



VEuPathDB BRC contract HHSN75N93019C00077

Usage Metrics Report

Reporting Period: June 1-30, 2021

Submission Date: July 12, 2021

Revision History

Date	Version/release	Description
7/12/2021	1	VEuPathDB Performance Metrics for June 2021

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 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 (June 1-30, 2021)

Metric	Result
Total visits	78,633
Total unique visitors	32,214
Total pageviews	11,648,697
Avg. pages / visit	148.14
Avg. visits / visitor	2.44
Avg. visit duration (seconds)	602
Bandwidth (GB)	391.10

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 (June 1-30, 2021)

<i>Taxa</i>	Domain	Page Views	# of Species	# of Genome Seqs
<i>Plasmodium</i>	Protozoa	198715	22	45
<i>Toxoplasma</i>	Protozoa	54897	1	15
<i>Trypanosoma</i>	Protozoa	47296	8	25
<i>Saccharomyces</i>	Fungi	22336	1	1
<i>Leishmania</i>	Protozoa	17815	15	22
<i>Aedes</i>	Vectors	16014	2	3
<i>Cryptococcus</i>	Fungi	13340	5	10
<i>Aspergillus</i>	Fungi	12044	23	28
<i>Anopheles</i>	Vectors	10767	19	22
<i>Cryptosporidium</i>	Protozoa	6357	7	11
<i>Fusarium</i>	Fungi	6275	6	12
<i>Neurospora</i>	Fungi	4077	3	3
<i>Giardia</i>	Protozoa	3977	4	6
<i>Entamoeba</i>	Protozoa	3684	5	9
<i>Biomphalaria</i>	Vectors	2777	1	1
<i>Rhodnius</i>	Vectors	2385	1	1
<i>Pyricularia</i>	Fungi	2352	1	2
<i>Babesia</i>	Protozoa	2177	6	6
<i>Ustilago</i>	Fungi	2029	1	1
<i>Phytophthora</i>	Fungi	2029	7	7
<i>Trichomonas</i>	Protozoa	2020	1	1
<i>Eimeria</i>	Protozoa	1955	8	8
<i>Neospora</i>	Protozoa	1849	1	1

<i>Candida</i>	Fungi	1577	8	15
<i>Acanthamoeba</i>	Protozoa	1318	1	1
<i>Theileria</i>	Protozoa	1161	4	4
<i>Ixodes</i>	Vectors	1135	1	2
<i>Culex</i>	Vectors	1048	1	1
<i>Glossina</i>	Vectors	1000	6	6
<i>Drosophila</i>	Vectors	966	1	1
<i>Besnoitia</i>	Protozoa	841	1	1
<i>Bodo</i>	Protozoa	830	1	1
<i>Crithidia</i>	Protozoa	747	1	1
<i>Naegleria</i>	Protozoa	675	2	3
<i>Paratrypanosoma</i>	Protozoa	608	1	1
<i>Angomonas</i>	Protozoa	598	1	1
<i>Trichoderma</i>	Fungi	575	2	2
<i>Sarcocystis</i>	Protozoa	514	1	2
<i>Lutzomyia</i>	Vectors	495	1	1
<i>Mucor</i>	Fungi	484	2	2
<i>Leptomonas</i>	Protozoa	453	2	2
<i>Hammondia</i>	Protozoa	435	1	1
<i>Blechomonas</i>	Protozoa	390	1	1
<i>Coccidioides</i>	Fungi	390	2	5
<i>Endotrypanum</i>	Protozoa	374	1	1
<i>Cyclospora</i>	Protozoa	374	1	2
<i>Botrytis</i>	Fungi	371	1	1
<i>Cimex</i>	Vectors	369	1	1
<i>Encephalitozoon</i>	Protozoa	366	4	9
<i>Hepatocystis</i>	Protozoa	360	1	1

