Assessment Information

Repository:	Immune Epitope Database and Analysis Resource
Website:	http://www.iedb.org/
Certification period:	01 October 2023 - 30 September 2026
Requirements version:	CoreTrustSeal Requirements 2020-2022

This repository is owned by: **La Jolla Institute for Immunology**

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background Information

Repository Type

Please provide context for your repository. You can select one or multiple options.

Compliance level:

Not Applicable - 0

Response:

- Domain or subject-based repository

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Description of Repository

Provide a short overview of the repository.

Compliance level:

Not Applicable - 0

Response:

Immune Epitope Database and Analysis Resource

The Immune Epitope Database and Analysis Resource (IEDB) is the most extensive, centralized resource for searching and analyzing data related to antibody and T cell epitopes for humans, non-human primates, rodents, and other animal species. Established in 2003 under the leadership of Dr. Alessandro Sette at the La Jolla Institute for Immunology, the IEDB (http://www.iedb.org) has been funded by the National Institute of Allergy and Infectious Diseases (NIAID), a component of the National Institutes of Health in the U.S. Department of Health and Human Services, since its inception.

At its core, the IEDB is a freely available epitope database and prediction resource. An epitope, also referred to as an antigenic determinant, is the portion of an antigen that is recognized by the adaptive immune system. It is a chemical structure recognized by specific receptors - antibodies, major histocompatibility (MHC) molecules, and T cell receptors. As such, epitopes play critical roles in many diseases, including infectious diseases, autoimmune diseases, and allergies. They are also involved in organ transplants and blood transfusions. Epitopes can be characterized as continuous and discontinuous. Continuous, or linear, epitopes are sequences of amino acids. Discontinuous epitopes, sometimes referred to as conformational, are composed of discontinuous segments of amino acids of one or more chains. The definitions of continuous and discontinuous also apply when one or more amino acids are replaced with a molecular entity that is not a peptide, such as a lipid. The study of epitopes has important applications in understanding the triggers of adaptive immune responses and in developing new vaccines, diagnostics, and therapeutics.

The IEDB has two avenues by which data is collected; biocuration of published, peer-reviewed articles and external data submissions from the research community. As of November 2020, the IEDB data has been derived from over 21,650 peer-reviewed journal articles. In addition, the IEDB contains over 300 direct submissions (corresponding to approximately 30% of the total data) from several NIH-funded large-scale epitope discovery programs and from researchers who directly approach the IEDB to deposit their data. This also includes negative data, which might typically appear in supplemental tables or might not be published at all. Slightly more than half of the references in the IEDB relate to infectious diseases and about one quarter relate to autoimmune diseases, apart from HIV, which is captured separately in the Los Alamos HIV database. The remainder includes allergy, transplant, and other categories. Given that the data in the IEDB resides in the public domain, researchers can freely access, analyze, and publish work using this data. It should be stressed that the database exclusively contains information about epitopes that have been derived from experiments. No predicted or modeled data resides in the database.

The curation of scientific literature started in 2004, requiring the curation of past and current relevant epitope literature in available peer-reviewed journals. As the IEDB has evolved, it has been necessary to change how biological concepts are captured in order to maximize accuracy. In addition, automated validation is continuously added. Consequently, there is a significant ongoing "recuration" effort of revising existing entries to improve data quality and consistency. The IEDB is now current with the published literature, and targeted PubMed queries are run biweekly, with the goal of making the data available in the IEDB within eight weeks of publication.

Attached is the current organizational and team structure of the IEDB; a resource developed under the leadership of the La Jolla Institute for Immunology, consisting of 5 key teams with assigned leads, and 2 subcontractors for additional expertise.

Recent articles provide additional context for the IEDB more broadly:

Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 2018 Oct 24. doi: 10.1093/nar/gky1006. PMID: 30357391; PMCID: PMC6324067.

Fleri W, Vaughan K, Salimi N, Vita R, Peters B, Sette A. The Immune Epitope Database: How Data Are Entered and Retrieved. J Immunol Res. 2017;2017:5974574. doi: 10.1155/2017/5974574. Epub 2017 May 29. Review. PMID: 28634590; PMCID: PMC5467323.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Designated Community

Provide a clear definition of the Designated Community

Compliance level:

Not Applicable - 0

Response:

The Immune Epitope Database and Analysis Resource (IEDB) is the single largest repository of immune epitope data in the world. It has an international user community that includes biomedical researchers in academia, non-profit research institutes, and industry. The users include immunologists, microbiologists, virologists, and bioinformaticians who are interested in developing new vaccines, diagnostics, and therapeutics, and who want to study the adaptive immune system. All data is freely available to users via http://www.iedb.org, and the data can also be downloaded in various, user-friendly formats.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Level of Curation

Select all relevant types of curation.

- Content distributed as deposited
- Basic curation – e.g., brief checking, addition of basic metadata or documentation
- Enhanced curation – e.g., conversion to new formats, enhancement of documentation
- Data-level curation – as above, but with additional editing of deposited data for accuracy

Compliance level:

Not Applicable - 0

Response:

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

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Level of Curation - explanation

Please add the description for your Level(s) of Curation.

Compliance level:

Not Applicable - 0

Response:

The IEDB data procurement requires systematic identification, categorization, curation and quality■checking processes, all of which have been documented in the publicly available IEDB curation manual (http://curationwiki.iedb.org/wiki/index.php/Curation_Manual2.0). The curation manual is regularly updated by our Lead Ontology and Quality Manager to reflect the latest curation procedures.

The information curated in the IEDB is derived from the scientific literature catalogued in PubMed, and from direct data submissions from users, or more often from various NIH■sponsored research efforts. Accordingly, for each epitope, it is necessary to extract the scientific details defining the assays in

which the epitopes are defined and studied. The same epitope might be studied in multiple publications or submissions, and the same epitope is commonly tested in multiple assays, sometimes with different outcomes (e.g. a virus■specific antibody might bind in an enzyme■linked immunosorbent assay format but will not neutralize the live virus). Such nuances in the assay parameters accompanying the epitopes are objectively made available to view, based on the end■user's defined query.

A highly specialized expert curation process is pivotal to ensuring the consistency of such highly contextual data. This is achieved by PhD-level curators who review the data, and systematically extract and deposit the relevant information into a highly structured, computer■operable format. For instance, the relevant information is not confined to a single location of the manuscript. Rather, the structure of the epitope as well as the contextual assay details may be reported in the methods section, while the data and interpretation of the results are often found in figures, tables and text. The curator is therefore challenged to faithfully synthesize these disjointed elements of a publication into a concise format for user consumption. To ensure that the data is accurately represented, all curations are peer reviewed by a second PhD-level curator to ensure accuracy, and are updated as required. External data submissions are also reviewed and updated by curators to ensure the data is converted into the IEDB format. Of course, we require final approval from the external submitter prior to publishing the data.

Our curation process certainly adds value by enhancing the content to ensure it is understandable by users in a simple, tabulated form, rather than distributed through one or more manuscripts. Our curators also contact authors to obtain any additional information for data completeness (e.g. additional results, supplementary figures, etc.), which may not be within the user's purview to do so.

The below recent article provides additional context to the curation process:

Salimi N, Edwards L, Foos G, Greenbaum JA, Martini S, Reardon B, Shackelford D, Vita R, Zalman L, Peters B, Sette A. A behind-the-scenes tour of the IEDB curation process: an optimized process empirically integrating automation and human curation efforts. Immunology. 2020 Jul 2;161(2):139–47. doi: 10.1111/imm.13234. Epub ahead of print. PMID: 32615639; PMCID: PMC7496777.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Insource/Outsource Partners

If applicable, please list them.

