



## **Assessment Information**

[CoreTrustSeal Requirements 2017–2019](#)

Repository: ImmPort Repository  
Website: <http://www.immport.org>  
Certification Date: 18 December 2017

This repository is owned by: **DAIT, NIAID, NIH**

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)



# Core Trustworthy Data Repository Requirements

---

## BACKGROUND INFORMATION

### Context

R0. Please provide context for your repository.

**Repository Type. Select all relevant types from:**

Domain or subject-based repository

**Other (please describe)**

### Comments

Research supported and conducted by the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), strives to understand, treat, and ultimately prevent immunologic, infectious, and allergic diseases. The NIAID Division of Allergy, Immunology, and Transplantation (DAIT) supports extramural basic, pre-clinical and clinical research that focuses on immune system development and function and the origin, prevention and treatment of immune-mediated diseases through a variety of research grants and contracts. The missions of DAIT include support for clinical research programs and individual research projects to evaluate the safety and efficacy of therapeutic and preventive approaches and agents and to elucidate the underlying mechanisms of such approaches and agents. Increasingly, bioinformatics and data science support for scientific data management and analysis is becoming an essential component of immunological research. Examples of this support include development and maintenance of stable and scalable computing systems and infrastructure to support long-term archival and dissemination of data and tools; integration and curation of disparate data sets and data types in specific domains and across clinical, laboratory and computational areas; data sharing follow the FAIR Data Principles (Findability, Accessibility, Interoperability and Reusability, datafairport.org); facilitating the use of novel analytical tools that allow the immunology community broad access to and facilitate analyses of existing data; providing online workspace for user collaborations; development of and adherence to standards and best practices for data collection, presentation and exchange; and training in the deposition and the use of data and tools, cutting-edge and cost-efficient computational infrastructure such as cloud computing platforms. ImmPort has provided a variety of bioinformatics resources in support of



immunology data sharing, integration, and reuse of DAIT-supported basic research, clinical studies and clinical trials. The data of these studies are curated by the ImmPort team through collaborations with research programs to ensure that data in ImmPort are adequately documented, of high quality and analyzable. ImmPort, as a primary immunology data sharing repository, will continue to provide and improve means for easy access, storage, integration, visualization, analyses and sharing of complex, high-quality immunology- related data sets through the use of user-friendly web interfaces and tools. Also, ImmPort continues to serve as a platform for collaborative scientific discoveries, integrative research, and data sharing. More details can be found in "Bhattacharya, S., Andorf, S., Gomes, L., et al., (2014). ImmPort: disseminating data to the public for the future of immunology. *Immunologic Research*, May;58(2-3), p. 234-9. doi: 10.1007/s12026-014-8516-1. PMID: 24791905".

**Brief Description of the Repository's Designated Community.**

The Immunology Database and Analysis Portal (ImmPort) provides advanced information technology and curation support in the archiving and exchange of scientific data for the diverse community of life science researchers supported by NIAID/DAIT and serves as a long-term, sustainable archive of immunology research and clinical data.

**Level of Curation Performed. Select all relevant types from:**

D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

**Comments**

The curation team at Northrup Grumman ensure that the data is converted to the ImmPort data model incorporating relevant metadata to allow for data deposition and linking to partner databases (<http://www.immport.org/immport-open/public/schema/schemaTree>). Also, to comply with all required laws and regulations, and domain-specific scientific guidelines including human subject de-identification process, sensitive genotypic information. The documentation of the processes and guideline is available publically at <http://www.immport.org/immport-open/public/home/documentation>. ImmPort adopts, adapt and create community standards and practices for data types relevant to immunology research and clinical studies. A list of information and data type collected in the form of templates is at <http://www.immport.org/immport-open/public/home/dataTemplates>.

**Outsource Partners. If applicable, please list them.**

ImmPort uses Federal contract mechanism to support its operation. Northrup Grumman (NG) is the primary contractor. NG subcontracts UCSF, ESAC, Inc, Yale University, and others as needed to acquire complementary leadership, scientific and technical expertise. Dr. Atul Butte of UCSF serves the Principal Investigator of the contract.



**Other Relevant Information.**

The ImmPort data storage is currently supported by a combination of federal facilities for protected and unpublished data and AWS for non-sensitive and non-sensitive data like metadata to facilitate search and retrieval of publicly shared data is distributed via the web using Amazon AWS.

**Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## ORGANIZATIONAL INFRASTRUCTURE

### I. Mission/Scope

Compliance Level: 4

**R1. The repository has an explicit mission to provide access to and preserve data in its domain.**

The goals of ImmPort are stipulated by DAIT, the funder of the contract. They include the following:

- Provide an open-access platform for research data sharing.
- Create an integrated environment that broadens the usefulness of scientific data and advances hypothesis-driven and hypothesis-generating research.
- Accelerate scientific discovery while extending the value of scientific data in all areas of immunological research.
- Promote rapid availability of important findings, making discoveries available to the research community for further analysis and interpretation • Provide analysis tools to advance research in basic and clinical immunology.

#### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## II. Licenses

Compliance Level: 4

**R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.**

ImmPort provides a limited license as stipulated in its User Agreement (<http://www.immport.org/agreement>). ImmPort de-identifies human subject data to comply with applicable laws and regulations. Users are permitted to use the data in any lawful way. The agreement is governed by the federal courts in the District of Columbia should users not comply with the User Agreement. Any updates regarding the User Agreement are relayed to users via email they used when registering to access the repository. The content is always available at the ImmPort Website.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



### III. Continuity of access

Compliance Level: 3

#### R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

ImmPort has been funded by contract mechanism by DAIT for three five year cycles since 2002 and was renewed for the fourth cycle in 2017. In short to medium-term (up to a 5-year cycle of a contract), the funding commitment will ensure the continued availability and accessibility of the data. Nonetheless, the nature of budgetary decisions and obligations by the Federal Government cannot be ascertained beyond the specified contract period; therefore, ensuring data availability beyond that commitment is not guaranteed. DAIT ascertains the performance of the contract internally two to three years before the contract period ends to determine whether to support an additional contract period. The current contract contains a continuity plan to cover transition in the event another contract cycle is undertaken, and the current contractor is not selected for the additional cycle of performance. Such a clause documents a process to ensure continuity of access and transference of all data to the new contractor and no lapse of access or loss of data.

#### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## IV. Confidentiality/Ethics

Compliance Level: 4

**R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.**

The ImmPort database complies with disciplinary and ethical norms by ensuring de-identification of the data to protect privacy as stipulated by laws such as HIPAA and the Common Rule. Funded research that generates the data must also meet such obligations including IRB review and compliance with human subjects regulations and laws such as informed consent. The User Agreement also specifies the legal ways in which the data may be used (<http://www.import.org/agreement>) which concur with disciplinary and ethical norms.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## V. Organizational infrastructure

Compliance Level: 4

**R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.**

The repository is currently funded through a contract from NIAID DAIT to Northrup Grumman (NG). NG subcontracts UCSF, ESAC, Inc, Yale University, and others as needed to acquire complementary leadership, scientific and technical expertise. NIAID DAIT oversees a portfolio of research on fundamental understanding of the immune system as part of NIAID. NIAID is one of the 27 Institutes and Centers of the NIH, the largest funder of biomedical research in the world, and is widely recognized as a leader in the area of immunology research. NG has extensive experience as a contractor organization working on Federal government IT initiatives. The NG staff includes programmers, bioinformaticians, database administrators, and project managers. Also, the ImmPort's scientific direction includes the input of Dr. Atul Butte, the Principal Investigator of the contract, who is a recognized leader in the area of biomedical informatics and biotechnology and who provides scientific input on how to reuse the data. The members of scientific teams attend scientific meetings and symposia to maintain up to date understanding and training in biomedical informatics and immunology. Please see link for more information about the current ImmPort team:

<http://www.immport.org/immport-open/public/home/teamCurrent>.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## VI. Expert guidance

Compliance Level: 4

**R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).**

The repository has a yearly steering committee meeting which involves input from the program staff at the NIH who are experts in their fields but also external, recognized experts in clinical science, immunology, and biomedical informatics. At this meeting, the ImmPort team receives feedback and suggestions on the technical and scientific progress made in the previous year. Beyond this steering committee meeting, ImmPort staff attend ad-hoc or regularly scheduled scientific conferences and symposia to solicit expert feedback on data deposition, curation, and reuse. One example is the first User Group meeting that was held on 5/7/2017 to solicit input from the research community. The ImmPort user community consists of all registered ImmPort users who agree to abide by the User Agreement. There are no defined roles assigned to users and no responsibilities associated with using the service except for abiding by the User Agreement. The User Group meeting is organized on an as-needed basis. The ImmPort team also attends weekly and monthly progress meetings with the program staff at DAIT to receive guidance and feedback. Finally, there is a UH2 funding mechanism (<https://grants.nih.gov/grants/guide/pa-files/PAR-16-253.html>) to permit researchers to apply for funding towards research on how to use the ImmPort database to develop new tools or facilitate secondary research. These researchers will regularly engage with the ImmPort team on how to improve the resource.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## DIGITAL OBJECT MANAGEMENT