<i>Schizosaccharomyces</i>	Fungi	341	3	3
<i>Sclerotinia</i>	Fungi	293	1	1
<i>Pediculus</i>	Vectors	289	1	1
<i>Musca</i>	Vectors	287	1	1
<i>Cystoisospora</i>	Protozoa	284	1	1
<i>Sarcoptes</i>	Vectors	283	1	1
<i>Phlebotomus</i>	Vectors	278	1	1
<i>Histoplasma</i>	Fungi	267	1	5
<i>Coprinopsis</i>	Fungi	250	1	1
<i>Spizellomyces</i>	Fungi	250	1	1
<i>Stomoxys</i>	Vectors	246	1	1
<i>Nosema</i>	Protozoa	244	2	3
<i>Culicoides</i>	Vectors	223	1	1
<i>Clavispora</i>	Fungi	208	1	1
<i>Penicillium</i>	Fungi	196	1	1
<i>Homo</i>	Host	186	1	1
<i>Chromera</i>	Protozoa	176	1	1
<i>Leptotrombidium</i>	Vectors	158	1	1
<i>Gregarina</i>	Protozoa	154	1	1
<i>Paracoccidioides</i>	Fungi	133	2	3
<i>Rhizopus</i>	Fungi	127	1	1
<i>Ascosphaera</i>	Fungi	122	1	1
<i>Monocercomonoides</i>	Protozoa	119	1	1
<i>Zymoseptoria</i>	Fungi	119	1	2
<i>Cytauxzoon</i>	Protozoa	118	1	1
<i>Thermothelomyces</i>	Fungi	112	1	1
<i>Anncaliia</i>	Protozoa	109	1	2

<i>Phycomyces</i>	Fungi	108	1	1
<i>Spironucleus</i>	Protozoa	107	1	1
<i>Scedosporium</i>	Fungi	105	1	1
<i>Malassezia</i>	Fungi	101	2	3
<i>Kwoniella</i>	Fungi	99	3	3
<i>Phanerochaete</i>	Fungi	97	1	1
<i>Edhazardia</i>	Protozoa	96	1	1
<i>Sordaria</i>	Fungi	83	1	1
<i>Lomentospora</i>	Fungi	81	1	1
<i>Macaca</i>	Host	77	2	2
<i>Trichosporon</i>	Fungi	69	1	1
<i>Nematocida</i>	Protozoa	68	3	5
<i>Yarrowia</i>	Fungi	67	1	2
<i>Puccinia</i>	Fungi	63	4	4
<i>Sporisorium</i>	Fungi	62	1	1
<i>Tremella</i>	Fungi	62	1	1
<i>Mus</i>	Host	61	1	1
<i>Vitrella</i>	Protozoa	60	1	1
<i>Sporothrix</i>	Fungi	59	2	2
<i>Penicillium</i>	Fungi	59	1	1
<i>Colletotrichum</i>	Fungi	57	1	1
<i>Melampsora</i>	Fungi	56	1	1
<i>Globisporangium</i>	Fungi	56	3	4
<i>Enterocytozoon</i>	Protozoa	56	2	2
<i>Aphanomyces</i>	Fungi	54	2	2
<i>Saprolegnia</i>	Fungi	50	2	2
<i>Podospira</i>	Fungi	46	1	1

<i>Mitosporidium</i>	Protozoa	45	1	1
<i>Ophiostoma</i>	Fungi	45	1	1
<i>Pythium</i>	Fungi	44	2	2
<i>Bos</i>	Host	41	1	1
<i>Spraguea</i>	Protozoa	41	1	1
<i>Albugo</i>	Fungi	37	2	2
<i>Talaromyces</i>	Fungi	36	2	2
<i>Cladophialophora</i>	Fungi	35	2	2
<i>Trachipleistophora</i>	Protozoa	33	1	1
<i>Cenococcum</i>	Fungi	32	1	1
<i>Phytopythium</i>	Fungi	32	1	1
<i>Fonsecaea</i>	Fungi	31	1	1
<i>Blastomyces</i>	Fungi	29	3	3
<i>Hepatospora</i>	Protozoa	29	1	2
<i>Bremia</i>	Fungi	29	1	1
<i>Vittaforma</i>	Protozoa	28	1	1
<i>Rhizophagus</i>	Fungi	27	1	2
<i>Hyaloperonospora</i>	Fungi	27	1	1
<i>Batrachochytrium</i>	Fungi	27	1	1
<i>Vavraia</i>	Protozoa	20	1	1
<i>Pseudoloma</i>	Protozoa	18	1	1
<i>Pichia</i>	Fungi	18	1	1
<i>Enterospora</i>	Protozoa	16	1	1
<i>Allomyces</i>	Fungi	15	1	1
<i>Exophiala</i>	Fungi	15	3	3
<i>Hanseniaspora</i>	Fungi	15	2	2
<i>Ordospora</i>	Protozoa	13	1	1