Compliance level:

Not Applicable - 0

Response:

The website and data reside at the La Jolla Institute for Immunology (LJI) in La Jolla, San Diego, California. There is a back-up site at the San Diego Supercomputer Center (SDSC), an organized research unit of the University of California, San Diego (UCSD). There is no SDSC certification, nor requirement to certify, that we are aware of.

After further review, we have attached the SDSC SLA agreement. They have Colocation (COLO) service terms, which can be found on their website here: https://www.sdsc.edu/assets/docs/COLO_terms_of_service_2022.pdf. We have also attached the formal agreement between the La Jolla Institute and SDSC, which has been in place since 2010, and abides by these COLO terms.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Significant Changes

Summary of Significant Changes Since Last Application if applicable.

Compliance level:

Not Applicable - 0

Response:

-

Links:

Reviews

Reviewer 1:

Immune Epitope Database and Analysis Resource

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Other Relevant Information

You may provide other relevant information that is not covered by the requirements.

Compliance level:

Not Applicable - 0

Response:

The IEDB websites typically receive a median of 20,000 visits per month, and this has grown dramatically in 2020 to 32,000 monthly visits based on Q1-3 data. This can likely be attributed to the growth in interest since the discovery of SARS-CoV-2 and use of the database and tools for research in this novel area. In 2020 thus far, the geographic breakdown of users (measured by visits to the main website) included Asia (39.6%), the Americas (36.7%), Europe (19.6%), Africa (2.2%), and Oceania (1.6%).

In terms of citations, the IEDB received 2,676 individual citations in 2019 (excluding self-citations), which is an increase of 462 citations from 2018. As of 2019, authors have been asked to cite one IEDB publication; "The Immune Epitope Database (IEDB): 2018 update", which was published in Nucleic Acids Research. According to Google Scholar as of November 2020, this paper has been cited 267 times. Prior to this, authors were asked to cite one of three different papers published by the IEDB team in 2005, 2009, and 2014. According to Google Scholar as of November 2020, the original paper, "The Immune Epitope Database and Analysis Resource: From Vision to Blueprint", has received 435 citations. The second paper, "The immune epitope database 2.0", has received 655 citations. The third paper, "The immune epitope database (IEDB) 3.0", has received 747 citations. In addition, many authors will cite the IEDB and/or its URL in their article without citing one of these three papers. We will complete the 2020 citation analysis in July 2021, and expect another considerable increase based on the growth of website visits in 2020. In addition, 72 US patent families cited or used the IEDB in 2019, which is 8 more than in 2018. The IEDB is projected to be cited by 84 US patent families in 2020, based on 56 records retrieved at the end of August 2020.

Overall, these statistics show that the IEDB has a global user base and its positive impact is far reaching into the scientific community.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Organizational Infrastructure

R1 Mission/Scope

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

The goals of the IEDB are stipulated in the contract with the National Institute for Allergy and Immunology (NIAID). According to the Statement of Work, the contractor (LJI) is to:

- Maintain and further enhance a central web-based source of information on T cell epitopes and linear and conformational antibody/B cell epitopes (e.g., carbohydrates, lipids, and modified peptides) through curation of existing literature and direct submissions by the broader research community.
- Maintain and further enhance a central web-based source of data on ligand binding to MHC class I, class II, non-classical, and MHC-related molecules, including ligands shown experimentally not to bind to any of these molecules (i.e., negative binding data).
- Maintain and further enhance a central source of data on BCR and TCR repertoire information associated with T cell and antibody/B cell epitopes located within the IEDB.
- Foster further development of an Analysis Resource within the IEDB composed of more robust algorithms, mathematical models, and other predictive tools that support:
- Identification of novel antibody/B cell and T cell epitopes from genome or protein sequence information and predicting host responses to specific pathogens or immune-mediated diseases;
- Draw connections between both BCR and/or TCR repertoire sequence data, epitope binding and computational identification of epitopes from TCR/BCR sequence; and/or
- Facilitate identification of antibody and T cell epitopes associated with infectious or immune-mediated diseases for their use as targets for vaccine candidates and/or immune-based therapies.

Furthermore, the Statement of Work lists four additional components for fulfilling the aforementioned scope:

- Maintain, further develop, and improve the IEDB's web-based relational database populated with antibody/B cell epitope and T cell epitope information. The IEDB will be freely accessible to the scientific community via an internet website and immune epitope information will be obtained primarily through curation of the scientific literature (relevant journal articles) and direct submissions from the broader research community.
- Maintain, enhance, further develop, and optimize the Analysis Resource for the IEDB. This includes online access to: (1) tools to help researchers locate and analyze information contained in the IEDB; (2) other relevant databases and related information; (3) data mining algorithms, mathematical models, and other sophisticated analytical tools to help researchers.

- Community Outreach activities to expand the user base and utility of the Immune Epitope Database resource for the broader research community.
- Interact with both current and future NIAID programs, which minimally include:
- Contractors supported by the B cell Epitope Discovery and Mechanisms of Action, Large-scale T cell Epitope Discovery, and the Allergen Epitope Discovery programs;
- Contractors supported by the Bioinformatics Integration Support Contract (BISC), Bioinformatics Resource Centers (BRCs) and the HIV Molecular Immunology Database.

It is a contractual obligation for the IEDB team to preserve and continue providing access to the data for the duration of our contract, and at the culmination of our contract, we will ensure all data is transferred to the incumbent (or back to NIAID) efficiently (see further details on this in R3).

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R2 Licenses

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

Data contained in the IEDB website is within the public domain, free of all copyright restrictions, and made fully and freely available for both non-commercial and commercial use. IEDB data is manually curated, either from experimental data shown in publications or submitted datasets, and the published data is linked to a PubMed identifier, and the submitted data is linked to a submission identifier. All data is attributed to the publishing or submitting authors. This work is licensed under a Creative Commons Attribution 4.0 International License and has been in place since May 2017. Users of the IEDB database are simply asked to cite the IEDB when using the resource, which can be found here. More information on our Creative Commons license can also be found via the citation website, but, in summary, the license enables users to:
- Share — copy and redistribute the material in any medium or format

- Adapt — remix, transform, and build upon the material for any purpose, even commercially.

As stipulated in the Creative Commons license, these actions apply under the following terms:

- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

- No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Further information on the IEDB's use is available on the IEDB Terms of Use web page.

The IEDB Analysis Resource (tools) is freely available to academic users through an open-source license, whilst commercial licenses are available to those wanting to utilize the tools within their private network. As agreed with the reviewers, this licensing scheme is outside the scope of this question, but more information can be found on this website if required.

Links:

- [IEDB Analysis Resource (Tools) Open-Source License](#)
- [How to Cite the IEDB](#)
- [Creative Commons Attribution 4.0 International License](#)
- [IEDB Terms of Use](#)

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R3 Continuity of access

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

Immune Epitope Database and Analysis Resource

The IEDB is in its third funding cycle with NIAID, which is a seven-year contract from December 2018 to December 2025. Prior to this, the IEDB was funded by a contract mechanism with NIAID for two additional funding cycles of eight and seven years. A fully developed transition plan has been a deliverable for all three contracts to enable a smooth transition of the database, and all respective user interfaces, from the incumbent to a new awardee or to the government, in the event of the incumbent not being selected for the contract renewal.

