VEuPathDB
Eukaryotic Pathogen, Vector & Host Informatics Resources

VEuPathDB BRC contract HHSN75N93019C00077

Usage Metrics Report

Reporting Period: July 1-31, 2021
Submission Date: August 10, 2021

Notes & Change Log

Date	Version/release	Description & Notes
8/10/2021	1	VEuPathDB Performance Metrics for July 2021 ● Response to COR feedback - 'Total registered users' has been moved from Table 6 to Table 1. ● Response to COR feedback - The 'Completed' column in Table 4 has been removed based on feedback that there was overlap between Table 4 of this report and Table 2 of the performance metrics report. Table 4 now more accurately represents usage (jobs submitted) while performance (jobs completed) is represented in the performance metrics report. ● In response to NIAID's request, we are working with our sister BRC to provide jointly agreed plots showing accumulative metrics over time. These will be included in the next reporting period.

Joint-BRC Common Usage Metrics Plan

This report will be made available from all VEuPathDB sites, *e.g.*, https://veupathdb.org/, from the About menu.

This monthly usage metrics report provides a summary of the VEuPathDB BRC usage for the current reporting period in accordance with the Joint-BRC Common Usage Metrics Plan developed by the BRCs and subsequently approved by NIAID.

As per the plan, each BRC will aggregate metrics for their constituent parts, *i.e.* FungiDB, PlasmoDB, OrthoMCL-DB, VectorBase, *etc.* for VEuPathDB. These metrics will serve as a basis for collecting quantitative measures of usage of the BRC resources to identify trends, areas that are performing well, and areas for improvement. Usage metrics will be reported to NIAID individually by each BRC on a monthly basis, and in combination on the BRC Gateway website once this is publicly available. Annual summaries will be included in the Annual Progress Reports.

It is important to note that metrics across the two BRCs are highly dependent on the relative sizes of the respective research communities, the associated quantities and types of available public data, and how each of the resources delivers the data and tools to the user. Thus, cross-BRC comparisons of individual metrics are not necessarily indicative of relative usage or performance.

Common usage metrics covering both BRCs (note that this list is subject to modification, based on feasibility of collection, changes in availability technologies, BRC website development, suggestions from NIAID program and other stakeholders, *etc.*):

Website Usage Metrics

Website usage is a key measure for evaluating use of the resource by the research communities. The number of website sessions unique users in a given period provide insights into trends, such as increased traffic resulting from outreach activities and prominent research topics and endeavors. Both the BRCs will use AWStats to monitor and track website usage by and report the number of unique visitors, visits, page views, pages/visit and visits/visitors for a given reporting period, aggregated across all constituent BRC websites, as summarized in the table below. For VEuPathDB, live website usage statistics pages generated by AWStats from individual websites can be accessed at https://veupathdb.org/awstats/awstats.pl, https://plasmodb.org/awstats/awstats.pl, etc. by replacing individual site names in the URL. These links provide more detailed usage statistics by day of the week/month, country, browser / operating system, and more.

- **Total registered users**
  - *Definition - Total cumulative number of users who have registered with the BRC via the website registration mechanism, from inception to the specified date.*
  - *Measurement mechanism - The registration process creates an entry in the registered user database for each BRC. Total number of registered users is queried from the database at the specified date.*
  - *Measure - Total number of registered users (cumulative).*
- **Total visits**
  - *Definition* - Number of visits made by all visitors. Think "session" here, say a unique IP accesses a page, and then requests three other pages within an hour. All of the "pages" are included in the visit; therefore, you should expect multiple pages per visit and multiple visits per unique visitor (assuming that some of the unique IPs are logged with more than an hour between requests).
  - *Measurement mechanism* - AWStats.
  - *Measure* - Total number of visits per month.