<i>Pseudogymnoascus</i>	Fungi	11	1	1
<i>Uncinocarpus</i>	Fungi	11	1	1
<i>Verruconis</i>	Fungi	8	1	1
<i>Pneumocystis</i>	Fungi	7	1	1
<i>Cyphellophora</i>	Fungi	3	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 June 1-30, 2021)

Data Type	Domain	Page Views	Searches
Taxonomy	VEuPathDB	439583	569
Genomes	VEuPathDB	439583	1689
Genome sequences	VEuPathDB	439583	3644
Genes/Proteins	VEuPathDB	439583	63253
Transcriptomics	VEuPathDB	376099	33521
Proteomics	VEuPathDB	319577	2109
Variation data	VEuPathDB	276844	5212
Epigenomics	VEuPathDB	246829	10
Enzyme commission	VEuPathDB	148875	88
Gene Ontology	VEuPathDB	305229	280
Protein domains	VEuPathDB	439583	250
Immunology	VEuPathDB	381224	117
Gene Orthology	VEuPathDB	430132	1249
Synteny	VEuPathDB	439583	NA
Metabolic pathways	VEuPathDB	1523	272
Phenotype	VEuPathDB	57668	3259
Isolate data	VEuPathDB	603	4985
Subcellular localization	VEuPathDB	296810	1312

ESTs	VEuPathDB	412313	130
Compounds	VEuPathDB	456	1299

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 and successfully completed 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 (June 1-30, 2021)

Tool/Service	BRC Domain	Submitted	Completed
Sequence retrieval tool	VEuPathDB	14622	12323
BLAST	VEuPathDB	13788	13629
Enrichment Analyses	VEuPathDB	2220	2220
Web services	VEuPathDB	6631	6251
Boolean operations	VEuPathDB	4668	4668
Apollo (Access)	VEuPathDB	967	755
Site Search	VEuPathDB	176786	176754
Galaxy Jobs	VEuPathDB	1949	1918
Genome Browser	VEuPathDB	487844	487844
User Comments	VEuPathDB	8398	8398

Multiple sequence alignment (isolates)	VEuPathDB	5812	5496
Results downloads	VEuPathDB	6681	6681
<i>Data analysis searches (breakdown below)</i>			
Annotation searches	VEuPathDB	4527	4527
Epigenomics	VEuPathDB	10	10
Function prediction	VEuPathDB	368	368
Gene models	VEuPathDB	98	98
Genetic variation	VEuPathDB	198	198
Genomic Location	VEuPathDB	183	183
Immunology	VEuPathDB	117	117
Orthology and synteny	VEuPathDB	1249	1249
Pathways and interactions	VEuPathDB	73	73
Phenotype	VEuPathDB	3259	3259
Protein features and properties	VEuPathDB	3470	3470
Protein targeting and localization	VEuPathDB	1312	1312
Proteomics	VEuPathDB	2109	2109
Sequence analysis	VEuPathDB	10933	10933
Structure analysis	VEuPathDB	27	27
Taxonomy	VEuPathDB	569	569
Text	VEuPathDB	1230	1230
Transcriptomics	VEuPathDB	33521	33521
Popset Isolate Sequences	VEuPathDB	4985	4985
Genomic Sequences	VEuPathDB	3409	3409
Genomic Segments	VEuPathDB	235	235
SNPs	VEuPathDB	5014	5014
ESTs	VEuPathDB	130	130
Metabolic Pathways	VEuPathDB	272	272
Compounds	VEuPathDB	1299	1299

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+CryptDB+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	674	10551
Citations of BRC Resources	1370	23000

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 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 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 (June 1-30)

Metric	Results (reporting period)
Total registered users	23266
VEuPathDB integrated user data	~51G
Galaxy user data	~11T

User requests for help (some fit multiple categories and total may be >100%)	75 (21% bugs, 43% help, 12% new data, 19% new feature, 5% other)
Webinar/workshop events and participants	1 Webinars, 39 participants 1 workshop 50 participants
Followers on social media: (reported as total)	
FaceBook @VEuPathDB	1825
FaceBook @FungiDB	558
FaceBook @VectorBase	2144
Twitter @VEuPathDB	2805
Twitter @FungiDB	3175
Twitter @VectorBase	1915
YouTube	531