Hence, as stipulated in the Statement of Work in the subsection "Information Technology (IT) Resources, Facilities and Security", the IEDB has a Continuity of Operations Plan, which includes a comprehensive Operational Recovery Plan (ORP) and Disaster Recovery Plan (DRP) that specifies the procedures used to restore operations following a natural or man-made disaster. This was submitted to NIAID in January 2019.

In agreement with the NIAID Contracting Officer, Emily Dubbaneh Bannister, and our IEDB Program Officer, Joseph Breen, we have now made the following documents publicly available in our Solutions Center:
- Operational Recovery Plan (ORP) and Disaster Recovery Plan (DRP) -
https://help.iedb.org/hc/en-us/articles/4406591519515--IEDB-Operational-Recovery-Plan-ORP-and-Disaster-Recovery-Plan-DRP-
- IEDB Contract 3 Statement of Work - https://help.iedb.org/hc/en-us/articles/4406597263771-IEDB-Statement-of-Work-SOW-with-NIAID

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text: Are these operational plans available on your public website?

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R4 Confidentiality/Ethics

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

The data collected and distributed by the IEDB is considered public data and does not present ethical disclosure risks. The IEDB does not distribute author contact details in excess of what is already publicly available in PubMed; for example, curated literature contact information for the corresponding author is provided by the journal in which the article appears, hence is also available in the IEDB. Data submitted to the IEDB, as opposed to data curated from literature, are not promoted to the IEDB website for public access until the submitter releases them via a web interface. Therefore, submitters also review and approve contact details prior to publishing the data. This action must be performed by the submitter and cannot be performed by the IEDB team. The delay in the public release of data is typically done to allow time for the data to be published.

In regards to ensuring that deposited data was obtained under ethical conditions, the IEDB does not undertake any additional ethics clearance checks. This is because all externally submitted data is from NIH epitope contracts whose projects undergo ethical screening prior to data collection, especially when human or live specimens are in question. Therefore, there is no further assessment to be done by the IEDB. Similarly, when curating published literature from PubMed, these studies have already passed ethics approval and peer review, hence we do not perform additional checks.

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R5 Organizational infrastructure

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:

The guideline has been fully implemented in the repository - 4

Response:

The repository is currently funded through a contract (75N93019C00001) from the National Institute of Allergy and Infectious Diseases (NIAID) to the La Jolla Institute for Allergy and Immunology (LJI). The IEDB is in its third funding cycle with NIAID, which is a seven-year contract from December 2018 to December 2025. Prior to this, the IEDB was funded by a NIAID contract mechanism for two funding cycles of eight and seven years, therefore the repository has adequate funding and continuity.

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. NIAID research, in particular, strives to understand, treat, and ultimately prevent the myriad infectious, immunologic, and allergic diseases that threaten millions of human lives. LJI has extensive experience as a contractor organization working on NIAID grants and contracts. There is a clear governance structure from NIAID, with our Program Officer (PO), Dr. Joseph Breen, overseeing all major funding decisions regarding the IEDB. The IEDB leadership team also presents monthly updates to the PO and written report updates on IEDB goals on a quarterly basis to both the PO and Contracting Office (CO) at NIAID.

The IEDB's scientific direction includes the leadership of Dr. Alessandro Sette, the Principal Investigator of the contract, who is a recognized leader in the area of immunology. He has considerable knowledge of immune epitopes, with a focus on the identification and biology of immune epitopes for infectious and immune-mediated diseases. In this respect, Dr. Sette has been the PI of several NIAID contracts for almost 30 years, including large-scale epitope identification contracts targeting smallpox/vaccinia virus, arenaviruses, dengue virus, mycobacterium tuberculosis, pertussis, and allergies. Dr. Sette has been the PI of the IEDB since its inception in 2003. Assisting him is Dr. Bjoern Peters, co-Principal Investigator, a bioinformatician who has been working on the IEDB since early 2004. His training in computer science, mathematics, and quantitative modeling, coupled with almost 20 years of working directly with clinicians, immunologists, and biochemists, uniquely qualifies him to integrate the computational and experimental components of the IEDB. Therefore, the leadership team is well-qualified to lead the IEDB.

At the team level, a sufficient number of qualified staff is managed through a clear governance structure to carry out the mission. The IEDB team includes PhD-level biocurators, bioinformaticians, database administrators, IT specialists, and project managers. The IEDB project is structured into 5 key teams; Curation (led by Dr. Alessandro Sette), Query & Reporting (led by Dr. Bjoern Peters), Tools (led by Dr. Bjoern Peters), IT Infrastructure (led by Dr. Jason Greenbaum - LJI Bioinformatics Core Director) and Outreach (led by the IEDB Project Manager). In addition, LJI has subcontracted the Technical University of Denmark (DTU) and Leidos Inc. to acquire complementary leadership, scientific and technical expertise.

Overall, the IEDB is both governed and comprised of highly skilled individuals in a structured manner, to ensure that the mission is executed effectively. More information about the current IEDB team can be found at https://help.iedb.org/hc/en-us/articles/115000071491-Acknowledgements. The IEDB team is actively working to improve our support materials, including the acknowledgments. We are in the process of implementing a new support platform, Discourse, which will be available in 2022, and will increase the visibility of this information.

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R6 Expert guidance

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

The IEDB participates in an annual epitope meeting sponsored by NIAID that includes the NIAID large-scale epitope discovery contracts, which has ranged from 10-20 projects over the years. At this meeting, the IEDB presents a status update and future plans. Feedback from this collection of epitope experts is solicited to improve data quality, expand query and reporting features, and facilitate their data submission process. The meeting is also attended by NIAID staff with expertise in the field. In addition, since 2007 the IEDB team at LJI has published almost 30 meta-analyses of the data in the IEDB relating to a particular field, such as influenza A, tuberculosis, Ebola virus, and diabetes. These studies have created opportunities to interact with domain experts in specific fields of interest that have resulted in improved data quality and completeness.

The IEDB team also has access to the expertise of over 20 immunology faculty at LJI, immunological and disease experts at local research organizations, including Salk Institute, Scripps Research Institute, Sanford Burnham Prebys Medical Discovery Institute, and UC San Diego. Several team members are active participants on a variety of scientific advisory boards where they interact with domain experts in immunology, biocuration, ontology, and diseases, and bring back new ideas for enhancing the IEDB.

In addition, as part of our outreach efforts, IEDB staff attend scientific conferences and meetings to present information about the IEDB, its data, and its uses, and to gather feedback from the experts in attendance. The IEDB team also hosts an annual user workshop, which commenced in 2012. In addition to educating new and experienced users from a variety of backgrounds, one of the stated goals is to solicit feedback and comments on current and future features. These workshops have been a valuable source of new ideas for further development and prioritization for the team. The most recent 3 user workshops (2020-2022) can be seen in our Solutions Center via the following links:
- 2020 - https://help.iedb.org/hc/en-us/articles/360052475011-2020-IEDB-Virtual-User-Workshop-Presentations
- 2021 - https://help.iedb.org/hc/en-us/articles/4409650396571-2021-IEDB-Virtual-User-Workshop-Presentations
- 2022 - https://help.iedb.org/hc/en-us/articles/10097021647131-2022-IEDB-Virtual-User-Workshop-Presentations
All workshop recordings can be accessed via our IEDB YouTube channel. The latest information for the 2023 user workshops can be found here. Finally, the IEDB has also contracted usability engineering consultants to improve the user experience in accessing data.

