VEuPathDB BRC contract HHSN75N93019C00077

Usage Metrics Report

Reporting Period: April 1-30, 2021
Submission Date: May 10, 2021

Revision History

Date	Version/release	Description
5/10/2021	1	April 2021 VEuPathDB Usage Metrics 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 (April 1-30, 2021)

Metric	Result
Total visits	77,121
Total unique visitors	31,979
Total pageviews	11,448,759
Avg. pages / visit	148.45
Avg. visits / visitor	2.41
Avg. visit duration (seconds)	590
Bandwidth (GB)	447.81

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 (April 1-30, 2021)

Taxa	**Domain**	**Page Views**	**# of Species**	**# of Genome Seqs**
Plasmodium	Protozoa	378968	22	45
Toxoplasma	Protozoa	72552	1	15
Trypanosoma	Protozoa	56574	8	23
Leishmania	Protozoa	34141	15	22
Cryptococcus	Fungi	33052	5	10
Anopheles	Vectors	14803	19	22
Saccharomyces	Fungi	12861	1	1
Aedes	Vectors	11221	2	3
Aspergillus	Fungi	10326	23	28
Cryptosporidium	Protozoa	9505	7	11
Neurospora	Fungi	8550	3	3
Fusarium	Fungi	7020	6	12
Entamoeba	Protozoa	4296	5	9
Pyricularia	Fungi	3149	1	2
Giardia	Protozoa	2889	4	6
Babesia	Protozoa	2272	6	6
Trichomonas	Protozoa	1513	1	1
Eimeria	Protozoa	1408	8	8
Culex	Vectors	1383	1	1
Neospora	Protozoa	1371	1	1

Phytophthora	Fungi	1254	7	7
Homo	Host	1130	1	1
Glossina	Vectors	1068	6	6
Crithidia	Protozoa	1031	1	1
Rhodnius	Vectors	1031	1	1
Ixodes	Vectors	943	1	2
Candida	Fungi	926	8	15
Cystoisospora	Protozoa	778	1	1
Naegleria	Protozoa	715	2	3
Musca	Vectors	657	1	1
Lutzomyia	Vectors	534	1	1
Culicoides	Vectors	515	1	1
Ustilago	Fungi	489	1	1
Cimex	Vectors	481	1	1
Sarcocystis	Protozoa	470	1	2
Histoplasma	Fungi	453	1	5
Biomphalaria	Vectors	440	1	1
Phlebotomus	Vectors	437	1	1
Theileria	Protozoa	433	4	4
Pediculus	Vectors	432	1	1
Acanthamoeba	Protozoa	427	1	1
Botrytis	Fungi	422	1	1
Bodo	Protozoa	393	1	1
Coccidioides	Fungi	391	2	5

Stomoxys	Vectors	382	1	1
Paracoccidioides	Fungi	366	2	3
Leptomonas	Protozoa	357	2	2
Paratrypanosoma	Protozoa	350	1	1
Schizosaccharomyces	Fungi	334	3	3
Phycomyces	Fungi	312	1	1
Endotrypanum	Protozoa	264	1	1
Blechomonas	Protozoa	259	1	1
Hepatocystis	Protozoa	255	1	1
Trichoderma	Fungi	241	2	2
Mucor	Fungi	211	2	2
Sclerotinia	Fungi	203	1	1
Hammondia	Protozoa	192	1	1
Besnoitia	Protozoa	187	1	1
Gregarina	Protozoa	186	1	1
Cyclospora	Protozoa	173	1	2
Nosema	Protozoa	171	2	3
Encephalitozoon	Protozoa	150	4	9
Malassezia	Fungi	142	2	3
Phanerochaete	Fungi	136	1	1
Sordaria	Fungi	134	1	1
Chromera	Protozoa	134	1	1
Zymoseptoria	Fungi	116	1	1
Cytauxzoon	Protozoa	110	1	1

Vitrella	Protozoa	109	1	1
Sporisorium	Fungi	106	1	1
Rhizopus	Fungi	104	1	1
Yarrowia	Fungi	100	1	2
Sarcoptes	Vectors	97	1	1
Leptotrombidium	Vectors	95	1	1
Vavraia	Protozoa	88	1	1
Clavispora	Fungi	86	1	1
Mus	Host	85	1	1
Albugo	Fungi	84	2	2
Kwoniella	Fungi	75	3	3
Coprinopsis	Fungi	73	1	1
Nematocida	Protozoa	64	3	5
Talaromyces	Fungi	63	2	2
Thermothelomyces	Fungi	62	1	1
Globisporangium	Fungi	59	3	4
Saprolegnia	Fungi	52	2	2
Cenococcum	Fungi	46	1	1
Trachipleistophora	Protozoa	41	1	1
Spizellomyces	Fungi	41	1	1
Pythium	Fungi	38	2	2
Anncaliia	Protozoa	38	1	2
Rhizophagus	Fungi	37	1	2
Macaca	Host	37	1	1

