



ImmPort Repository

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository, National repository system; including governmental

Reviewer Entry

Reviewer 1

Comments:
accept

Reviewer 2

Comments:
accept

Brief Description of Repository

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) and TRUST Principles (Transparency, Responsibility, User Focus, Sustainability and Technology); 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. DAIT serves as the funder and is the entity applying for the certification of the ImmPort repository. 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 "ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, Hu Z, Zalocusky KA, Shankar RD, Shen-Orr SS, Thomson E, Wiser J, Butte AJ. Sci Data. 2018 Feb 27;5:180015. doi: 10.1038/sdata.2018.15".

Reviewer Entry

Reviewer 1

Comments:
accept

Reviewer 2

Comments:
accept

Brief Description of the Repository's Designated Community.

The ImmPort data repository serves the broader immunology research community to explore the wide spectrum of clinical and basic research data and associated findings. The community consists of national and international immunological clinical, translational and basic science researchers from a variety of institutional affiliations including academia, government, pharmaceutical and non-profits. Immunology researchers utilize ImmPort to explore the possibilities of reproducibility, reuse and secondary analyses of immunology data. Open-access datasets are being applied in the validation of the methods used in the original studies, leveraging studies for meta-analysis, or generating new hypotheses. To promote these goals, ImmPort provides advanced information technology and curation support in the archiving and exchange of scientific data and serves as a long-term, sustainable archive of immunology research and clinical data.

Reviewer Entry

Reviewer 1

Comments:
accept

Reviewer 2

Comments:
accept

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

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
accept

Comments

The ImmPort curation team ensures that the data is converted to the ImmPort data model incorporating relevant metadata to allow for data deposition and linking to partner databases (<https://www.immport.org/shared/dataModel>). This also ensures compliance with all required laws and regulations, as well as domain-specific scientific guidelines including human subject de-identification process and sensitive genotypic information. The documentation of the processes and guideline is available publicly at <https://www.immport.org/resources/documentation> . ImmPort adopts, adapts and creates 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 ImmPort Data Upload Templates (<https://downloads.immport.org/data/upload/templates/ImmPortTemplates.zip>).

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

accept

Outsource Partners. If applicable, please list them.

- ImmPort uses Federal contract mechanism to support its operation (<https://sam.gov/opp/bb587d89cc25c38706812ed5dcb25c50/view>). DAIT provides funding, governance, Statement of Work, and program administration for the ImmPort contract. An internal division within NIAID, Office of Cyber Infrastructure and Computational Biology (OCICB) provides hosting services for the repository's backend production system. Peraton is the primary contractor for the ImmPort repository. Peraton subcontracts UCSF, ESAC Inc, Yale University, and others as needed to acquire complementary leadership, scientific and technical expertise. Dr. Atul Butte of UCSF serves as the Principal Investigator of the contract. The contract covers all operational aspects of running a repository including mechanisms for data submission, data curation and data availability to the community following the FAIR principles. Additionally, outreach to the community is also undertaken by the contractors.
- The Request for Proposal (RFP) for BISC (previous name for the ImmPort contract) has more information about the contract. The RFP is at <https://sam.gov/opp/bb587d89cc25c38706812ed5dcb25c50/view>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

accept

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.
- ImmPort aspires to collaborate and connect with additional partners. Currently, ImmPort makes its metadata searchable via Vivli (<https://vivli.org/members/ourmembers/>) and is a member.
- ImmPort serves the immunology community for data sharing including a growing number of large consortia [Human Immunology Project Consortium (HIPC; <https://www.immport.org/resources/hipc>) from DAIT, and scientific programs from other Divisions of NIAID, and other NIH Institutes, such as National Cancer Institute (NCI) such as Seronet (<https://www.immport.org/resources/seronet/>), National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) such as Accelerating Medicines Partnership (AMP; <https://www.immport.org/resources/amp-ra-sle>)] and non-profits such as Bill and Melinda Gates Foundation (<https://www.immport.org/resources/bmgf>), March of Dimes (<https://www.immport.org/resources/mod>).
- As of February 1st, 2022, ImmPort has 525 studies, including 146 clinical trial data shared. Further, data shared via ImmPort has been reused and has resulted in 44 peer-reviewed publications to date (<https://www.immport.org/shared/publications>). The current counts of dataset and publication are at <https://www.immport.org/shared/home>

Summary of Significant Changes Since Last Application:• ImmPort has made progress in completing requirements to fully implement guidelines by the CoreTrustSeal. This is specifically true with reference to continuity of access to the repository. ImmPort has a continuity plan in place as described under the response to Section R3.

