VEuPathDB BRC contract HHSN75N93019C00077

Usage Metrics Report

Reporting Period: February 1-28, 2021
Submission Date: March 10, 2021

Revision History

Date	Version/release	Description
3/10/2021	1	Initial VEuPathDB Usage Metrics Report

Joint-BRC Common Usage Metrics Plan

This report is 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. In addition, we will also provide links to the live website usage statistics pages generated by AWStats from respective BRC websites, which will 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 (Feb 1-28, 2021)

Metric	Result
Total visits	65,970
Total unique visitors	28,611
Total pageviews	10,906,185
Avg. pages / visit	165.32
Avg. visits / visitor	2.3
Avg. visit duration (seconds)	613
Bandwidth (GB)	327.94

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 pageviews 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 (Feb 1-28, 2021)

Taxa	Domain	Page Views	# of Species	# of Genome Seqs
Plasmodium	Protozoa	189325	22	45
Toxoplasma	Protozoa	80257	1	15
Trypanosoma	Protozoa	79986	8	23
Leishmania	Protozoa	19198	15	22
Cryptococcus	Fungi	16828	5	10
Cryptosporidium	Protozoa	15803	7	11
Aspergillus	Fungi	12939	23	27
Entamoeba	Protozoa	11533	5	9
Aedes	Vectors	11267	2	2
Saccharomyces	Fungi	10039	1	1
Candida	Fungi	9106	7	9
Anopheles	Vectors	7789	19	22
Trichomonas	Protozoa	5219	1	1
Fusarium	Fungi	4510	5	12
Neurospora	Fungi	4320	3	3
Mucor	Fungi	2861	1	2
Giardia	Protozoa	2739	4	6
Biomphalaria	Vectors	1256	1	1
Coccidioides	Fungi	1252	2	5
Homo	Host	1195	1	1

Culex	Vectors	1095	1	1
Cystoisospora	Protozoa	1033	1	1
Crithidia	Protozoa	984	1	1
Magnaporthe	Fungi	949	1	2
Acanthamoeba	Protozoa	896	1	1
Eimeria	Protozoa	874	8	8
Neospora	Protozoa	757	1	1
Babesia	Protozoa	567	6	6
Histoplasma	Fungi	543	1	5
Rhodnius	Vectors	533	1	1
Theileria	Protozoa	533	4	4
Phytophthora	Fungi	511	7	7
Glossina	Vectors	487	6	6
Ixodes	Vectors	448	1	2
Phycomyces	Fungi	411	1	1
Besnoitia	Protozoa	401	1	1
Bodo	Protozoa	398	1	1
Schizosaccharomyces	Fungi	354	3	3
Naegleria	Protozoa	344	2	3
Leptomonas	Protozoa	322	2	2
Sordaria	Fungi	314	1	1
Hammondia	Protozoa	299	1	1
Paratrypanosoma	Protozoa	289	1	1
Pediculus	Vectors	275	1	1

Blechomonas	Protozoa	275	1	1
Sarcocystis	Protozoa	273	1	2
Musca	Vectors	272	1	1
Lutzomyia	Vectors	269	1	1
Zymoseptoria	Fungi	240	1	1
Ustilago	Fungi	238	1	1
Spizellomyces	Fungi	221	1	1
Cimex	Vectors	218	1	1
Endotrypanum	Protozoa	217	1	1
Cyclospora	Protozoa	201	1	2
Phlebotomus	Vectors	197	1	1
Pythium	Fungi	195	6	7
Trichoderma	Fungi	192	2	2
Vitrella	Protozoa	186	1	1
Stomoxys	Vectors	181	1	1
Botrytis	Fungi	173	1	1
Encephalitozoon	Protozoa	171	4	9
Nematocida	Protozoa	166	3	5
Chromera	Protozoa	162	1	1
Nosema	Protozoa	154	2	3
Paracoccidioides	Fungi	150	2	3
Malassezia	Fungi	149	2	2
Phanerochaete	Fungi	135	1	1
Allomyces	Fungi	126	1	1

Puccinia	Fungi	122	2	2
Sporothrix	Fungi	104	2	2
Talaromyces	Fungi	98	2	2
Gregarina	Protozoa	96	1	1
Coprinopsis	Fungi	95	1	1
Batrachochytrium	Fungi	88	1	1
Scedosporium	Fungi	86	1	1
Kwoniella	Fungi	78	3	3
Sclerotinia	Fungi	76	1	1
Enterocytozoon	Protozoa	76	2	2
Sporisorium	Fungi	75	1	1
Tremella	Fungi	73	1	1
Pneumocystis	Fungi	69	1	1
Rhizophagus	Fungi	68	1	2
Sarcoptes	Vectors	68	1	1
Leptotrombidium	Vectors	66	1	1
Rhizopus	Fungi	65	1	1
Cladophialophora	Fungi	63	2	2
Penicillium	Fungi	63	1	1
Spironucleus	Protozoa	61	1	1
Clavispora	Fungi	59	1	1
Melampsora	Fungi	57	1	1
Monocercomonoides	Protozoa	56	1	1
Fonsecaea	Fungi	55	1	1