Furthermore, in 2020, we established an IEDB Expert Committee, comprised of 17 power users of the IEDB database and tools. This group ranges from graduate and postdoctoral students to PI and NIAID-level representatives, specializing in both B and T cell research. We engage with this committee on a monthly basis, demonstrate new work, and solicit feedback to improve key features. Details of the Expert Committee can be found here in our Solutions Center (https://help.iedb.org/hc/en-us/articles/360057231112-IEDB-Expert-Committee), and it provides links to the 2 projects they have provided input on (IEDB Filter Options and the IEDB Query API - see links below).
- https://help.iedb.org/hc/en-us/articles/360053990892-What-are-the-IEDB-Filter-Options-
- https://help.iedb.org/hc/en-us/articles/4402872882189-Immune-Epitope-Database-Query-API-IQ-API-

However, the IEDB team is actively working to improve our support materials, including the Expert Committee page. We are in the process of implementing a new support platform, Discourse, which will be available in the future, and will increase the visibility of this information. We will be updating the details of the members and including headshots, as well as a path to get involved with the group if desired. At this stage, the exact deliberations of the group are not published, but this is something we can consider in our updated support platform.

Lastly, as part of the IEDB contract, we host monthly teleconferences with our NIAID Program Officer, Dr. Joseph Breen. This provides an avenue to receive expert input and advice based on NIAID strategic priorities. This is imperative to ensure that we continue to meet NIAID expectations, as well as our users' expectations.

Overall, we have multiple mechanisms to seek feedback, both internally and externally to the LJI team, and have avenues to access user feedback, strategic feedback from NIAID, and scientific input from other leaders in the field.

Links:

- [IEDB YouTube Channel](#)
- [IEDB User Workshop Website](#)

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text: Are the deliberations of the expert advisories and workshops published or available on your website? I could only find the membership for the advisory committee online

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Digital Object Management

R7 Data integrity and authenticity

The repository guarantees the integrity and authenticity of the data.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

Overall, IEDB data is stringently controlled and revised to guarantee the integrity and authenticity of the data. Members of the IEDB curation staff utilize a web-based curation system to input all data from published literature, which cannot be accessed by unauthorized or external users. Data from this system is exported weekly to a separate production site, which updates the IEDB so users can query the latest data. Upon releasing the new data, our developers follow a clear procedure of confirming the latest build has not altered or corrupted the data or digital objects. If a change to existing data is required, the curation system automatically tracks the date that reference is modified, the curator implementing the change, and the reason for the revision. Changes in entries are typically initiated when an author or user identifies a perceived discrepancy between a paper and the data in the IEDB. Such concerns can easily be raised by contacting help@iedb.org, which is noted throughout the IEDB website. In this case, a curator is assigned to review the paper and make any required revisions, which are promoted to the external database within 2 to 3 weeks. The user is notified of the resolution and the expected promotion date. Revisions are also made when changes arise in controlled vocabularies or ontologies, such as the NCBI Taxonomy, Disease Ontology (DO), or Ontology for Biomedical Investigations (OBI) assays, and this follows our normal weekly build and change control process.

The IEDB website is re-built weekly via scripts that convert the latest curation data into de-normalized query tables, CSV exports, and JSON files for the IEDB website. These scripts provide logging output, which is validated weekly to ensure there are no errors encountered in the data processing. Weekly table counts and output file sizes are tracked and compared week-to-week to ensure weekly data growth in line with expectations. The newly built IEDB website is tested to make sure all the default homepage queries and finders are functioning properly prior to a new build being released live. Each page of the website is tagged in the lower right-hand corner with a "Last Updated" date which tracks the specific build release. Additionally, the IEDB website codebase is maintained via a source control system and deployment packages which are thoroughly tested prior to production release to ensure accurate data presentation. Once the data is on the publicly accessible website, it is read-only and non-modifiable by users.

In regards to ontology updates, each release of IEDB data uses a specific, dated version of each supporting ontology. When a new version of a supporting ontology is released, we update IEDB data to use that new version, in a timely manner. For example, when the NCBI Taxonomy merges taxon X into taxon Y, we automatically update IEDB data using taxon X to use taxon Y instead. When NCBI Taxonomy deletes taxon Z, we manually review IEDB data using taxon Z and find a suitable replacement. When NCBI Taxonomy adds a new taxon W, we automatically add it as an option for curators to use when curating new data. This is done on an ongoing basis. In addition, we annually review the subset of the NCBI Taxonomy used in IEDB data,

looking for problems and outliers, and update data as required. Similar procedures are used for each supporting ontology.

These procedures also apply to data submitted by external researchers. Data that is submitted undergoes a comprehensive review by IEDB curation staff to ensure that it is in the proper format and complies with IEDB standards for data completeness. The data is submitted to machine-automated data validation checks, as well as the curator manual checks, before being approved for dissemination to the public. Submitters are granted access to the IEDB data submission system upon request. Submitters are also asked to state the type and amount of data they plan to submit so IEDB staff can verify that the data is within the IEDB's scope. Once approved, they can access the IEDB user center and a variety of CSV templates that can be used to submit their data. The templates allow the users to conveniently use spreadsheet programs, such as Excel, or a text editor, to input data in the required format. The user center contains documentation on the data fields. There is also a description of data fields available in the Curation Wiki URL. The curation manual, which documents the procedures used by the IEDB curators, is also available at the URL titled "Curation Manual 2.0). The curation manual is updated as needed when new data types or assays are encountered. This often occurs as technologies evolve, and as community standards are updated. One such example is our team's regular contact with the International Society for Biocuration, giving us insight into major updates in the field.

The repository does maintain provenance data and links to the metadata in 2 ways (i) the original data source is specified with PubMed identities and by specifying locations within the journal article where the data is derived from, and (ii) adding that metadata was authored by an IEDB curator as machine-readable data in JSON-LD format, following Google's structured data guidelines (using the provenance authoring and versioning ontology).

It is important to note that IEDB team members do not place their own quality assessment on data, or make any decisions to modify the data. It is inputted into the system exactly as the data is described in the peer-reviewed publication (after data synthesis). If any aspects of the data are unclear, the author or submitter may be contacted for clarification to ensure that the data is accurately represented before public dissemination.

Links:

- [Curation Wiki](#)
- [Curation Manual 2.0](#)

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text: Are released data retrospectively revised to address changes in the NCBI taxonomy, Disease Ontology (DO), or Ontology of Biological Investigation (OBI)?

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R8 Appraisal

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

The IEDB uses automated tools, followed by manual inspection by a senior immunologist, to ensure that published data is relevant to the IEDB. On a biweekly basis, the IEDB Document Specialist queries the PubMed database, and the output (list of PMIDs) is run through an automated document classifier. The classifier generates a binary curatability assessment, referred to as 'curatable' and 'uncuratable', and assigns broad subject matter categories and subcategories (i.e. Allergy; pollen or Infectious Disease; poxviruses). This automation ensures that the IEDB accepts data in a consistent and regulated manner. Finally, the Document Specialist and senior immunologist jointly review the abstracts of the curatable manuscripts for classification accuracy to confirm data relevancy.

All literature articles that are curated for the IEDB must contain the minimum list of required data, as stated in the Curation Manual. If data is missing, such as protein sequences, the corresponding author is contacted with a request to provide the missing data. If the author does not respond within 2 weeks, the article is either deemed uncuratable and set aside, or whatever data does meet the inclusion criteria is captured and published. The data undergoes an automated validation process that flags errors that are then fixed by the curator. Curated articles are then reviewed by another experienced curator as a quality control measure. Any discrepancies are discussed between the curators and are resolved according to the curation manual. Differences in interpretation of the data as stated in the article are elevated to the biweekly curation meeting, where the curation staff and PIs meet to discuss such issues and reach a consensus on how to proceed. If needed, the curation manual is revised accordingly.