- **Total unique visitors**
  - *Definition* - A unique visitor is a person or computer (host) that has made at least 1 hit on 1 page of your web site during the current period shown by the report. If this user makes several visits during this period, it is counted only once. Visitors are tracked by IP address, so if multiple users are accessing your site from the same IP (such as a home or office network), they will be counted as a single unique visitor
  - *Measurement mechanism* - AWStats.
  - *Measure* - Total number of unique visitors per month.
- **Total page views**
  - *Definition* - The number of "pages" viewed by visitors. Pages are usually HTML, PHP or ASP files, not images or other files requested as a result of loading a "Page" (like js, css... files).
  - *Measurement mechanism* - AWStats.
  - *Measure* - Total pageviews per month.
- **Average pages per visit**
  - *Definition* - The average number of pages viewed during a visit. Repeated views of a single page are counted.
  - *Measurement mechanism* - AWStats.
  - *Measure* - Average number of pages per visit per month.
- **Average visits per visitor**
  - *Definition* - The average number of visits per visitor.
  - *Measurement mechanism* - AWStats.
  - *Measure* - Average number of visits per visitor per month.
- **Average visit duration**
  - *Definition* - The average time a visitor spent on the site for each visit, measured in seconds.
  - *Measurement mechanism* - AWStats.
  - *Measure* - Average visit duration per month.
- **Total bandwidth**
  - *Definition* - Total number of bytes for pages, images and files downloaded by web browsing. This number includes traffic for web only (or mail only, or ftp only depending on value of LogType). This number does not include technical header data size used inside the HTTP or HTTPS protocol or by protocols at a lower level (TCP, IP...). Note that this number is often lower than the bandwidth usually reported by internet providers as it is counted at a lower level and includes all IP and UDP traffic.
  - *Measurement mechanism* - AWStats.
  - *Measure* - Total bandwidth per month.

Table 1 VEuPathDB Website Usage Metrics (July 1-31, 2021)

Metric	Result
Total registered users	23,450
Total visits	94,385
Total unique visitors	34,972
Total pageviews	10,401,162
Avg. pages / visit	110.19
Avg. visits / visitor	2.69
Avg. visit duration (seconds)	488

Bandwidth (GB)	509.08

Website Usage by Taxa

BRCs support a variety of organism taxa containing human pathogens and their vectors, along with related genomic and other omics data types. These taxa vary widely in the number of species and genomes they contain, availability of omics data, as well as the size of the research communities studying them. Measuring the BRC website usage by taxa allows us to understand how BRC resources are used by various organism communities. We will report the number of website page views by taxa, which will be measured by querying the website usage statistics in Google Analytics by taxa name.

Table 2 VEuPathDB Website Usage by Taxa (July 1-31, 2021)

Taxa	Domain	Page Views	# of Species	# of Genome Seqs
Plasmodium	Protozoa	114908	22	51
Trypanosoma	Protozoa	79596	8	25
Pyricularia	Fungi	61214	1	2
Toxoplasma	Protozoa	41023	1	15
Cryptococcus	Fungi	20472	5	10
Saccharomyces	Fungi	15181	1	1
Leishmania	Protozoa	14713	15	22
Aedes	Vectors	12062	2	3
Anopheles	Vectors	7757	19	24
Aspergillus	Fungi	5637	23	28
Fusarium	Fungi	4613	7	13
Cryptosporidium	Protozoa	4471	7	11
Neurospora	Fungi	3843	3	3
Entamoeba	Protozoa	3381	5	9
Giardia	Protozoa	2165	4	6
Culex	Vectors	1835	1	1
Babesia	Protozoa	1795	6	6
Rhodnius	Vectors	1783	1	1
Phytophthora	Fungi	1691	7	7
Trichomonas	Protozoa	1592	1	1
Ixodes	Vectors	1404	1	2

Eimeria	Protozoa	1361	8	8
Neospora	Protozoa	1148	1	1
Glossina	Vectors	1139	6	7
Lomentospora	Fungi	1040	1	1
Naegleria	Protozoa	1005	2	3
Drosophila	Vectors	966	1	1
Candida	Fungi	921	8	15
Biomphalaria	Vectors	854	1	1
Angomonas	Protozoa	741	1	1
Encephalitozoon	Protozoa	655	4	9
Acanthamoeba	Protozoa	641	1	1
Coccidioides	Fungi	606	2	6
Homo	Host	499	1	1
Theileria	Protozoa	496	4	4
Sarcocystis	Protozoa	432	1	2
Crithidia	Protozoa	430	1	1
Chromera	Protozoa	416	1	1
Ustilago	Fungi	393	1	1
Paratrypanosoma	Protozoa	362	1	1
Schizosaccharomyces	Fungi	339	3	3
Bodo	Protozoa	333	1	1
Lutzomyia	Vectors	330	1	1
Leptomonas	Protozoa	318	2	2
Histoplasma	Fungi	294	1	5
Endotrypanum	Protozoa	291	1	1
Blechomonas	Protozoa	291	1	1
Hepatocystis	Protozoa	265	1	1
Zymoseptoria	Fungi	259	1	2
Besnoitia	Protozoa	251	1	1
Mucor	Fungi	248	2	2
Nosema	Protozoa	232	2	3
Clavispora	Fungi	217	1	1