Cyphellophora	Fungi	54	1	1
Yarrowia	Fungi	53	1	2
Exophiala	Fungi	52	2	2
Cenococcum	Fungi	52	1	1
Cytauxzoon	Protozoa	45	1	1
Anncaliia	Protozoa	45	1	2
Bos	Host	39	1	1
Lomentospora	Fungi	34	1	1
Uncinocarpus	Fungi	31	1	1
Hepatospora	Protozoa	29	1	2
Vittaforma	Protozoa	27	1	1
Macaca	Host	22	1	1
Aphanomyces	Fungi	19	2	2
Albugo	Fungi	18	2	2
Hyaloperonospora	Fungi	11	1	1
Saprolegnia	Fungi	10	2	2
Spraguea	Protozoa	10	1	1
Mus	Host	10	1	1
Mitosporidium	Protozoa	4	1	1
Pseudoloma	Protozoa	1	1	1
Ordospora	Protozoa	1	1	1
Edhazardia	Protozoa	1	1	1
Amphiamblys	Protozoa	1	1	1
Enterospora	Protozoa	1	1	1

Trachipleistophora	Protozoa	1	1	1
Vavraia	Protozoa	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 (Feb 1-28, 2021)

Data Type	BRC Domain	Page Views	Searches
Taxonomy	VEuPathDB	478764	519
Genomes	VEuPathDB	478764	1527
Genome sequences	VEuPathDB	478764	3413
Genes/Proteins	VEuPathDB	478764	17932
Transcriptomics	VEuPathDB	397825	842
Proteomics	VEuPathDB	374767	131
Variation data	VEuPathDB	338373	183
Epigenomics	VEuPathDB	270854	8
Enzyme commission annotation	VEuPathDB	205673	55
Gene Ontology	VEuPathDB	309013	384
Protein domains (InterPro)	VEuPathDB	478764	298
Immunology	VEuPathDB	414110	101
Orthology	VEuPathDB	478764	626
Synteny	VEuPathDB	478764	NA
Metabolic pathways	VEuPathDB	590	125
Phenotypic	VEuPathDB	100208	40
Subcellular localization	VEuPathDB	336107	652
Isolate data	VEuPathDB	720	183
ESTs	VEuPathDB	449800	167

Compounds	VEuPathDB	246	66

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 (Feb 1-28, 2021)

Tool/Service	BRC Domain	Submitted	Completed
Sequence retrieval tool	VEuPathDB	11724	11724
BLAST	VEuPathDB	10570	10447
Enrichment Analyses	VEuPathDB	1607	1607
Web services	VEuPathDB	27955	27215
Boolean operations	VEuPathDB	2615	2615
Apollo (Access)	VEuPathDB	271	271
Site Search	VEuPathDB	153596	153576
Galaxy Jobs	VEuPathDB	4261	3849
Genome Browser	VEuPathDB	455650	455650
User Comments	VEuPathDB	24	24
Multiple sequence alignment (isolates)	VEuPathDB	5789	5789
Results downloads	VEuPathDB	5940	5940

Data analysis searches (all, see below for breakdown)	VEuPathDB	21960	21960
Annotation searches	VEuPathDB	4418	4418
Epigenomics	VEuPathDB	8	8
Function prediction	VEuPathDB	430	430
Gene models	VEuPathDB	133	133
Genetic variation	VEuPathDB	109	109
Genomic Location	VEuPathDB	333	333
Immunology	VEuPathDB	101	101
Orthology and synteny	VEuPathDB	626	626
Pathways and interactions	VEuPathDB	89	89
Phenotype	VEuPathDB	40	40
Protein features and properties	VEuPathDB	346	346
Protein targeting and localization	VEuPathDB	652	652
Proteomics	VEuPathDB	131	131
Sequence analysis	VEuPathDB	7917	7917
Structure analysis	VEuPathDB	102	102
Taxonomy	VEuPathDB	519	519
Text	VEuPathDB	1155	1155
Transcriptomics	VEuPathDB	842	842
Popset Isolate Sequences	VEuPathDB	183	183
Genomic Sequences	VEuPathDB	3202	3202
Genomic Segments	VEuPathDB	211	211
SNPs	VEuPathDB	74	74
ESTs	VEuPathDB	167	167
Metabolic Pathways	VEuPathDB	125	125
Compounds	VEuPathDB	66	66

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	247	10131
Citations of BRC Resources	471	23300

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**

- o *Definition* - Total cumulative number of users who have registered with the BRC via the website registration mechanism, from inception to the specified date.
- o *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.
- o *Measure* - Total number of registered users (cumulative).

- **Total storage used for user data**
  - o *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.
  - o *Measurement mechanism* - Inspection of disk usage via query or automated script.
  - o *Measure* - Total terabytes (TB) currently in use.

- **User requests for help**
  - o *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.
  - o *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.
  - o *Measure* - Requests per month.

- **Webinar/workshop events and participants**
  - o *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.
  - o *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.
  - o *Measure* - Cumulative number of participants per month

- **Followers on social media**
  - o *Definition* - Total number of followers, by social media outlet, at the specified date. Current active BRC social media outlets are Twitter, Facebook, and YouTube.
  - o *Measurement mechanism* - Inspection of the number of followers reported by the media outlet at the specified date.
  - o *Measure* - Total number of followers, by media outlet.

Table 6: VEuPathDB User Activities (**Feb 1-28, 2021)**

Metric	Results (reporting period)
Total registered users	22340
VEuPathDB integrated user data	~35G
Galaxy user data	~13T
User requests for help	49
Webinar/workshop events and participants	0
Followers on social media: (reported as total)	

FaceBook @VEuPathDB	1770
FaceBook @FungiDB	551
FaceBook @VectorBase	2078
Twitter @VEuPathDB	2696
Twitter @FungiDB	3005
Twitter @VectorBase	1808
YouTube	467