VEuPathDB BRC contract HHSN75N93019C00077

Usage Metrics Report

Reporting Period: March 1-31, 2021
Submission Date: April 12, 2021

Revision History

Date	Version/release	Description
4/12/2021	1	March 2021 VEuPathDB Usage Metrics Report Note that in addition to current metrics, the report has been updated to reflect NIAID feedback from last month's report.

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. 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 (March 1-31, 2021)

Metric	Result
Total visits	84,604
Total unique visitors	34,302
Total pageviews	12,955,282
Avg. pages / visit	153.12
Avg. visits / visitor	2.5
Avg. visit duration (seconds)	580
Bandwidth (GB)	386.93

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 (March 1-31, 2021)

Taxa	**Domain**	**Page Views**	**# of Species**	**# of Genome Seqs**
Plasmodium	Protozoa	174944	22	45
Trypanosoma	Protozoa	163760	8	23
Toxoplasma	Protozoa	127846	1	15
Leishmania	Protozoa	20497	15	22
Cryptococcus	Fungi	18119	5	10
Aspergillus	Fungi	13736	23	28
Anopheles	Vectors	12694	19	22
Aedes	Vectors	12274	2	3
Saccharomyces	Fungi	7370	1	1
Cryptosporidium	Protozoa	7268	7	11
Neurospora	Fungi	5994	3	3
Fusarium	Fungi	5015	6	12
Entamoeba	Protozoa	4250	5	9
Trichomonas	Protozoa	2999	1	1
Neospora	Protozoa	2992	1	1
Crithidia	Protozoa	2976	1	1
Giardia	Protozoa	2874	4	6
Candida	Fungi	2784	8	15
Pyricularia	Fungi	2644	1	2
Rhodnius	Vectors	2364	1	1
Culex	Vectors	1497	1	1

Eimeria	Protozoa	1480	8	8
Naegleria	Protozoa	1092	2	3
Homo	Host	1071	1	1
Babesia	Protozoa	976	6	6
Biomphalaria	Vectors	962	1	1
Ixodes	Vectors	899	1	2
Phytophthora	Fungi	788	7	7
Glossina	Vectors	764	6	6
Sordaria	Fungi	754	1	1
Acanthamoeba	Protozoa	726	1	1
Ustilago	Fungi	630	1	1
Lutzomyia	Vectors	616	1	1
Theileria	Protozoa	597	4	4
Coccidioides	Fungi	592	2	5
Musca	Vectors	583	1	1
Cimex	Vectors	579	1	1
Schizosaccharomyces	Fungi	529	3	3
Botrytis	Fungi	514	1	1
Sclerotinia	Fungi	499	1	1
Leptomonas	Protozoa	490	2	2
Phlebotomus	Vectors	480	1	1
Paracoccidioides	Fungi	480	2	3
Paratrypanosoma	Protozoa	407	1	1
Stomoxys	Vectors	384	1	1

Bodo	Protozoa	372	1	1
Zymoseptoria	Fungi	354	1	1
Pediculus	Vectors	353	1	1
Blechomonas	Protozoa	349	1	1
Malassezia	Fungi	341	2	3
Histoplasma	Fungi	339	1	5
Endotrypanum	Protozoa	330	1	1
Besnoitia	Protozoa	325	1	1
Mucor	Fungi	314	2	2
Trichoderma	Fungi	312	2	2
Coprinopsis	Fungi	288	1	1
Cyclospora	Protozoa	279	1	2
Hammondia	Protozoa	271	1	1
Nosema	Protozoa	261	2	3
Leptotrombidium	Vectors	249	1	1
Spizellomyces	Fungi	248	1	1
Phycomyces	Fungi	239	1	1
Hepatocystis	Protozoa	231	1	1
Chromera	Protozoa	224	1	1
Sarcocystis	Protozoa	224	1	2
Kwoniella	Fungi	220	3	3
Sarcoptes	Vectors	219	1	1
Cystoisospora	Protozoa	216	1	1
Phanerochaete	Fungi	192	1	1

Spironucleus	Protozoa	176	1	1
Monocercomonoides	Protozoa	173	1	1
Tremella	Fungi	171	1	1
Puccinia	Fungi	164	3	3
Encephalitozoon	Protozoa	156	4	9
Yarrowia	Fungi	156	1	2
Batrachochytrium	Fungi	152	1	1
Cenococcum	Fungi	151	1	1
Clavispora	Fungi	150	1	1
Melampsora	Fungi	148	1	1
Talaromyces	Fungi	147	2	2
Globisporangium	Fungi	145	3	4
Nematocida	Protozoa	136	3	5
Gregarina	Protozoa	131	1	1
Allomyces	Fungi	122	1	1
Sporisorium	Fungi	121	1	1
Mus	Host	115	1	1
Scedosporium	Fungi	106	1	1
Sporothrix	Fungi	101	2	2
Culicoides	Vectors	99	1	1
Cladophialophora	Fungi	97	2	2
Rhizopus	Fungi	95	1	1
Penicillium	Fungi	94	1	1
Vitrella	Protozoa	86	1	1