With respect to data submitted to the IEDB, the data submission templates that depositors use indicate required data fields. As with the curated literature, automated validation is performed. All submissions are reviewed by the IEDB curators who work with the depositors to ensure that all data needed for inclusion in the IEDB are provided. Only at that point will a submission be ready for release in the database.

Lastly, we always aim to ensure the data is understandable for our users, by adhering to the FAIR data principles of findability, accessibility, interoperability, and reusability. We utilize community-derived ontologies (e.g., NCBI Taxonomy, Disease Ontology (DO), Ontology for Biological Investigation (OBI)) to ensure that users can find data based on the multiple synonyms used for immunological terms.

Additional information on the IEDB can be access via our Solutions Center (Help > Support) and some links are copied below to illustrate this:
- IEDB Curation Manual Article - https://help.iedb.org/hc/en-us/articles/114094147051-Curation-Manual
- IEDB Tutorials and Reference Materials - https://help.iedb.org/hc/en-us/sections/114094004331-Tutorials-and-Reference-Materials
- IEDB & FAIR Principles - https://help.iedb.org/hc/en-us/articles/360001387171-IEDB-FAIRness
- IEDB Data Submission Process:
- https://help.iedb.org/hc/en-us/articles/115001729911-Data-submission-process-overview
- https://help.iedb.org/hc/en-us/articles/114094146791-Data-submission-policy-for-the-general-public

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text: There is a wealth of information on the IEDB (http://wwwiedborg) website, but surprisingly little information describing the details of the curation process such as the details presented in items 7 and 8 Am I missing this information, or is it not provided in detail?

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R9 Documented storage procedures

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

The IEDB has all relevant processes and procedures documented online and they are managed by the team and Information Technology staff at LJI. As part of our start-of-contract deliverables, the IEDB team submitted 3 key documents related to data management and archival storage in January 2019 to NIAID; IT Backup Plan, Information System Security Plan and the Operational and Disaster Recovery Plan. These documents outline the process by which we manage data storage, and an abridged version is detailed below:
- External servers monitor all independent IEDB resources to verify that they are reachable every 120 seconds, 24 hours a day, 365 days a year.
- In the event of any failure of connectivity for primary IEDB resources at LJI, automated failover to a secondary site hosted at the San Diego Supercomputer Center (SDSC) on the University of California, San Diego (UCSD) campus is initiated. The SDSC IEDB infrastructure hosts a mirrored build of the IEDB, ensuring feature parity as well in the event of downtime. The build update process for the IEDB also takes advantage of this failover capacity. While upgrades are performed on the production site, the SDSC site will take over, resulting in no downtime for users.
- All IEDB virtual machines are snapshotted on a regular basis so that any individual VM failure can be quickly rolled back and recovered from.

In addition, LJI uses Confluence to document all internal processes, and the IEDB uses their own Confluence space to document system architecture and database design. Here we document how the internal curation system data is stored and protected:
- The Oracle 12c database used to store curation data is backed up daily using Oracle's export utility. Through a defined process, these backups can be used to restore the curation database, if needed, or used to refresh the test curation database to synchronize it with the production site.
A version of this documentation is available externally to users in our IEDB Solutions Center
https://help.iedb.org/hc/en-us/articles/114094150691-IEDB-System-Architecture-and-Design.

By way of curated data, the curation process is documented in the Curation Manual. Only the IEDB team has write access to the database, therefore this data is managed very closely and to a high degree of accuracy. All data in the IEDB either have been published or have been released by the data depositors for public dissemination.

The links to System Architecture and the Curation Manual have been checked and do resolve for further review. The Operational and Disaster Recovery Plan has now been made available in agreement with the NIAID Contracting Officer, Emily Dubbaneh Bannister, and our IEDB Program Officer, Joseph Breen, in our Solutions Center:
- Operational Recovery Plan (ORP) and Disaster Recovery Plan (DRP) -
https://help.iedb.org/hc/en-us/articles/4406591519515--IEDB-Operational-Recovery-Plan-ORP-and-Disaster-Recovery-Plan-DRP-
Due to security reasons, the IT Backup Plan and Information System Security Plan cannot be made available at this time.

Links:

- Curation Manual 2.0

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text: The links to the curation manual v2 and architectural manual are broken Searching your site I was able to find the earlier curation manual but I was unable to access it from your site (timeout) The link to the manual in Section 12 is live
The disaster recovery/security documents do not appear to be online so it is really not possible to evaluate any of the details of these plans

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

R10 Preservation plan

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

Compliance level:

The repository is in the implementation phase - 3

Response:

The level of responsibility for preservation is clearly outlined in the contract with NIAID, as are the future mitigation to address the threat of obsolescence. Since the repository has been funded by a government contract, the repository assumes responsibility for long-term preservation as long as the funding continues. NIAID 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 (or to the government), either way, as per the contract documentation. More specifically, the contractor shall coordinate with the incumbent contractor and NIAID to implement an orderly, secure and efficient transition of contract activities and contract-generated data, systems, analytical tools, and other documents and material during the three-month transition period. In preparation, a draft and final transition plan will be prepared to describe all transition activities, timelines, and assigned staff, which will then be approved by the Contracting Office, as described in the contract. In addition to this, in order to further mitigate the threat of obsolescence, the IEDB uses as much open-source software as possible, keeps systems up to date, and reviews the system architecture annually.

Finally, in regards to the contract between depositor and repository, all curated data is already in the public domain, so there are no contracts required between the depositor and respiratory. For submitted data, there is an implicit contract whereby the depositor must approve their submission and release their data through their personal account in the Data Submission Tool. The IEDB does have the right to copy, transform, and store the data, as well as provide access to them. We ensure that all IT plans and depositor data follow the correct protocol through strict governance and project management.

To further clarify, our approach to achieving long-term preservation in terms of format preservation, bit preservation, and storage migration is as follows:
1) Format preservation: The formats we have chosen (SQL dumps generated by MySQL and PostGres) are expected to remain readable for the foreseeable future, as SQL is one of the most fundamental IT formats. Moreover, the SQL files are in text format that is fully human-readable.
2) Bit preservation & storage migration: Bit preservation on our primary storage servers is ensured at multiple levels including the use of SSDs as opposed to spinning media in our Dell/EMC PowerVault ME4084, using a filesystem (ZFS) that checksums and corrects each read, and the redundancy of data in the RAID. Daily backups using Veeam, which has its own internal checksum methods, extends the bit preservation protection to our backups. We are currently in the process of migrating to a next-generation storage system using a Ceph/BlueStore object backend that similarly checksums and corrects each file access. Migrating data from our current system to the next will be achieved using rsync, which will handle checksum comparisons and ensure the faithful transfer of data between systems.