Botrytis	Fungi	206	1	1
Cyclospora	Protozoa	200	1	2
Phlebotomus	Vectors	189	1	1
Trichoderma	Fungi	181	2	2
Cystoisospora	Protozoa	157	1	1
Paracoccidioides	Fungi	139	2	3
Coprinopsis	Fungi	116	1	1
Cimex	Vectors	112	1	1
Penicillium	Fungi	110	1	1
Hammondia	Protozoa	105	1	1
Rhizopus	Fungi	105	1	1
Thermothelomyces	Fungi	103	1	1
Nematocida	Protozoa	95	3	5
Scedosporium	Fungi	94	1	1
Kwoniella	Fungi	93	4	4
Puccinia	Fungi	89	4	5
Vitrella	Protozoa	85	1	1
Cladophialophora	Fungi	85	2	2
Phanerochaete	Fungi	85	1	1
Talaromyces	Fungi	81	2	2
Gregarina	Protozoa	80	1	1
Sordaria	Fungi	77	1	1
Mus	Host	77	1	1
Anncaliia	Protozoa	77	1	2
Fonsecaea	Fungi	76	1	1
Malassezia	Fungi	75	3	4
Pichia	Fungi	70	1	1
Colletotrichum	Fungi	69	1	1
Sporothrix	Fungi	69	2	2
Musca	Vectors	68	1	1
Pleurotus	Fungi	66	1	1
Blastomyces	Fungi	64	3	4

Sclerotinia	Fungi	63	1	1
Podospora	Fungi	62	1	1
Spizellomyces	Fungi	60	1	1
Exophiala	Fungi	60	3	3
Yarrowia	Fungi	55	1	2
Pediculus	Vectors	55	1	1
Uncinocarpus	Fungi	53	1	1
Enterocytozoon	Protozoa	53	2	2
Trichosporon	Fungi	52	1	1
Phycomyces	Fungi	52	1	1
Sporisorium	Fungi	51	1	1
Melampsora	Fungi	49	1	1
Tremella	Fungi	47	1	1
Penicilliopsis	Fungi	46	1	1
Stomoxys	Vectors	45	1	1
Enterospora	Protozoa	44	1	1
Cenococcum	Fungi	44	1	1
Allomyces	Fungi	43	1	1
Mitosporidium	Protozoa	39	1	1
Ophiostoma	Fungi	39	1	1
Edhazardia	Protozoa	37	1	1
Cyphellophora	Fungi	36	1	1
Rhizophagus	Fungi	34	1	2
Leptotrombidium	Vectors	34	1	1
Hepatospora	Protozoa	34	1	2
Bos	Host	34	1	1
Sarcoptes	Vectors	34	1	1
Verruconis	Fungi	33	1	1
Hanseniaspora	Fungi	33	2	2
Monocercomonoides	Protozoa	30	1	1
Culicoides	Vectors	30	1	1
Hyaloperonospora	Fungi	29	1	1

Batrachochytrium	Fungi	28	1	1
Pythium	Fungi	25	2	2
Globisporangium	Fungi	23	3	4
Claviceps	Fungi	23	1	1
Cytauxzoon	Protozoa	23	1	1
Vittaforma	Protozoa	22	1	1
Spraguea	Protozoa	22	1	1
Trachipleistophora	Protozoa	16	1	1
Spironucleus	Protozoa	15	1	1
Albugo	Fungi	15	2	2
Kluyveromyces	Fungi	14	1	1
Macaca	Host	14	2	2
Ascosphaera	Fungi	14	1	1
Pseudoloma	Protozoa	13	1	1
Pseudogymnoascus	Fungi	13	1	1
Aphanomyces	Fungi	12	2	2
Bremia	Fungi	11	1	1
Phytopythium	Fungi	10	1	1
Vavraia	Protozoa	9	1	1
Saprolegnia	Fungi	9	2	2
Pneumocystis	Fungi	8	1	1
Ordospora	Protozoa	4	1	1
Blumeria	Fungi	3	1	1
Peronospora	Fungi	2	1	1
Amphiamblys	Protozoa	1	1	1
Rhinocladiella	Fungi	1	1	1

Website Usage by Data Types

BRCs support genomic and a variety of other omics data types, providing an integrated view of these multi-omics data and related analysis tools. Tracking the website usage by primary data types allows us to understand how these data types are used. We will report the number of website pageviews by primary data types, which will be measured by querying the website usage statistics in Google Analytics by data type.