Rhizophagus	Fungi	84	1	2
Exophiala	Fungi	84	3	3
Lomentospora	Fungi	83	1	1
Uncinocarpus	Fungi	75	1	1
Anncaliia	Protozoa	71	1	2
Fonsecaea	Fungi	70	1	1
Cytauxzoon	Protozoa	69	1	1
Thermothelomyces	Fungi	63	1	1
Hyaloperonospora	Fungi	51	1	1
Cyphellophora	Fungi	50	1	1
Hanseniaspora	Fungi	46	1	1
Enterocytozoon	Protozoa	46	2	2
Hepatospora	Protozoa	37	1	2
Pneumocystis	Fungi	34	1	1
Blastomyces	Fungi	33	2	2
Bos	Host	33	1	1
Penicilliopsis	Fungi	26	1	1
Aphanomyces	Fungi	25	2	2
Pythium	Fungi	23	2	2
Albugo	Fungi	23	2	2
Spraguea	Protozoa	23	1	1
Phytopythium	Fungi	21	1	1
Saprolegnia	Fungi	20	2	2
Enterospora	Protozoa	14	1	1

Mitosporidium	Protozoa	14	1	1
Macaca	Host	12	1	1
Ordospora	Protozoa	12	1	1
Vittaforma	Protozoa	9	1	1
Edhazardia	Protozoa	8	1	1
Pseudoloma	Protozoa	7	1	1
Trachipleistophora	Protozoa	6	1	1
Vavraia	Protozoa	4	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 (March 1-31, 2021)

Data Type	BRC Domain	Page Views	Searches
Taxonomy	VEuPathDB	586336	1025
Genomes	VEuPathDB	586336	1863
Genome sequences	VEuPathDB	586336	3585
Genes/Proteins	VEuPathDB	586336	23980
Transcriptomics	VEuPathDB	483778	715
Proteomics	VEuPathDB	477014	142
Variation data	VEuPathDB	418899	251
Epigenomics	VEuPathDB	370447	6
Enzyme commission annotation	VEuPathDB	273696	34
Gene Ontology	VEuPathDB	384145	363
Protein domains (InterPro)	VEuPathDB	586336	309
Immunology	VEuPathDB	496519	176
Orthology	VEuPathDB	586336	834

Synteny	VEuPathDB	586336	NA
Metabolic pathways	VEuPathDB	940	237
Phenotypic	VEuPathDB	216135	125
Subcellular localization	VEuPathDB	429308	1462
Isolate data	VEuPathDB	1003	162
ESTs	VEuPathDB	545991	203
Compounds	VEuPathDB	182	26

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 (March 1-31, 2021)

Tool/Service	BRC Domain	Submitted	Completed
Sequence retrieval tool	VEuPathDB	13647	13647
BLAST	VEuPathDB	14070	13934
Enrichment Analyses	VEuPathDB	1655	1655
Web services	VEuPathDB	5007	5007
Boolean operations	VEuPathDB	3930	3930
Apollo (Access)	VEuPathDB	576	576

Site Search	VEuPathDB	187114	187062
Galaxy Jobs	VEuPathDB	5358	4956
Genome Browser	VEuPathDB	499845	499845
User Comments	VEuPathDB	32	32
Multiple sequence alignment (isolates)	VEuPathDB	5638	5638
Results downloads	VEuPathDB	5868	5868
Data analysis searches (all, see below for breakdown)	VEuPathDB	28313	28313
Annotation searches	VEuPathDB	4339	4339
Epigenomics	VEuPathDB	6	6
Function prediction	VEuPathDB	397	397
Gene models	VEuPathDB	112	112
Genetic variation	VEuPathDB	251	251
Genomic Location	VEuPathDB	171	171
Immunology	VEuPathDB	176	176
Orthology and synteny	VEuPathDB	834	834
Pathways and interactions	VEuPathDB	747	747
Phenotype	VEuPathDB	125	125
Protein features and properties	VEuPathDB	343	343
Protein targeting and localization	VEuPathDB	1462	1462
Proteomics	VEuPathDB	142	142
Sequence analysis	VEuPathDB	11080	11080
Structure analysis	VEuPathDB	29	29
Taxonomy	VEuPathDB	1025	1025
Text	VEuPathDB	2026	2026
Transcriptomics	VEuPathDB	715	715
Popset Isolate Sequences	VEuPathDB	162	162
Genomic Sequences	VEuPathDB	3466	3466
Genomic Segments	VEuPathDB	119	119
SNPs	VEuPathDB	120	120

ESTs	VEuPathDB	203	203
Metabolic Pathways	VEuPathDB	237	237
Compounds	VEuPathDB	26	26

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	361	10242
Citations of BRC Resources	690	23500

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 (**March 1-31)**

Metric	**Results (reporting period)**
Total registered users	22592
VEuPathDB integrated user data	~37G
Galaxy user data	~13T
User requests for help (some fit multiple categories and total may be >100%)	78 (22% bugs, 44% help, 12% new data, 4% new feature, 23% other)
Webinar/workshop events and participants	3 Webinars, 186 participants
Followers on social media: (reported as total)	
FaceBook @VEuPathDB	1776
FaceBook @FungiDB	554
FaceBook @VectorBase	2092
Twitter @VEuPathDB	2709
Twitter @FungiDB	3084
Twitter @VectorBase	1843
YouTube	481