Links:

Reviews

Reviewer 1:

Compliance level:

The repository is in the implementation phase - 3

Comments:

Reviewer 2:

Compliance level:

The repository is in the implementation phase - 3

Comments:

Accepted

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text: The link to the architecture document is broken

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

R11 Data quality

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:

The guideline has been fully implemented in the repository - 4

Response:

The IEDB houses data curated from published literature, which has already been through a peer review process in order to be accepted to a scientific journal. The content of the articles must contain, at a minimum, the required data fields that are stated in the Curation Manual in order to be captured in the IEDB. To ensure high-quality curation, automated data validations are run in the curation system, and the curations themselves are peer-reviewed by another curator, prior to publishing the data. Pertinent information that must be included in the curation is the PubMed ID of the article, the authors and their affiliations, article title, and year of publication. In essence, the IEDB is a database of epitopes and the assays in which they were tested, and all relevant details about each assay are included. In this way, users can make their own judgments about the data they want to incorporate in their own analyses and studies. Users are also directed to the original paper (in the 'References' tab) for further quality-related assessments. For example, they might want to use only data generated from a particular type of assay, laboratory, and date range.

There is no method in place for users to comment on or rate data or metadata, and there are no plans to implement such a feature into the IEDB. From the outset, it has been the policy of the IEDB to curate all literature and direct data submissions that meet our published standards for inclusion, as described previously. The IEDB does not set out to make quality assessments on the published data; rather its goal is to simply aggregate this information and make it publicly available to the scientific community, so researchers can use their own critical judgment. However, by default, the IEDB data is sorted by the number of references, whereby the epitope with the highest number of references is shown first. This does not provide users a quality indicator, but it does easily highlight which epitopes have been curated most (and hence, which ones have the most data).

Immune Epitope Database and Analysis Resource

In an effort to maintain utmost transparency, the IEDB publishes multiple papers throughout the year outlining major changes to the database or process updates in capturing information. This serves as a great resource for users wanting to know how data is captured, aiding them in their own quality assessment. In addition, we maintain the IEDB Solutions Center (https://help.iedb.org/hc/en-us), which houses help information, post video tutorials on our YouTube channel and provide help guides for each tool in the Analysis Resource, which link back to the associated publication for more information. This provides users with sufficient information to make their own quality assessments. In the event an error is found, users are directed to help@iedb.org where the issue is investigated by a team member and rectified. If it is related to curation, the data entry is temporarily removed from the database and recurated, prior to re-publishing.

Links:

* [IEDB YouTube Channel](#)

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Explicitly no value judgment on quality - but this is fine, the systemic checks are comprehensive.

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

R12 Workflows

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

Given the IEDB has been in operation since 2003, it has well-established and documented procedures for curating data from the literature and processing data submitted from external researchers. Any revisions to data records are tracked by the online curation system that is accessed only by the IEDB team members involved with curation. Additionally, external data submissions must also conform to the overall mission of the IEDB. For example, at present, the IEDB would not accept HIV data submissions because such data is outside the current scope of the IEDB. The IEDB curation team works closely with data depositors to ensure that data and metadata are complete and in the correct format for submission.

Data is promoted to the IEDB website on a weekly basis via a well-established and documented process. Prior to each release, checks are performed to ensure that all data has been correctly transferred from the internal curation system, running Oracle, to the user-facing IEDB website. In addition to this, any updates made to the Analysis Resource are documented in the IEDB-AR Release Notes, with a description of changes for the user's awareness. Data submissions are not made public until the depositor initiates their release. As a result, there are both qualitative and quantitative checks of outputs; system checks to ensure that servers are running properly and data is being transferred accurately, and human review to ensure data is correct. Currently, we do not have the procedural workflows of the weekly build process and checks documented in our Solutions Center. This process is clearly articulated in our internal Confluence site for IEDB team members, however, we will investigate creating an external resource for users in the next update of our help material.

At certification renewal, we will be sure to share more information publicly on our procedural workflows.

Links:

- [Curation Manual 2.0](#)
- [External Data Submission Guidelines](#)

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text: The curation manuals are quite clear and rich in scientific detail
I could not find any documentation of the the procedural workflows of the weekly update process and associated checking processes online,

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R13 Data discovery and identification

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

The IEDB strives to make all data discoverable by users with stable, static, and persistent identifiers. By presenting curated data in a searchable database, we have liberated it from the tables and figures of journal articles, making it more accessible and usable by immunologists. From the IEDB

website home page, users can perform 98% of queries, which can be further refined from the 'Results' page using the 'Filter Options'. The home page web interface has been optimized over many iterations to be intuitive to users, ensuring that the data is easily queried and discoverable.

The IEDB assigns unique identifiers, with the most fundamental being the identifier for a specific assay. For example, URL http://www.iedb.org/assay/1288921 identifies the IEDB record for an experimental assay. In addition, a collection of assays curated from a single reference has a separate identifier. For example, http://www.iedb.org/reference/1001817 identifies the set of 34 assays curated from the journal article that included the experiment. By utilizing the full uniform resource locator (URL) as the identifier, we ensure that the identifier is globally unique and that someone with an identifier can find more information about the resource using a web browser or other common tools. CURIE syntax can be used to define a mapping from 'IEDB_ref' to the URL http://www/iedb.org/reference/, allowing two-way translation between the CURIE 'IEDB_ref: 1001817' and the full URL http://www.iedb.org/reference/1001817. This gives us all the benefits of a compact ID and a findable URL. In terms of persistence, the IEDB is committed to retaining these identifiers, and if there are changes to the identifier scheme, utilizing HTTP redirects to ensure that the URLs will continue to resolve. While there are dependencies on the continued control of the iedb.org domain name for long-term persistence, we believe that scenarios in which the IEDB continues to be available, but the domain does not are highly unlikely. Additionally, these IDs are not dependent on the specific software used as they are in both a URL and CURIE format. Thus, we believe that the IEDB data is persistent. In the event that a reference is removed (due to error or otherwise), a message is added to the database advising users why this reference has been temporarily or permanently removed.

In regards to using DOIs or universal PIDs/PURLs as epitope identifiers, this has been discussed extensively within the group. Whilst DOIs and PURLs may provide more flexibility, it does come at the cost of another layer of complexity and delegating responsibility. We believe that the IEDB URLs are simple and effective, and they were designed to be resource agnostic and easy to redirect. Furthermore, in the event that the IEDB contract must be transferred, the iedb.org domain will be handed over (along with the IEDB name and other "trademarks"). We can also establish a URL redirect of the existing identifiers as requested by the incumbent. We feel confident that all IEDB data can be transferred to the incumbent without compromising the existing information or identifiers. This will also be documented in the 'Final Transition Plan', which is outlined in the Statement of Work provided by NIAID, whereby the contractor (IEDB) shall "ensure an orderly, secure, and efficient transition of contract-related materials and activities to the successor contractor or to the Government". The IEDB will prepare a draft and final plan, to be reviewed by NIAID, no later than 12 months prior to the completion date of the contract. The plan will detail the transition activities to be carried out, provide a timeline for the implementation of each transition activity, and describe the capabilities and responsibilities of Contractor staff who shall be assigned to implement the plan. As a result, we are confident that the IEDB URLs are not only effective for current use but can be easily transferred in the future if needed.

In addition to this, the IEDB has worked tirelessly to ensure that we are FAIR-compliant, adhering to the principles set by Wilkinson et al. (2016). Below, we describe how the IEDB meets these community guidelines:

1. Findable: The IEDB uses unique identifiers, as explained above, which are included in the full uniform resource locator (URL), so it can be easily searched for in the web browser (e.g. http://www.iedb.org/reference/1001817; http://www.iedb.org/assay/1288921). This gives us all the benefits of a compact ID and a findable URL. In terms of persistence, the IEDB is committed to retaining these identifiers, and if there are changes to the identifier scheme, we utilize HTTP redirects to ensure that the URLs will continue to resolve. The IEDB also identifies the data being described by linking to the relevant journal publication in terms of citation information (journal, author, title, year, volume, pages, etc.) and, more importantly for machine-readable linkage, by the PubMed ID.