- To facilitate discovery of ImmPort datasets by generic search engines, ImmPort has implemented Schema.org metadata markup on its dataset description web pages. Please see Section R11 and the links provided in paranthesis for details (<https://www.immport.org/shared/publications>; <https://www.immport.org/shared/news>).
- ImmPort has also streamlined metrics to track reuse of the publicly shared datasets: one mechanism adopted by ImmPort is to track all publications that acknowledge the reuse of ImmPort shared data. Please see response to Section R14.
- The primary contractor of ImmPort is Peraton Corporation, which acquired the unit used to operate ImmPort at Northrop Grumman. The announcement of the sale is at <https://www.businesswire.com/news/home/20201207005967/en/Veritas-Capital-to-Acquire-the-Federal-IT-and-Mission-Support-Services-Business-of-Northrop-Grumman>.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
accept

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository
Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository
accept

Response:

The goals of ImmPort are stipulated by DAIT, the contract funder and are articulated on the ImmPort website (<https://www.immport.org/home> , <https://www.immport.org/about>). 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

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

II. Licenses

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

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. These are the federal, state and local laws such as HIPAA law (<https://www.hhs.gov/hipaa/for-professionals/privacy/index.html>) and Common rule adopted by federal agencies (<https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html>). The ImmPort repository follows all guidelines set forth by the National Institutes of Health for adhering to privacy rules as detailed here (https://privacyruleandresearch.nih.gov/pr_05.asp). Users are permitted to use the data in any lawful way as detailed in the User Agreement (<http://www.immport.org/agreement>). The agreement is governed by the federal courts in the District of Columbia should users not comply with the User Agreement. Any updates to the User Agreement are relayed to users via email provided at the time of registration for repository access. The license is available via the ImmPort website at all times. While the ImmPort team does not have appropriate resources to monitor compliance of data use by its users, the authors of datasets have the ability to pursue legal redress pursuant to federal law as applied by the federal courts in the District of Columbia.

Reviewer Entry

Reviewer 1

Comments:

While understanding that it might be challenging to identify instances of non-compliance, for recertification in three years please give some consideration to the consequences if noncompliance is detected. For now accept.

Reviewer 2

Comments:

accept

III. Continuity of access

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

ImmPort is funded by US Government contract. 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.

However, 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.

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 to continue the operation until 2022. The NIAID has committed to renew the ImmPort program for another five years after the current cycle ends at 2022 to extend the program at least to 2027.

In the Statement of Work (SOW) of the ImmPort contract, DAIT required the contractor to set aside a budget for initial and final transition. The contractor shall coordinate with the incumbent contractor and the NIAID 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. This aspect is documented as Option 6, Final transition (Page 16) of the SOW document available at <https://sam.gov/opp/bb587d89cc25c38706812ed5dcb25c50/view> .

The technical requirement details of ImmPort system and architecture, software design, data model is publicly available at <https://www.immport.org/resources/documentation> . The expected outcome of the initial transition is for the new contractor to completely take over the operation of the current system after 3 months of the award of contract. NIAID shall support the incumbent during the same 3 months by handing over inventory, providing necessary training, and connecting with the relevant contacts to continue the ImmPort operations.

Reviewer Entry

Reviewer 1

Comments:

It is well described that plans are available to ensure continuity in the repository.