### VII. Data integrity and authenticity

Compliance Level: 4

#### R7. The repository guarantees the integrity and authenticity of the data.

The ImmPort database maintains a regular release schedule where new studies, data, or information is shared. Each data release contains release notes on what is being shared or any data changes being made (<http://www.immport.org/immportopen/public/home/dataReleaseNotes>). It is also the case for any changes to the software powering the ImmPort resource (<http://www.immport.org/immport-open/public/home/softwareReleaseNotes>). The database schema is publicly available (<http://www.immport.org/immport-open/public/schema/schemaDefinition/study>). The database checks the identity of data providers and maintains a list of the studies and datasets that were shared by each provider (<http://www.immport.org/immportopen/public/home/dataProviders>). Data provided by data depositors are given a ticket and enters an audit trail that monitors changes to the uploaded data. The database uses community standards including the use of metadata identifiers to linked databases such as GEO where much of the gene expression data resides for example. Previous data releases are stored by the data release version number, and the data can be accessed to compare with any updates to the most recent data release. ([http://www.immport.org/immportopen/resources/docs/BISC\\_System\\_Architecture\\_and\\_Software\\_Design\\_Specification\\_v5.0.pdf](http://www.immport.org/immportopen/resources/docs/BISC_System_Architecture_and_Software_Design_Specification_v5.0.pdf)).

#### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## VIII. Appraisal

Compliance Level: 4

**R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.**

The repository has developed data standards and procedures by working with the large research consortia and programs that generate the bulk of the data (e.g., HIPC <https://www.immuneprofiling.org/hipc/page/show>) as well as continuous outreach and feedback (e.g., the User Group meeting). The ImmPort teams listen to the community's needs to continuously develop data standards for storage long term. The repository has a series of data submission templates that facilitate data submission to ensure that any deposited data have the minimal level of metadata for interpretability and long-term storage. There are QC checks in place (e.g., the data validator <http://www.immport.org/documentation/data/upload/validator>) at the time of submission to ensure the minimal level of metadata is available and successfully curated from the data in the preferred format. If data is submitted in a non-preferred format, the data submission team at ImmPort will work with the submitters to wrangle the data into a preferred format (<http://www.immport.org/immportopen/public/home/dataSubmission>).

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## IX. Documented storage procedures

Compliance Level: 4

### R9. The repository applies documented processes and procedures in managing archival storage of the data.

Relevant processes and procedures are documented online and are managed by the ImmPort team in consultation with program staff at NIAID who oversee the management of the contract funding this resource. The database host two types of data (private and public). Private data, which is in preparation for public release, is hosted in a secure facility at NIAID and is only available to the data depositors and their collaborators. Since private data is hosted on government servers, it must maintain strict security and fault tolerance capabilities as specified by FISMA law

([https://en.wikipedia.org/wiki/Federal\\_Information\\_Security\\_Management\\_Act\\_of\\_2002](https://en.wikipedia.org/wiki/Federal_Information_Security_Management_Act_of_2002)). Public data is hosted in the cloud on Amazon AWS, which also provides industry gold standard for security and fault tolerance capabilities (<http://www.immport.org/immport-open/public/home/dataSubmission>).

#### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## X. Preservation plan

Compliance Level: 4

**R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.**

Since the repository has been funded by a government contract, the repository assumes responsibility for long-term preservation as long as the funding continues. DAIT internally reviews current contract performance before the end of the existing contract cycle to decide whether another contract period will be supported. In the event another contract period will be supported, the current contract includes a clause to handle continuity of access to repository data whether the current contractor is selected for the next contract period or not. In such a hypothetical situation, the data will be preserved and transferred to the next contractor, either way, as per the contract documentation. More specifically, the contractor shall coordinate with the incumbent contractor and the Government to implement an orderly, secure and efficient transition of contract activities and contract-generated data, systems, analytical tools and other documents and materials, and to ensure a seamless continuation of contract services and operation of ImmPort during the transition period. The technical requirement details of ImmPort system and architecture, software design, data model are published publically at <http://www.immport.org/immport-open/public/home/documentation>.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## XI. Data quality

Compliance Level: 4

**R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.**

The repository handles metadata and data quality through conversations with the communities generating the data to determine the level of quality necessary for data storage. The repository documents how metadata and data quality are ensured via the data deposition procedure. It also has automated methods to determine whether the data quality adheres to the schema strategy (e.g., the data validator). At the moment, the repository does not offer functionality to rate the metadata or data quality although this has been internally discussed and may be a possibility if there is user interest. ImmPort contains metadata on any citations that generated the data as well as links to those publications (<http://www.immport.org/immport-open/public/home/publications>). This is relevant for citation purposes to link the data via associated ImmPort study ID to publication.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## XII. Workflows

