VEuPathDB BRC contract HHSN75N93019C00077

Usage Metrics Report

Reporting Period: May 1-31, 2021
Submission Date: June 10, 2021

Revision History

Date	Version/release	Description
6/10/2021	1	The following updates are included to provide answers to questions from the COR sent via email on 5/14/2021. Specifically ● The 'Website Usage Metrics' section has been updated to provide links to AWStats pages for individual sites.

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 (May 1-31, 2021)

Metric	Result
Total visits	84,559
Total unique visitors	34,023
Total pageviews	12,272,782
Avg. pages / visit	145.13
Avg. visits / visitor	2.48
Avg. visit duration (seconds)	586
Bandwidth (GB)	453.47

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 (May 1-31, 2021)

Taxa	**Domain**	**Page Views**	**# of Species**	**# of Genome Seqs**
Plasmodium	Protozoa	287138	22	45
Toxoplasma	Protozoa	77707	1	15
Trypanosoma	Protozoa	53274	8	25
Saccharomyces	Fungi	20982	1	1
Cryptococcus	Fungi	17571	5	10
Leishmania	Protozoa	15939	15	22
Aspergillus	Fungi	15301	23	28
Anopheles	Vectors	15141	19	22
Aedes	Vectors	13098	2	3
Cryptosporidium	Protozoa	8847	7	11
Fusarium	Fungi	7221	6	12
Neurospora	Fungi	7162	3	3
Entamoeba	Protozoa	5832	5	9
Giardia	Protozoa	5043	4	6
Phytophthora	Fungi	4035	7	7
Pyricularia	Fungi	3156	1	2
Candida	Fungi	1950	8	15
Trichomonas	Protozoa	1907	1	1
Eimeria	Protozoa	1608	8	8
Glossina	Vectors	1601	6	6
Babesia	Protozoa	1541	6	6
Neospora	Protozoa	1482	1	1
Ixodes	Vectors	1145	1	2

Culex	Vectors	998	1	1
Rhodnius	Vectors	968	1	1
Theileria	Protozoa	877	4	4
Biomphalaria	Vectors	715	1	1
Acanthamoeba	Protozoa	705	1	1
Zymoseptoria	Fungi	648	1	2
Paratrypanosoma	Protozoa	640	1	1
Crithidia	Protozoa	596	1	1
Coccidioides	Fungi	587	2	5
Cimex	Vectors	575	1	1
Sarcocystis	Protozoa	518	1	2
Leptomonas	Protozoa	512	2	2
Bodo	Protozoa	508	1	1
Lutzomyia	Vectors	507	1	1
Histoplasma	Fungi	449	1	5
Ustilago	Fungi	428	1	1
Naegleria	Protozoa	417	2	3
Culicoides	Vectors	393	1	1
Blechomonas	Protozoa	368	1	1
Phlebotomus	Vectors	367	1	1
Schizosaccharomyces	Fungi	353	3	3
Phycomyces	Fungi	343	1	1
Cyclospora	Protozoa	342	1	2
Musca	Vectors	333	1	1
Cystoisospora	Protozoa	324	1	1
Stomoxys	Vectors	321	1	1
Hammondia	Protozoa	313	1	1

Gregarina	Protozoa	306	1	1
Endotrypanum	Protozoa	304	1	1
Hepatocystis	Protozoa	295	1	1
Lomentospora	Fungi	289	1	1
Cytauxzoon	Protozoa	243	1	1
Botrytis	Fungi	236	1	1
Malassezia	Fungi	236	2	3
Besnoitia	Protozoa	208	1	1
Mucor	Fungi	206	2	2
Chromera	Protozoa	202	1	1
Trichoderma	Fungi	197	2	2
Sclerotinia	Fungi	192	1	1
Penicillium	Fungi	186	1	1
Pediculus	Vectors	177	1	1
Angomonas	Protozoa	166	1	1
Drosophila	Vectors	162	1	1
Saprolegnia	Fungi	154	2	2
Encephalitozoon	Protozoa	144	4	9
Paracoccidioides	Fungi	137	2	3
Nosema	Protozoa	137	2	3
Rhizopus	Fungi	119	1	1
Spizellomyces	Fungi	116	1	1
Sordaria	Fungi	111	1	1
Batrachochytrium	Fungi	102	1	1
Thermothelomyces	Fungi	93	1	1
Monocercomonoides	Protozoa	91	1	1
Homo	Host	91	1	1

Phanerochaete	Fungi	88	1	1
Allomyces	Fungi	82	1	1
Puccinia	Fungi	79	4	4
Sarcoptes	Vectors	76	1	1
Kwoniella	Fungi	75	3	3
Exophiala	Fungi	73	3	3
Talaromyces	Fungi	73	2	2
Sporisorium	Fungi	70	1	1
Coprinopsis	Fungi	69	1	1
Melampsora	Fungi	60	1	1
Vitrella	Protozoa	58	1	1
Clavispora	Fungi	57	1	1
Blastomyces	Fungi	56	3	3
Leptotrombidium	Vectors	51	1	1
Mitosporidium	Protozoa	46	1	1
Colletotrichum	Fungi	43	1	1
Tremella	Fungi	43	1	1
Hepatospora	Protozoa	42	1	2
Spironucleus	Protozoa	41	1	1
Uncinocarpus	Fungi	40	1	1
Mus	Host	39	1	1
Ascosphaera	Fungi	38	1	1
Trichosporon	Fungi	37	1	1
Sporothrix	Fungi	37	2	2
Rhizophagus	Fungi	35	1	2
Penicilliopsis	Fungi	35	1	1
Aphanomyces	Fungi	33	2	2