Hyaloperonospora	Fungi	37	1	1
Sporothrix	Fungi	36	2	2
Puccinia	Fungi	35	3	3
Blastomyces	Fungi	35	2	2
Lomentospora	Fungi	35	1	1
Batrachochytrium	Fungi	35	1	1
Cladophialophora	Fungi	34	2	2
Phytopythium	Fungi	33	1	1
Scedosporium	Fungi	32	1	1
Enterocytozoon	Protozoa	31	2	2
Tremella	Fungi	31	1	1
Melampsora	Fungi	30	1	1
Spraguea	Protozoa	30	1	1
Hanseniaspora	Fungi	29	1	1
Bos	Host	26	1	1
Aphanomyces	Fungi	25	2	2
Vittaforma	Protozoa	23	1	1
Allomyces	Fungi	21	1	1
Hepatospora	Protozoa	19	1	2
Fonsecaea	Fungi	18	1	1
Monocercomonoides	Protozoa	17	1	1
Penicillium	Fungi	15	1	1
Exophiala	Fungi	15	3	3
Uncinocarpus	Fungi	15	1	1

Spironucleus	Protozoa	11	1	1
Mitosporidium	Protozoa	11	1	1
Amphiamblys	Protozoa	10	1	1
Pneumocystis	Fungi	10	1	1
Edhazardia	Protozoa	9	1	1
Cyphellophora	Fungi	7	1	1
Penicilliopsis	Fungi	6	1	1
Enterospora	Protozoa	5	1	1
Ordospora	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 (April 1-30, 2021)

Data Type	Domain	Page Views	Searches
Phenotype	VEuPathDB	86146	33
Genes/Proteins	VEuPathDB	664057	18843
Enzyme commission	VEuPathDB	294312	67
Isolate data	VEuPathDB	1048	149
Genome sequences	VEuPathDB	664057	4248
ESTs	VEuPathDB	622135	25
Variation data	VEuPathDB	520060	280
Genomes	VEuPathDB	664057	1753
Gene Orthology	VEuPathDB	655306	1157
Immunology	VEuPathDB	585012	74
Epigenomics	VEuPathDB	455326	8
Taxonomy	VEuPathDB	664057	587
Compounds	VEuPathDB	191	33

Synteny	VEuPathDB	664057	NA
Gene Ontology	VEuPathDB	482211	317
Proteomics	VEuPathDB	564472	128
Metabolic pathways	VEuPathDB	1166	175
Protein domains	VEuPathDB	664057	442
Transcriptomics	VEuPathDB	601255	1145
Subcellular localization	VEuPathDB	514838	785

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 (April 1-30, 2021)

Tool/Service	BRC Domain	Submitted	Completed
Sequence retrieval tool	VEuPathDB	12451	12451
BLAST	VEuPathDB	11858	11759
Enrichment Analyses	VEuPathDB	1514	1514
Web services	VEuPathDB	1052	1052
Boolean operations and colocation	VEuPathDB	3372	3372
Apollo (Access)	VEuPathDB	964	964
Site Search	VEuPathDB	173413	173372

Galaxy Jobs	VEuPathDB	2728	2496
Genome Browser	VEuPathDB	463146	463146
User Comments	VEuPathDB	53	53
Multiple sequence alignment (isolates)	VEuPathDB	5977	5977
Results downloads	VEuPathDB	4984	4984
Data analysis searches (all, see below for breakdown)	VEuPathDB	23619	23619
Annotation searches	VEuPathDB	4059	4059
Epigenomics	VEuPathDB	8	8
Function prediction	VEuPathDB	384	384
Gene models	VEuPathDB	132	132
Genetic variation	VEuPathDB	134	134
Genomic Location	VEuPathDB	132	132
Immunology	VEuPathDB	74	74
Orthology and synteny	VEuPathDB	1157	1157
Pathways and interactions	VEuPathDB	60	60
Phenotype	VEuPathDB	33	33
Protein features and properties	VEuPathDB	476	476
Protein targeting and localization	VEuPathDB	785	785
Proteomics	VEuPathDB	128	128
Sequence analysis	VEuPathDB	8500	8500
Structure analysis	VEuPathDB	25	25
Taxonomy	VEuPathDB	587	587
Text	VEuPathDB	1024	1024
Transcriptomics	VEuPathDB	1145	1145
Popset Isolate Sequences	VEuPathDB	149	149
Genomic Sequences	VEuPathDB	3910	3910
Genomic Segments	VEuPathDB	338	338
SNPs	VEuPathDB	146	146
ESTs	VEuPathDB	25	25

Metabolic Pathways	VEuPathDB	175	175
Compounds	VEuPathDB	33	33

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	437	10323
Citations of BRC Resources	941	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 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 (**April 1-30)**

Metric	**Results (reporting period)**
Total registered users	22806
VEuPathDB integrated user data	~37G
Galaxy user data	~14T
User requests for help (some fit multiple categories and total may be >100%)	69 (25% bugs, 49% help, 17% new data, 11% new feature, 6% other)
Webinar/workshop events and participants	1 Webinar, 33 participants
Followers on social media: (reported as total)	
FaceBook @VEuPathDB	1781
FaceBook @FungiDB	557
FaceBook @VectorBase	2106
Twitter @VEuPathDB	2724
Twitter @FungiDB	3117
Twitter @VectorBase	1865
YouTube	494