2. Accessible: The main protocol used to obtain data from the IEDB is simply through HTTP, therefore no authentication is required to access IEDB data, making the data very accessible. If a published article is retracted (or redacted), the metadata captured in the IEDB would remain available, adhering to this principle.

3. Interoperable: Whenever possible, the IEDB utilizes externally developed vocabularies to describe a given domain, primarily through the use of Open Biomedical Ontologies (OBO) Foundry ontologies. The principles and practices of the OBO Foundry ensure that member ontologies are findable through the OBO registry, accessible through standardized interfaces, interoperable through the common use of the OWL standard and reproducible through the persistent availability of versioned copies of ontologies over time. If an existing OBO ontology in a domain does not provide a term needed by the IEDB, we submit new term requests.

4. Reusable: IEDB data is richly described with a plurality, per the guidelines, with the metadata describing individual experiments of up to 400 attributes. IEDB data is also associated with detailed provenance in 2 ways; (i) the original data source is specified with PubMed identities and by specifying locations within the journal article where the data is derived from, and (ii) adding that metadata was authored by an IEDB curator as machine-readable data in JSON-LD format, following Google's structured data guidelines (using the provenance authoring and versioning ontology).

In addition to this, the IEDB has been working to improve data harvesting of the metadata. Users can access data in the IEDB by running queries on the website and exporting a set of query results in spreadsheet format, or by downloading the entirety of the IEDB either in XML format or as a SQL database. We have also added machine-readable metadata to our IEDB web pages, beginning with provenance data. The metadata is encoded in JSON-LD format, following Google's structured data recommendations, and can be easily translated into other concrete Resource Description Framework (RDF) formats.

Overall, the IEDB certainly does enable users to discover the data and refer to it in a persistent way through proper citation.

Links:

- [Open Biomedical Ontologies (OBO) Foundry ontologies](#)

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text: Has there been any consideration of assigning DOIs to IEDB data sets This may provide a more enduring identifier in the event that your data must be served under another URL at some future point

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R14 Data reuse

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:

The guideline has been fully implemented in the repository - 4

Response:

The domain of the IEDB is immunological investigations of epitope reactivity, for which there are no formal standards established. As a result, the design, implementation, and curation guidelines of the IEDB have been, instead, vetted by the scientific immunology community through interactions with domain experts, publications, and outreach activities, including conference booths, annual workshops, and user surveys. As described in R6, there are formal opportunities to discuss advancements in this space, such as the annual NIAID large-scale epitope discovery contracts meeting, and informal avenues, such as through our IEDB Expert Committee and user feedback. Ultimately, the IEDB adopts community standards via its direct work and feedback from the community and continues to improve understandability through these avenues.

In addition to this, the IEDB provides as much metadata (when the data is accessed) as can be extracted from the publication. We provide up to 400 attributes in an easy-to-understand table format on the web, with some of the most important information relating to provenance, so that users can locate the original article for further analysis and reuse. The data is provided in formats used by the community; firstly, an easily searchable web interface that has been optimized for users, and secondly tabulated results, which can be exported in a spreadsheet format, useful for both researchers and bioinformaticians.

Elements of the IEDB are expressed using external metadata standards, such as the use of the Open Biomedical Ontologies (OBO) Foundry ontologies more generally, and NCBI taxonomy to describe organism species, more specifically. It is imperative that the IEDB upholds these already established ontologies and taxonomies, to make it easier for users to understand and reuse the data across a broad range of areas. In regards to ontology updates, each release of IEDB data uses a specific, dated version of each supporting ontology. When a new version of a supporting ontology is released, we update IEDB data to use that new version, in a timely manner. For example, when the NCBI Taxonomy merges taxon X into taxon Y, we automatically update IEDB data using taxon X to use taxon Y instead. When NCBI Taxonomy deletes taxon Z, we manually review IEDB data using taxon Z and find a suitable replacement. When NCBI Taxonomy adds a new taxon W, we automatically add it as an option for curators to use when curating new data. This is done on an ongoing basis. In addition, we annually review the subset of the NCBI Taxonomy used in IEDB data, looking for problems and outliers, and update data as required. Similar procedures are used for each supporting ontology.

These updates follow a very structured weekly process. The IEDB website is re-built weekly via scripts that convert the latest curation data into de-normalized query tables, CSV exports, and JSON files for the IEDB website. These scripts provide logging output, which is validated weekly to ensure there are no errors encountered in the data processing. Weekly table counts and output file sizes are tracked and compared week-to-week to ensure weekly data growth in line with expectations. The newly built IEDB website is tested to make sure all the default homepage queries and finders are functioning properly prior to a new build being released live.

We are cognizant that technologies and data are constantly evolving, therefore continuous feedback and outreach efforts are imperative to our success and adoption of these new advancements. We are proactive in attending annual conferences, hosting booths, and running the IEDB user workshop, which facilitates this feedback loop. Dr. Bjoern Peters is a leader in developing and maintaining ontologies, therefore we have unique insight into developments in this space. We also have regular interactions with team members of other databases, such as UniProt, enabling us to adapt when new changes are being introduced. Ultimately, the IEDB is constantly evolving to ensure that we keep up with changing technologies and standards, and by extension, ensure that our data is reusable long into the future.

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text: Given the extent of the external references in the fully curated IEDB data,
how do you manage changes in the external data resources in released datasets Is there some regular update procedure

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Technology

R15 Technical infrastructure

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:

The guideline has been fully implemented in the repository - 4

Response:

Immune Epitope Database and Analysis Resource

Since its inception in 2003, the IEDB has functioned on well-supported operating systems and other core infrastructure software supported by the LJI Information Technology department. Below is a summary of hardware and software technologies currently used, and how they are appropriate for use by our community.

The IEDB has three basic systems that reside in different locations.

1. Curation - The curation system uses an Oracle Database 12C, only accessible by LJI employees for data curation. We have 2 database servers that run Oracle (production and development) and two application servers (production and development).

2. External Database - We have 5 database servers running MySQL (two local production, two remote production, and one local development) and 5 web servers (local production, remote production, and development).

3. Tools - We have 3 application servers for the tools (local production, remote production, and development) and 3 corresponding compute clusters for data processing.

All servers are virtual machines running in a high-availability VMware vSphere cluster with 5 physical hosts and utilize a redundant Storage Area Network (SAN). All local production virtual machines are replicated offsite at the San Diego Supercomputer Center (SDSC) in La Jolla, CA, (excluding the curation servers). The virtual servers in the production set are duplicated, whereby one is used for staging the weekly update while the other actually serves as the production machine visible to the public. The staging and production environments are swapped weekly at the end of the update on the staging machine to ensure there is no downtime for users.

It is clear that the IEDB has reliable infrastructure to ensure that users can utilize the repository most effectively. We maintain community standards for technical implementation and adopt new technologies and methods as they arise (and are proven to be effective).

LJI uses Confluence to document all internal processes, and the IEDB uses its own Confluence space to document system architecture, database design and hardware/software. This documentation is reviewed on an annual basis by key team members. A version of this documentation is available externally to users in our IEDB Solutions Center https://help.iedb.org/hc/en-us/articles/114094150691-IEDB-System-Architecture-and-Design. We also have automated monthly software updates to ensure the latest software is in use, and hardware is managed centrally by the LJI IT department.