Compliance Level: 4

### R12. Archiving takes place according to defined workflows from ingest to dissemination.

ImmPort has documented processes for each stage of the data archiving process from ingest to dissemination. First, the determination of what data to collect in the repository is governed by the database mission. The number one priority is any immunology data from the large clinical or research networks funded by NIAID DAIT. Immunology data from other partner organizations are the second priority and involves a discussion and determination by the program staff at DAIT. Ingest involves outreach and discussion with the depositors by the ImmPort team to point them in the direction of tools such as the submission templates and data validator. After data ingestion, data is given a ticket number and audited through the parsing and curation process to ensure quality. The data itself is first stored in a private workspace in an NIAID secure hosting facility since many datasets are generated by researchers who have specific data sharing needs before public data sharing (e.g., primary publication, QC/QA, curation, etc.). Publicly shared data is also stored in this Federal secure hosting facility; the metadata to search and find public studies of interest is hosted on Amazon AWS. The publicly shared data is made available through several interfaces: 1) directly downloadable from the public-facing website (Amazon AWS), 2) via MySQL database 3) via interfaces such as RImmport, an R package offering the data in the CDISC standard ([http://www.immport.org/immportopen/resources/docs/BISC\\_System\\_Architecture\\_and\\_Software\\_Design\\_Specification\\_v5.0.pdf](http://www.immport.org/immportopen/resources/docs/BISC_System_Architecture_and_Software_Design_Specification_v5.0.pdf)) (<https://www.bioconductor.org/packages/release/bioc/html/RImmPort.html>).

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



### XIII. Data discovery and identification

Compliance Level: 4

**R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.**

The ImmPort repository has some ways to access the data, as mentioned before, 1) the public facing website running on Amazon AWS, 2) the MySQL database, 3) RImmport. The public facing website allows for users to do queries through a search interface (<http://www.immport.org/immport-open/public/home/studySearch>). The website also offers the ability for filtering of data by important demographic variables such as species, whether the data is from a clinical trial, or what sample the data came from. The repository exposes the metadata via a web interface on Amazon AWS for web crawlers. An API is in development to allow for customized data calls and even great facility for machine harvesting. The repository is included in Nature Scientific Data's list of recommended repositories under Cytometry and Immunology (<https://www.nature.com/sdata/policies/repositories>). The repository offers persistent identifiers in the form of DOI for each study dataset to allow for data citations.

#### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## XIV. Data reuse

Compliance Level: 4

**R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.**

The repository adopts community standards via their direct work and feedback from the community. This is obtained in scientific meetings, user group meetings, and feedback during data submission. For instance, flow cytometry used the Flow Cytometry Standard (FCS) for any flow cytometry data submitted. Nevertheless, the technologies and data are constantly evolving; therefore, the continuous feedback and outreach are received. For instance, the ImmPort team attends weekly data standard calls for the HIPC program (Human Immunology Project Consortium) discussing new data types such as single cell RNA sequencing that will involve the generation of new metadata templates and minimal information standards for data reuse. The UH2 funding opportunity also provides resources for the community to come up with ideas on how to facilitate reanalysis of the data which could involve the generation or synthesis of new datasets incorporating new metadata for easier understanding and reuse of the data.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## TECHNOLOGY

### XV. Technical infrastructure

Compliance Level: 4

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

The data reside on NIAID hosting facility. The hardware, networking systems, and applications are maintained and administered by staff from the Office of Computational Infrastructure and Computational Biology. The NIH network is operated by the CIT (Center for Information Technology) of NIH. Staff at NIAID or NIH operate all standard hosting activities including OS administration of servers and network management according to well-defined laws and regulations (e.g., FISMA). Metadata to facilitate search and retrieval of publicly shared data is distributed via the web using Amazon AWS. Search is facilitated by indexing of study information with Apache Lucene SOLR (e.g., to search for studies on a topic using the public-facing website). Data is organized using a schema of tables in a MySQL database. Transfer of large data files is facilitated using Aspera connect technology. See [http://www.import.org/importopen/resources/docs/BISC\\_System\\_Architecture\\_and\\_Software\\_Design\\_Specification\\_v5.0.pdf](http://www.import.org/importopen/resources/docs/BISC_System_Architecture_and_Software_Design_Specification_v5.0.pdf).

#### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## XVI. Security

Compliance Level: 4

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Since the repository is a Federal information system, it complies with security policies and requirements for Federal and HHS Information Security Program, <http://csrc.nist.gov/drivers/documents/FISMA-final.pdf>.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## APPLICANT FEEDBACK

Comments/feedback

Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*