancel the operation.</p> <p>7. No operation is to be conducted</p>
Exceptions	
Included Use Cases	None
Extended Use Cases	
Priority	2
Frequency of Use	TBD
Business Rules	
Comments	<p>Please refer to the following URL at the GEO web site:</p> <p>http://www.ncbi.nlm.nih.gov/geo/</p>

Use Case ID	UP-9	
Use Case Name	Integrate ImmPort to hypothesize genes based on Gene Ontology, network modules and gene expression profile	
Created By	Jennifer Cai	Date: 4/25/2005
Updated By	Richard Scheuermann	Date: 4/27/2005
Actors	Any user	
Description	The use case of <i>Integrate ImmPort system to hypothesize genes based on Gene Ontology, network modules and gene expression profile</i> describes the flow of events for the actors to use the public data deposited in the ImmPort system	
Preconditions	<ol style="list-style-type: none"> 1. The actor is registered and authorized in ImmPort System. 2. The actor has generated experimental data from genotyping. 3. The actor has performed a population genetic study in which they have identified SNPs that correlate with a specific clinical parameter (e.g. serum antibody response to vaccination) 	
Postconditions	The results from data retrieval are displayed either in html format and viewed through the browser or are supplied in a report format (.rtf or .txt) based on actor's needs.	
Assumptions	In this use case, the actor is only interested in integrating the ImmPort genes to hypothesize gene related to certain of disease	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The actor is logged on 2. The actor inputs a dbSNP ID identified by a gene association study into the ImmPort system. 3. The ImmPort system generates a report with following information for the gene(s) located closest to the SNP: <ul style="list-style-type: none"> • Gene ID (with hyperlink to NCBI) • Gene symbol • Gene synonyms • Gene name • Chromosome position • Hyperlink to genome browser view of gene • Unigene ID (with hyperlink to NCBI) • HomoloGene group ID (with hyperlink to homologous gene record from other species) • mRNA Refseq ID (with hyperlink to sequence record in FASTA format) • protein Refseq ID (with hyperlink to sequence record in FASTA format) • Organism taxonomic name • Gene Ontology annotation terms and IDs • Metabolic pathway annotation terms and IDs (derived from the human BioCyc database) • Signal transduction network annotation terms and IDs (to be developed) • Protein-protein interaction network annotation and IDs (to be developed) 4. The ImmPort system will also ask the actor if they want to expand the gene list base on: 	

	<ul style="list-style-type: none"> • Gene Ontology annotation • Metabolic pathway annotation • Signal transduction network annotation • Protein-protein interaction network annotation <p>5. Based on the report generated from steps 3 & 4, the actor can select one of the options in step 4; the ImmPort system would then supply a list of GeneIDs, Gene symbols and Gene names annotated with the selected IDs from the ImmPort Gene List database</p> <p>6. The actor may then choose to extract marker SNPs from the ImmPort Gene List database for genes selected from this list</p>
Alternate Course	<p>Cancel action (after step 2)</p> <p>8. User selects to cancel the operation.</p> <p>9. No operation is to conducted</p>
Exceptions	
Included Use Cases	None
Extended Use Cases	
Priority	2
Frequency of Use	TBD
Business Rules	

Use Case ID	PD-1	
Use Case Name	Authorize research project data set access	
Created By	Jihong Chen	Date: 3/18/05
Updated By	Mark Dela Cruz	Date: 05/17/05
Actors	PI	
Description	This use case describes the workflow of adding a user to the access list of a research project.	
Preconditions	A user does not have access to research project data.	
Postconditions	A user is granted access to research project data.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1 The PI selects to change data set accessing user list. 2 The system checks the authority. 3 The system loads the user list that can access the research project. 4 The PI adds the applicant to the accessing list of the research project. 5 The system saves the changes. 6 The system provides a confirmation page of the change. 	
Alternate Course	TBD	
Exceptions	TBD	
Included Use Cases	TBD	
Extended Use Cases	TBD	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	PD-2	
Use Case Name	Revoke the research project data set access	
Created By	Jihong Chen	Date: 3/18/05
Updated By	Jihong Chen	Date: 4/1/05
Actors	PI	
Description	This use case describes the workflow of revoking the data set access from some external user.	
Preconditions	A user has access to the research project data.	
Postconditions	A user no longer has access to the research project data.	
Assumptions		
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. The PI selects to change data set accessing user list. 2. The system checks the authority. 3. The system loads the users who can access the research project data. 4. The PI removes the selected user from the accessing list. 5. The system saves the changes. 6. The system provides a confirmation page of the change. 	
Alternate Course	TBD	
Exceptions	TBD	
Included Use Cases	TBD	
Extended Use Cases	TBD	
Priority	1	
Frequency of Use	TBD	
Business Rules	TBD	