Availability, bandwidth, and connectivity are, indeed, sufficient to meet the needs of the community, and this is continually monitored by Zabbix. This is to ensure that 95% of database queries display results in 5 seconds or less, and so that the tools do not drop in speed of 2x or more. We have also added automated monitoring for usage spikes that may be indicative of misuse of the database or tools.

As described in R9, the IEDB submitted to NIAID a series of documents relating to business continuity and disaster planning in January 2019 (namely IT Backup Plan, Information System Security Plan, and the Operational and Disaster Recovery Plan). They describe that we receive contact from our monitoring servers every 120 seconds, 24 hours a day, 365 days a year. The majority of operational outages can be mitigated by redundancies in place at LJI. Failure of our primary internet circuit will result in immediate failover to a secondary circuit. Power failures are mitigated by uninterruptible power supplies and LJI's diesel power generator. IEDB servers are hosted in a VMWare environment on a 5-node cluster, ensuring that server hardware failure is highly unlikely to affect the IEDB's service, and computing resources are allocated dynamically to ensure that the IEDB always has the resources necessary to perform optimally.

In the event of a failure that renders an IEDB-critical resource unreachable, however, the monitoring service will automatically alert the Systems Administrator and Senior Director of the connection failure (the system is capable of e-mail, SMS, and Push notifications) and initiate an immediate DNS failover to the IEDB's replication site located at the San Diego Supercomputer Center (SDSC) on the University of California, San Diego (UCSD) campus. The replication site represents an exact mirror of production IEDB servers at the time of failure. Once failover has been verified as functional, the Systems Administrator will troubleshoot the server/application at the primary site, repair the issue, and restore the primary site service. The issue will be documented and reported to relevant IEDB personnel.

Infrastructure components are evaluated as an annual activity by the IT infrastructure team. The system's architecture is compared against future usage projections as well as planned new features to determine if changes to the underlying components are necessary to continue to serve the needs of our users. In this evaluation, solutions are drawn from both existing and new technologies.

Overall, the IEDB has excellent IEDB infrastructure, stringent procedures, and system monitoring to provide users with continued and stable access to the repository. This is reflected in consistently exceeding a self-imposed SLA of 99.9% since 2016.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text:

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R16 Security

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

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

The IEDB's IT Infrastructure team is dedicated to maintaining the integrity of the repository and ensuring no security breaches occur. We have set up firewalls to protect the aforementioned servers from unauthorized access, both at LJI and SDSC. We conduct system scans every 6 months to test security and patch any high-risk vulnerabilities. In 2020, we also added automated triggers to Google Analytics to monitor for usage spikes and traffic deviations. This acts as an additional layer of security, as we review traffic to the database and tools on a quarterly basis. Where usage exceeds expectations without clear cause, it may indicate a potential attack by users and will require further investigation.

As part of our start-of-contract deliverables, the IEDB team submitted an Information System Security Plan in January 2019 to NIAID. This document outlines 5 key focus areas; detect, identify, protect, recover and respond:

1. Detect: Data travelling through the LJI network is protected by our Fortigate 3600C firewall. All external traffic is denied by default, and all connections are logged and monitored for abnormalities. Security software (Javelin) is used to detect malicious behavior internal to the network and Cylance antivirus is installed on each workstation. This ensures that malware and other security threats are identified and quarantined immediately. Zabbix software is hosted both internally and externally and monitors nearly all of LJI's IT infrastructure.

2. Identify: All server and end-user hardware at LJI is tagged and inventoried, with major assets receiving an annual audit of their status and location. An annual cybersecurity assessment is performed to determine vulnerable areas and to measure cybersecurity program progress.

3. Protect: All LJI and IEDB resources are protected behind LJI's next-gen firewall, the Fortigate 3600C. Access to all resources is granted on the principle of least privilege and access from outside of the building requires VPN access granted to named user accounts (no generic logins are distributed for the VPN). All workstations are provided with antivirus, and Windows/Linux servers are patched regularly to ensure that they are not vulnerable to exploit.

4. Recover: In the event of any failure of connectivity for primary IEDB resources at LJI, automated failover to a secondary site hosted at the San Diego Supercomputer Center (SDSC) on the University of California, San Diego (UCSD) campus is initiated.

5. Respond: In the event of outages and issues with the IEDB, all core IEDB personnel are located at LJI's primary site, making communication and collaboration simple. The support site for the IEDB is an externally hosted application via Zendesk, ensuring that updates to IEDB users can be made independent of LJI and SDSC's site status.

Additionally, we have established IEDB Tools General Usage Guidelines, which are freely available online. If users do not abide by these guidelines, our team can block their access in order to maintain IEDB integrity. It is documented that our system security will be based on permissions granted to authenticated users, hence this is managed closely to ensure only those that require access are granted it.

Since the repository is a federal information system, it must follow IT system protection measures that meet Department of Health and Human Services (DHHS) requirements by complying with the HHS Automated Information Systems Security Program Handbook (http://ftp.fas.org/sgp/othergov/hhs-infosec.pdf).

In addition to the above information, the IEDB has documented approaches to resource preservation. The IEDB Disaster Recovery Plan can now be found in our Solution Center, along with the Operational Recovery Plan-
https://help.iedb.org/hc/en-us/articles/4406591519515--IEDB-Operational-Recovery-Plan-ORP-and-Disaster-Recovery-Plan-DRP-. In addition to this, we also have an IT Backup Plan and IT Information Security Plan, which were submitted to NIAID upon the start of the new contract cycle. Unfortunately, these plans cannot be publicly shared, but below is a brief overview of the preservation planning processes detailed within those documents.

The IT environment on which the IEDB runs has been built to standards in line with a high-performance bioinformatics or commercial informatics organization. IEDB servers and backup materials are located in a central data closet, connected to every IEDB workstation by at least 1 gigabit Ethernet, that uses 100% conditioned and emergency-backup power as well as electronic locks that log and restrict all but a few IT staff from entering. All of the IEDB suites and offices have extremely high densities of computing equipment and network infrastructure.

The Reduxio storage system of which all IEDB VMs reside takes snapshots continuously at the following intervals and with the indicated retention of each:
- Once per hour and retained for 72 hours
- Once per day and retained for 1 week
- Once per month and retained for as long as free space allows

Veeam backs up all IEDB virtual machines on a daily basis starting at 7pm. These daily backups are retained for a period of two weeks within Veeam itself. The backup files are stored on a dedicated ZFS volume that has its own snapshot policy that occurs hourly and is retained for 24 hours, daily and retained for 31 days, weekly and retained for 8 weeks, and monthly and retained as many as free space allows. The ZFS backup repository for Veeam is replicated to an offsite storage facility at the San Diego Supercomputer Center (SDSC) in La Jolla, CA, (excluding the curation servers). The virtual servers in the production set are duplicated, whereby one is used for staging the weekly update while the other actually serves as the production machine visible to the public. The staging and production environments are swapped weekly at the end of the update on the staging machine to ensure there is no downtime for users. Due to these daily snapshots and offsite backups, bit rot protection is being addressed by the IEDB. Format preservation is also being addressed by the fact that the IEDB stores the same dataset in a variety of open-source formats, including MySQL, PostGres, XML, all of which are constantly snapshotted and backed up.

Links:

- IEDB Tools General Usage Guidelines

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Response Text:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Applicant Feedback

R17 Applicant Feedback

We welcome feedback on the CoreTrustSeal Requirements and the Certification procedure.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

We thank the reviewers and Board for their thoughtful feedback and support!

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

This is a strong application from a mature repository. For overall Compliance Level 4 (R10) I recommend a little more public evidence online as well as links added to the evidence.