Reviewer 2

Comments:

accept

IV. Confidentiality/Ethics

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

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 (<https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html>). 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 IRBs typically exist at the institutions that conduct the research and appropriate letters from IRBs are used to check compliance. Additionally, the ImmPort data curators are trained in the process of de-identification as well as conducting compliance checks. If the data does not comply as expected, the data is not made publicly accessible until the required compliance is achieved. Data with disclosure risk is not disseminated via ImmPort. Such data is directed to specialized repositories (such as dbGAP) and links are provided within ImmPort to the dataset. The User Agreement also specifies the legal ways in which the data may be used (<http://www.immport.org/agreement>) which concur with disciplinary and ethical norms.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

V. Organizational infrastructure

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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

The repository is currently funded through a contract from NIAID DAIT to Peraton. Peraton subcontracts UCSF, ESAC Inc, Yale University, and others as needed to acquire complementary leadership, scientific and technical expertise. The contract funds about 10 full-time employees (FTEs), who have expertise in data curation, data management, software and network engineering as well as project management. The scientific and technical aspects of the contract is overseen by the Contract Officer Representative (COR) who bears responsibility for ensure that the work is completed in line with the contracts' statement of work. All decisions regarding the repository management and operations are under the purview of the COR in consultation with the ImmPort management team. 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. Peraton has extensive experience as a contractor organization working on Federal government IT initiatives. The Peraton 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: <https://www.immport.org/shared/teamCurrent> .

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

VI. Expert guidance

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository
accept

Response:

The repository has a yearly external advisory board (<https://import.org/shared/teamCurrent>) 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. 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 as needed basis. The ImmPort team also attends weekly and monthly progress meetings with the program staff at DAIT to receive guidance and feedback. The team interact closely with the consortia members that submit their data to ImmPort, some of whom are also users of the data in ImmPort and solicit periodical feedback from the community.

Reviewer Entry**Reviewer 1**

Comments:

Reviewer 2

Comments:

accept

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry**Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository
accept

Response:

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 (<https://www.immport.org/shared/news>). It is also the case for any changes to the software powering the ImmPort resource (https://www.immport.org/docs/BISC_System_Architecture_and_Software_Design_Specification_v5.0.pdf). The database schema is publicly available (<https://www.immport.org/shared/dataModelDocumentation?table=study>). The database checks the identity of data providers and maintains a list of the studies and datasets that were shared by each provider (<https://www.immport.org/shared/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 (<https://www.immport.org/shared/news>).

At the backend, ImmPort utilizes an Oracle database that encrypts the data stored on it. The system disk volumes that hold all the database data (ASM disks), are configured as ASM disks, whereby multiple disks make-up an ASM Data diskgroup, such that there are several diskgroups for the ImmPort database. All data is stratified across all of the ASK disks, such that even data in a table is across all the ASM disks, and not contained in one volume uniformly. Data access (and visibility) can only be done from within the Oracle database. The ASM disks are encrypted such that the data cannot be seen at the O/S level, and they are not readable, unless it is read or modified by connecting to the database via SQL*Plus, an Oracle utility, or via a standard database connection to the active Oracle ImmPort database (example - JDBC connection), The Oracle data that is contained with Oracle ASM diskgroups is not just "hashed", it is encrypted via Oracle's database storage methodology.

Further, during the time of a transaction (example: an insert, update, or delete), Oracle performs a data integrity check for all parent key and foreign key constraints, that apply to referential integrity. Any database operation to be performed which does not agree with this data integrity constraint, throws an error and the operation is not performed. The user receives the constraint violation error(s). Thus, the data integrity is maintained for that data, or data table(s), and the operation is not performed.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

VIII. Appraisal

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

ImmPorts' mandate is to serve as a repository for the immunological community. Therefore, it accepts data pertaining to the immunological domain from its' community members. 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. The repository is data-type agnostic and will accept all types of data that are relevant to immunology research and do not have a designated public repository. The human subject should be de-identified. The number of data types are evolving with the technology and user need. The current data types are at <https://www.immport.org/resources/dataTemplates> . Major data submitters to ImmPort have been members of large consortia project funded by the DAIT division of NIAID, NIH as well as non-profit organizations such as March of Dimes and Bill & Melinda Gates Foundation. 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.immunospace.org/project/home/begin.view?pageId=HIPC%20Standards>) 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 (<https://www.immport.org/resources/dataTemplates>) . There are QC checks in place (e.g., the data validator <https://immport.niaid.nih.gov/login#!/validateData>) 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 (<https://www.immport.org/resources/dataTemplates>).