Use Case ID	PD-3	
Use Case Name	Publishing of data set(s) from PPW to Public Data Warehouse	
Created By	Jawwad Cheema	Date: 3/15/05
Updated By	Mark Dela Cruz	Date: 5/17/05
Actors	PI	
Description	DS – 11 describes the flow of events for publishing the data set(s) from a PPW to public data warehouse.	
Preconditions	<ol style="list-style-type: none"> 1. User has successfully completed the Security/Access Module. 2. User is authorized to publish data. 3. User is logged into the ImmPort system. 4. Data set(s) to be published have already been created. 5. Data set(s) format has been validated. 	
Postconditions	Data set(s) is/are published to the data warehouse.	
Assumptions	<ol style="list-style-type: none"> 1. Data set(s) have already been created. 2. Data set(s) format has been validated. 	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. ImmPort system provides user with the options: <ul style="list-style-type: none"> • Publish new data set(s) to the data warehouse • View published data set(s) tag information 2. User selects to publish new data set(s) to the data warehouse. 3. ImmPort display all data set(s) ready to be published. 4. User selects from the list of data set(s) to be published. 5. ImmPort system validates data set(s) format and provides user with: <ul style="list-style-type: none"> • Format validated – Continue 6. User selects to continue. 7. ImmPort system display data set(s) tag information (Data set ID, Data set Name, user name, create date, project name, today’s date (publish date – read only)) and provides user with the options: <ul style="list-style-type: none"> • Accept displayed information. • Cancel data set(s) Publishing. 8. User accepts data set(s) tag information. 9. ImmPort system publishes data set(s) to the data warehouse. 10. ImmPort system notifies user with “publishing complete” message. 11. ImmPort system returns User to step 1. 	
Alternate Course	<p>View published Data Set(s) (at step 1)</p> <ol style="list-style-type: none"> 1. User selects to view published data set(s). 2. ImmPort system display all data set(s) published by the user for that research project. 3. User makes selection from the list. 4. ImmPort system display data set(s) information. <ul style="list-style-type: none"> • Data set ID • Data set Name • Data set project name • Data set published date • Data set creation date • Data set last modify date • Data set owner name (user who created the data set) 	

	<p>User selects to cancel data set(s) publishing (at step 7)</p> <ol style="list-style-type: none"> 1. User selects to cancel data set(s) publishing 2. ImmPort system prompt user to acknowledge the selection (are you sure you want to cancel? Yes/no) <ol style="list-style-type: none"> a. User makes 'yes' selection, return user to step 1 b. User selects no. User is returns to step 8 of Normal course
Exceptions	<p>ImmPort system display invalid format (step 5)</p> <ol style="list-style-type: none"> 1. Error Message – Invalid format. 2. ImmPort system notify the owner of data set(s) <ul style="list-style-type: none"> • Format not valid, please reformat data set (Data set ID, Data set Name, create date) 3. User is returned to Step 1. <p>User not authorized to publish data (at step 1)</p> <ol style="list-style-type: none"> 1. Display Error Message 2. Return to Step 1
Included Use Cases	TBD
Extended Use Cases	TBD
Priority	2
Frequency of Use	TBD
Business Rules	TBD