Albugo	Fungi	32	2	2
Cladophialophora	Fungi	30	2	2
Nematocida	Protozoa	27	3	5
Yarrowia	Fungi	27	1	2
Enterocytozoon	Protozoa	26	2	2
Hyaloperonospora	Fungi	24	1	1
Globisporangium	Fungi	23	3	4
Anncaliia	Protozoa	23	1	2
Hanseniaspora	Fungi	22	2	2
Cenococcum	Fungi	19	1	1
Edhazardia	Protozoa	18	1	1
Fonsecaea	Fungi	17	1	1
Cyphellophora	Fungi	14	1	1
Scedosporium	Fungi	12	1	1
Pneumocystis	Fungi	11	1	1
Vavraia	Protozoa	9	1	1
Pythium	Fungi	9	2	2
Podospora	Fungi	8	1	1
Enterospora	Protozoa	7	1	1
Phytopythium	Fungi	7	1	1
Pseudoloma	Protozoa	6	1	1
Vittaforma	Protozoa	6	1	1
Trachipleistophora	Protozoa	6	1	1
Ordospora	Protozoa	6	1	1
Spraguea	Protozoa	6	1	1
Ophiostoma	Fungi	5	1	1
Amphiamblys	Protozoa	3	1	1

Macaca	Host	3	2	2
Pichia	Fungi	2	1	1
Pseudogymnoascus	Fungi	2	1	1
Verruconis	Fungi	2	1	1
Bos	Host	2	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 (May 1-31, 2021)

Data Type	Domain	Page Views	Searches
Taxonomy	VEuPathDB	563895	595
Genomes	VEuPathDB	563895	1834
Genome sequences	VEuPathDB	563895	3797
Genes/Proteins	VEuPathDB	563895	23323
Transcriptomics	VEuPathDB	500269	2302
Proteomics	VEuPathDB	449081	70
Variation data	VEuPathDB	405986	475
Epigenomics	VEuPathDB	356399	4
Enzyme commission	VEuPathDB	184669	40
Gene Ontology	VEuPathDB	416655	290
Protein domains	VEuPathDB	563895	577
Immunology	VEuPathDB	496029	57
Gene Orthology	VEuPathDB	553412	2056
Synteny	VEuPathDB	563895	NA
Metabolic pathways	VEuPathDB	2202	221
Phenotype	VEuPathDB	80865	28
Isolate data	VEuPathDB	721	60
Subcellular localization	VEuPathDB	418455	1422

ESTs	VEuPathDB	537078	65
Compounds	VEuPathDB	286	34

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 (May 1-31, 2021)

Tool/Service	BRC Domain	Submitted	Completed
Sequence retrieval tool	VEuPathDB	12128	12128
BLAST	VEuPathDB	11684	11462
Enrichment Analyses	VEuPathDB	2575	2575
Web services	VEuPathDB	595	595
Boolean operations	VEuPathDB	6711	6711
Apollo (Access)	VEuPathDB	1247	1247
Site Search	VEuPathDB	186490	186461
Galaxy Jobs	VEuPathDB	2441	2145
Genome Browser	VEuPathDB	498638	498638
User Comments	VEuPathDB	46	46
Multiple sequence alignment (isolates)	VEuPathDB	6106	6106

Results downloads	VEuPathDB	5269	5269
Data analysis searches (all, see below for breakdown)	VEuPathDB	27761	27761
Annotation searches	VEuPathDB	5002	5002
Epigenomics	VEuPathDB	4	4
Function prediction	VEuPathDB	330	330
Gene models	VEuPathDB	123	123
Genetic variation	VEuPathDB	214	214
Genomic Location	VEuPathDB	199	199
Immunology	VEuPathDB	57	57
Orthology and synteny	VEuPathDB	2056	2056
Pathways and interactions	VEuPathDB	121	121
Phenotype	VEuPathDB	28	28
Protein features and properties	VEuPathDB	598	598
Protein targeting and localization	VEuPathDB	1422	1422
Proteomics	VEuPathDB	70	70
Sequence analysis	VEuPathDB	8644	8644
Structure analysis	VEuPathDB	437	437
Taxonomy	VEuPathDB	595	595
Text	VEuPathDB	1121	1121
Transcriptomics	VEuPathDB	2302	2302
Popset Isolate Sequences	VEuPathDB	60	60
Genomic Sequences	VEuPathDB	3559	3559
Genomic Segments	VEuPathDB	238	238
SNPs	VEuPathDB	261	261
ESTs	VEuPathDB	65	65
Metabolic Pathways	VEuPathDB	221	221
Compounds	VEuPathDB	34	34

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	549	10444
Citations of BRC Resources	1080	26000

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 (**May 1-31)**

Metric	**Results (reporting period)**
Total registered users	23123
VEuPathDB integrated user data	~49G
Galaxy user data	~11T
User requests for help (some fit multiple categories and total may be >100%)	67 (25% bugs, 48% help, 15% new data, 12% new feature, 12% other)

Webinar/workshop events and participants	2 Webinars, 121 participants 1 workshop 30 participants
Followers on social media: (reported as total)	
FaceBook @VEuPathDB	1794
FaceBook @FungiDB	557
FaceBook @VectorBase	2121
Twitter @VEuPathDB	2766
Twitter @FungiDB	3139
Twitter @VectorBase	1885
YouTube	511