Table 3 VEuPathDB Website Usage by Data Type July 1-31, 2021)

Data Type	Domain	Page Views	Searches
Taxonomy	VEuPathDB	404090	451
Genomes	VEuPathDB	404090	1350
Genome sequences	VEuPathDB	404090	3474
Genes/Proteins	VEuPathDB	404090	55674
Transcriptomics	VEuPathDB	355479	34096
Proteomics	VEuPathDB	262093	1748
Variation data	VEuPathDB	365480	5325
Epigenomics	VEuPathDB	147645	4
Enzyme commission	VEuPathDB	147284	30
Gene Ontology	VEuPathDB	264317	126
Protein domains	VEuPathDB	404090	323
Immunology	VEuPathDB	350433	26
Gene Orthology	VEuPathDB	397151	612
Synteny	VEuPathDB	404090	NA
Metabolic pathways	VEuPathDB	1283	265
Phenotype	VEuPathDB	41631	3293
Isolate data	VEuPathDB	531	5021
Subcellular localization	VEuPathDB	222038	564
ESTs	VEuPathDB	378959	58
Compounds	VEuPathDB	209	1347

Service/Tool Usage

Both BRC analysis services and tools allow users to analyze data pulled from the respective BRC databases and their own private data, compare to other datasets, and save the results in their private workspaces. Since the types of tools vary across the BRCs, we will report aggregated usage of all tools in each BRC, and also a breakdown by service/tool. We will also report the total amount of storage used for user data.

- **Total number of analysis tasks submitted and completed successfully by users**
  - *Definition* - The total number of analysis tasks submitted and completed successfully by users for a given month. An analysis task usually involves users providing input data/search terms and/or parameters to initiate a search or analysis task, which may perform one or more searches, data transformations, or data analysis steps, generate results that provide additional insights into the data and present it back to the user in structured view and/or file formats via web interface and/or user workspace.

- *Measurement mechanism* - Analysis tasks are recorded via website and server logs, which are used to tally the number.
- *Measure* - Analysis tasks submitted and completed successfully per month.

- **Analysis tasks submitted and successfully completed by service/tool**
  - *Definition* - A breakdown of total number of analysis tasks (see metric above), summarized by service/tool during the specified date range.
  - *Measurement mechanism* - Analysis tasks submitted by users are captured via website and server logs, which are used to tally the number.
  - *Measure* - Jobs per month, tallied by service/tool.

Table 4. VEuPathDB Tools/Services Usage Metrics (July 1-31, 2021)

Tool/Service	**BRC Domain**	**Submitted**
Sequence retrieval tool	VEuPathDB	11062
BLAST	VEuPathDB	11573
Enrichment Analyses	VEuPathDB	1599
Web services	VEuPathDB	5094
Boolean operations	VEuPathDB	2516
Apollo (Access)	VEuPathDB	956
Site Search	VEuPathDB	132948
Galaxy Jobs	VEuPathDB	3024
Genome Browser	VEuPathDB	364566
User Comments	VEuPathDB	37
Multiple sequence alignment (isolates)	VEuPathDB	4791
Results downloads	VEuPathDB	4446
Data analysis searches (breakdown below)		
Annotation searches	VEuPathDB	4147
Epigenomics	VEuPathDB	4
Function prediction	VEuPathDB	156
Gene models	VEuPathDB	100
Genetic variation	VEuPathDB	99
Genomic Location	VEuPathDB	173
Immunology	VEuPathDB	26

Orthology and synteny	VEuPathDB	612
Pathways and interactions	VEuPathDB	77
Phenotype	VEuPathDB	3293
Protein features and properties	VEuPathDB	348
Protein targeting and localization	VEuPathDB	564
Proteomics	VEuPathDB	1748
Sequence analysis	VEuPathDB	8822
Structure analysis	VEuPathDB	20
Taxonomy	VEuPathDB	451
Text	VEuPathDB	938
Transcriptomics	VEuPathDB	34096
Popset Isolate Sequences	VEuPathDB	5021
Genomic Sequences	VEuPathDB	3270
Genomic Segments	VEuPathDB	204
SNPs	VEuPathDB	5226
ESTs	VEuPathDB	58
Metabolic Pathways	VEuPathDB	265
Compounds	VEuPathDB	1347

Publications and Citations

Publications and citations provide insights into how the BRC is moving science and technology forward and how the resources are serving their respective research communities. Lists of BRC-generated publications (including publications supported by the BRC program in collaboration with various partners) are updated when new manuscripts are accepted and published. Citations to BRC resources are measured using Google Scholar and augmented using PubMed and custom queries as needed to identify citations to the resource that do not cite the official reference publication(s).

- **Citations to BRC publications**
  - *Definition* - Citations to the BRC as measured by citations to key BRC publications, which describe the overall BRC resources, new data and/or analysis tools, or novel use cases supported by them.
  - *Measurement mechanism* - Set up a common Google Scholar profile covering key BRC resource publications (grouped by BRC) and show aggregated citations for each group. The use of Google Scholar profile makes it easier to view the list of publications used to track citations, update the list with new publications, and provide citation counts for individual publications as well as aggregated counts for each resource. Below is the link to the common BRC Google Scholar Profile.