Use Case ID	UD-1	
Use Case Name	View data in Private Project Workspace (PPW)	
Created By	Jawwad Cheema	Date: 3/15/05
Updated By	Mark Dela Cruz	Date: 05/17/05
Actors	Users of a PPW	
Description	The UD-1 View data in PPW describes the flow of events for user to view data set(s) from PPW.	
Preconditions	<ol style="list-style-type: none"> 1. User has successfully completed the Security/Access Module. 2. User is logged into ImmPort. 	
Postconditions	ImmPort system displays data set(s) from PPW.	
Assumptions	<ol style="list-style-type: none"> 1. PPW format has been decided. 2. User has successfully completed Security/Access Module. 3. User is authorized to view data in PPW 	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. User selects to view data from PPW. 2. ImmPort system displays an estimated time to display data. 3. ImmPort system returns PPW data results for user to view. 	
Alternate Course		
Exceptions	ImmPort system returns no results (step 3). <ol style="list-style-type: none"> 1. ImmPort system displays message (no results found). 2. User is returned to step 1 (Normal Course). 	
Included Use Cases	TBD	
Extended Use Cases	TBD	
Priority	1	
Frequency of Use	High	
Business Rules	List of Data Sharing business rules.	

Use Case ID	UD-2	
Use Case Name	Query data in Private Project Workspace (PPW)	
Created By	Jawwad Cheema	Date: 3/15/05
Updated By	Mark Dela Cruz	Date: 05/17/05
Actors	Users of a PPW	
Description	The UD-2 Query data in a PPW describes the flow of events for user to query data set(s) from the PPW.	
Preconditions	User has successfully completed the Security/Access Module. User is logged into ImmPort.	
Postconditions	ImmPort system displays data set(s) from the PPW.	
Assumptions	<ol style="list-style-type: none"> 1. PPW format has been decided. 2. User has successfully completed Security/Access Module. 3. User is authorized to query data from PPW. 	
Flow of Events		
Normal Course	<ol style="list-style-type: none"> 1. User selects to query data from PPW. 2. ImmPort system display query options screen. 3. User selects query criteria and submits the request. 4. ImmPort system displays an estimated query time. 5. ImmPort system returns query results for user. 6. ImmPort provides user with options: <ul style="list-style-type: none"> View query result details Run new query Save query criteria Download Analyze. 	
Alternate Course	<p>ImmPort system returns no results (Step 5).</p> <ol style="list-style-type: none"> 1. ImmPort system displays message (no results found). 2. User is returned to step 2 (Normal Course) <p>Invalid selection criteria</p> <ol style="list-style-type: none"> 1. ImmPort system displays error message. 2. User is provided with option to correct error and rerun query. 3. User is returned to query options screen. 	
Exceptions	TBD	
Included Use Cases	TBD	
Extended Use Cases	TBD	
Priority	1	
Frequency of Use	TBD	
Business Rules	List of Data Sharing business rules.	

APPENDIX D BIOCHEMICAL ASSAYS

This table provides the same information in Section 7.5 Biochemical Assay Data but categorized according to the scientific purpose of conducting the experimental technique.

Table D1. Biochemical Assays To Be Supported by ImmPort

Experiment Type	Experimental Technique
Measure immune response	ELISPOT
Measure immune response	ELISA
SNP analysis and genotyping	Sequencing
SNP analysis and genotyping	HLA typing and other genotyping technologies
SNP analysis and genotyping	Illumina Bead Array
SNP analysis and genotyping	Microsatellite markers
SNP analysis and genotyping	GeneChip-based genotyping (Affymetrix SNP chip)
Measure mRNA expression	Microarray (Affymetrix cDNA array, Genpix array)
Measure protein expression	Flow Cytometry (FACS)
Measure protein expression	2D-GEL
Measure protein expression	ESI-LC
Measure protein expression	Mass Spectrometry
Measure mRNA expression	PCR
Measure protein function	Luminex Bead Assay
Measure protein function	Cytometric Bead Immunoassay
Quantify the function of protein and gene	Western Blot
Measure the DNA binding activity	EMSA
Measure the DNA binding activity	CHIP
Functional validation using siRNA	siRNA
Knockout/Knockin mice	Transgenic or Null Mutant Mice