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

IX. Documented storage procedures

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

Relevant processes and procedures are documented online

(https://www.immport.org/docs/BISC_System_Architecture_and_Software_Design_Specification_v5.0.pdf) 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). Public data is hosted in the cloud on Amazon AWS, which also provides industry gold standard for security and fault tolerance capabilities (https://www.immport.org/docs/BISC_System_Architecture_and_Software_Design_Specification_v5.0.pdf).

All the ImmPort data is housed within NIAID by the Office of Cyber Infrastructure and Computational Biology where it is managed appropriately including backups, disaster management etc.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

X. Preservation plan

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

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Accept for level 3

Reviewer 2

Comments:

3 – The repository is in the implementation phase

this will need to be addressed more clearly in the reapplication phase. I would accept it as an answer for compliance level 3, but in the future this will need to be more elaborated.

Response:

Since the repository has been funded by a government contract, the repository assumes responsibility for long-term preservation as long as the funding continues. The ImmPort User Agreement (<https://www.immport.org/agreement>) in place delineates some of the responsibilities and liabilities undertaken by the ImmPort repository. The agreement also makes it explicitly clear that the data provider remains the owner of the data within ImmPort. Since most of the within ImmPort are those that are generated by the research community using the US government funding, the data sharing activity is the US Government exercising its rights as a funder to a copy of the data (<https://www.grants.gov/learn-grants/grant-policies.html>). All data submitters to ImmPort are cognizant of the repository's right to transform the data to comply with submission standards as delineated in ImmPorts' data model (<https://www.immport.org/shared/dataModel>). ImmPort repository's approach for defining and implementing data

preservation activities is publicly documented at the links provided in parenthesis (Data Collection section of <https://www.nature.com/articles/sdata201815> ; <https://www.immport.org/resources/documentation>).

As part of the NIH Strategic Plan for Data Science (<https://datascience.nih.gov/nih-strategic-plan-data-science>), the NIH is committed to providing permanent and accessible solutions to house datasets resulting from NIH investigator publications. Under this initiative, the ImmPort repository has been able to generate redundancies of its data holdings by maintaining a copy of the data and metadata annotations in a cloud environment in addition to the on-prem storage provided by OCICB. Some approaches for long-term preservation that are currently being explored for ImmPort data holdings are:

- a) Usage of digital preservation software such as Rosetta (by Ex Libris),
- b) Storage of the data by National Archives (maintained by the US Government)
- c) Establishing MoUs with complementary repositories to archive ImmPorts' data holdings.

Further, a formal preservation policy and preservation plan that details the final preservation mechanism adopted by the ImmPort repository is being developed and will be made publicly available after review and approval by the appropriate authorities.

Reviewer Entry

Reviewer 1

Comments:

Throughout the application it is clear that the repository defines and implements preservation actions. This is also reflected in the provided "documentation" section of the "resource" -page, see last link provided by applicant under response statement R10. Currently accept for level 3, with the expectation that a preservation plan that integrates this pieces of evidence will be provided at the renewal in three years.

Reviewer 2

Comments:

accept

XI. Data quality

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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

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 (<https://www.immport.org/shared/publications>). This is relevant for citation purposes to link the data via associated ImmPort study ID to publication.

Additionally, ImmPort facilitates discovery of datasets by general search engines via metadata markup of the individual dataset page. ImmPort follows the Schema.org Dataset definition for this purpose. Schema.org, founded by Google, Microsoft, Yahoo and Yandex, is a vocabulary that is a collaborative and community activity that define structured data covering entities, relationships between entities and actions, and can easily be extended through a well-documented extension model.

Reviewer Entry**Reviewer 1**

Comments:

If ImmPort made progress on an ability to comment on, and/or rate data and metadata that would be good to see reflected in the upcoming recertification in three years. However, response statement clear enough to accept R11.