        - https://scholar.google.com/citations?user=kXLGwkYAAAAJ
    - *Measure* - Cumulative number of citations, year to date.

- **Citations to BRC resources**
    - *Definition* - Citations to the BRC resource as measured Google Scholar searches using predetermined set of keywords based on name and/or acronym of each of the BRC resources, and additional keywords to filter out any false positive or negative results to the extent possible. This is complementary to the citations to the BRC publications described above and necessary because, often, users cite BRC resources by mentioning the resource name or URL in the manuscript text, instead of citing relevant publications.
    - *Measurement mechanism* - Define set of keywords based on name and/or acronym of each of the BRC resources and additional keywords to filter out any false positive or negative results to the extent possible. Using these keywords as search terms, create Google Scholar URLs for each of the BRC resources, which will be checked every month to report a cumulative number of citations for each resource. Because of the limitations of the logical and advanced query operations supported by Google Scholar search interface, we are dividing BV-BRC query into three distinct sub queries as shown below.
        - VEuPathDB (merged DB, including legacy VectorBase, FungiDB & parasite resources): https://scholar.google.com/scholar?q=OrthoMCL+OR+PlasmoDB+OR+ToxoDB+OR+CryptoDB+OR+TrichDB+OR+GiardiaDB+OR+TriTrypDB+OR+AmoebaDB+OR+MicrosporidiaDB+OR+%22FungiDB%22+OR+PiroplasmaDB+OR+%22vectorbase%22+OR+veupathdb+OR+ApiDB+OR+EuPathDB+-encrypt+-cryptography+-hymenoptera
    - *Measure* - Cumulative number of citations, year to date.

Table 5: Citations

Metric	Year to date	Cumulative
Citations of BRC Publications	760	10662
Citations of BRC Resources	1540	24100

User Activities

Outreach activities provide additional channels to engage users. User requests for help typically come in through the help desk functionality available from both BRC websites and are tracked using ticketing software tools. Webinar and workshop participants are counted at the time of registration and participation at the event. Counts of access to recorded webinars may be used to augment the total. Followers on social media (Twitter, Facebook, YouTube) are counted using the built-in mechanisms those platforms provide.

- **Total storage used for user data**
    - *Definition* - Total amount of disk storage in use to host user data at the specified date. This metric provides an additional indication of resource usage that may not be reflected by website traffic or analysis jobs.
    - *Measurement mechanism* - Inspection of disk usage via query or automated script.
    - *Measure* - Total terabytes (TB) currently in use.

- **User requests for help**
    - *Definition* - Total number of user-initiated contacts to the BRC to request help or information during the specified date range. In addition to summarizing total user requests, we will also summarize them by the following categories: Requests for help, Bug reports, and New features / enhancements.

- *Measurement mechanism* - Manual tally of the auto-generated helpdesk tickets triggered by user requests. Tallies may be augmented with manual counts of interactions where the user bypassed the helpdesk system, *e.g.* via direct email or messaging to BRC team members.
- *Measure* - Requests per month. Note that because some emails fit into multiple categories the total percent can exceed 100.

- **Webinar/workshop events and participants**
  - *Definition* - Total number of outreach events (*i.e*., BRC webinars, workshops, and online courses) held per month and total number of participants who attended those events.
  - *Measurement mechanism* - Manual tally of participants in attendance at the time of the webinar or workshop, summed over all of the events held per month.
  - *Measure* - Cumulative number of participants per month
- **Followers on social media**
  - *Definition* - Total number of followers, by social media outlet, at the specified date. Current active BRC social media outlets are Twitter, Facebook, and YouTube.
  - *Measurement mechanism* - Inspection of the number of followers reported by the media outlet at the specified date.
  - *Measure* - Total number of followers, by media outlet.

Table 6: VEuPathDB User Activities (**July 1-31)**

Metric	**Results (reporting period)**
VEuPathDB integrated user data	~53G
Galaxy user data	~11T
User requests for help (some fit multiple categories and total may be >100%)	56 (16% bugs, 68% help, 13% new data, 9% new feature, 2% other)
Webinar/workshop events and participants	None
Followers on social media: (reported as total)	
FaceBook @VEuPathDB	1839
FaceBook @FungiDB	558
FaceBook @VectorBase	2165
Twitter @VEuPathDB	2822
Twitter @FungiDB	3194
Twitter @VectorBase	1939
YouTube	538