Reviewer 2

Comments:

accept

XII. Workflows

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

ImmPort has documented processes for each stage of the data archiving process from ingest to dissemination. The complete documentation of the process is available at <https://www.immport.org/resources/dataTemplates> . 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 (decision tree available here -

<https://www.immport.org/images/home/ImmPort.Data.Deposition.Decision.Tree.pdf>). 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) programmatic access via APIs (<https://docs.immport.org/#API/DataQueryAPI/dataqueryapi/>) 4) via interfaces such as RImmport, an R package offering the data in the CDISC standard

(https://www.immport.org/docs/BISC_System_Architecture_and_Software_Design_Specification_v5.0.pdf)

(<https://www.bioconductor.org/packages/release/bioc/html/RImmPort.html>).

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

XIII. Data discovery and identification

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

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) APIs, 4) RImmport. The public facing website allows for users to do queries through a search interface (<https://www.immport.org/shared/home>). 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 (<https://www.immport.org/shared/search>). The repository exposes the metadata via a web interface on Amazon AWS for web crawlers. Additionally, ImmPort facilitates discovery of datasets by general search engines via metadata markup of the individual dataset page. ImmPort follows the Schema.org Dataset definition for this purpose. Schema.org, founded by Google, Microsoft, Yahoo and Yandex, is a vocabulary that is a collaborative and community activity that define structured data covering entities, relationships between entities and actions, and can easily be extended through a well-documented extension model.

The repository is included in Nature Scientific Data's list of recommended repositories under Cytometry and Immunology (<https://www.nature.com/sdata/policies/repositories>), PLOS One under Biomedical Sciences (<https://journals.plos.org/plosone/s/recommended-repositories#loc-biomedical-sciences>) and NIH's Trans-NIH BioMedical Informatics Coordinating Committee (BMIC) list as a domain-specific repository (https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html). The repository offers persistent identifiers in the form of DOI for each study dataset to allow for data citations.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

XIV. Data reuse

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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

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, any flow cytometry data submitted to ImmPort follows the Flow Cytometry Standard (FCS). Nevertheless, with technologies and types of data evolving constantly, continuous feedback and outreach is maintained to stay up-to-date with community standards. 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 (<https://grants.nih.gov/grants/guide/pa-files/PAR-19-229.html>) 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.

A primary metric utilized by ImmPort to evaluate reuse of publicly shared data is in the area of Scientific Impact. A key surrogate metric for measuring scientific impact is published articles that cite reuse of ImmPort shared studies. To date, 44 research articles have been published by the scientific community that reuse ImmPort shared data and a listing of the same is available at <https://www.immport.org/shared/publications>. The increasing number of publications that have reused datasets shared via ImmPort over the years are testament to the fact that the immunology community is able to

utilize the data in the format that it's shared. Further, the data standards used within ImmPort are all open access standards and allow us to update as the formats and standards themselves evolve.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

TECHNOLOGY

XV. Technical infrastructure

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.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

accept

Response:

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 (OCICB). 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 <https://www.immport.org/resources/documentation> .

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

accept

XVI. Security

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

accept

Response:

The ImmPort system is hosted as a hybrid architecture with the metadata and backend system hosted on AWS and the data residing on premise on an NIAID server. 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> . The NIAID OCICB plans, implements, and manages NIAID's cyber technology infrastructure. The OCICB provides these services from multiple data centers that are geographically distributed. OCICB architects, implements, maintains and supports the infrastructure that hosts ImmPort applications, (<https://bioinformatics.niaid.nih.gov/services#Applicationhosting>). OCICB also manages the network and cyber security

via network firewalls, monitoring network activity, and ensuring that known security vulnerabilities are addressed according to NIAID policies and industry best practices.

Additionally, while browsing study metadata within the ImmPort repository is free, both upload and download of datasets requires authentication by the ImmPort system. Registration mechanism is available for ImmPort system. The "Hardware Infrastructure" of the following document provides more details,

https://immpport.org/docs/BISC_System_Architecture_and_Software_Design_Specification_v5.0.pdf .

Reviewer Entry

Reviewer 1

Comments:

Accepted based on high-level approaches to security risks/threats as outlined in the evidence provided

Reviewer 2

Comments:

accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: