BV-BRC

Bacterial and Viral (BV) -
Bioinformatics Resource Center (BRC)

Monthly Usage Metrics Report

Performance Period: December 1, 2022 – December 31, 2022

Issued to:

National Institute of Allergy and Infectious Diseases
National Institute of Health

Contract No.: 75N93019C00076
Contract Title: Bioinformatics Resource Centers for Infectious Diseases

Submission Date:
January 10, 2023

Submitted by:
Rick Stevens (contact PI)
Associate Laboratory Director
Argonne National Laboratory
Professor, Computer Science
University of Chicago
5801 South Ellis Avenue
Chicago, IL 60637-5418
630.252.3378 (phone)
630.252.6333 (fax)

BV-BRC Usage Metrics Report

Note: As per the recent request from NIAID, we are working with the other BRC to provide jointly agreed plots showing accumulative usage data over time. We will start including them in the monthly reports, starting with the next monthly report.

This monthly usage metrics report provides a summary of the BV-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 and report usage metrics for their constituent parts, *i.e.,* PATRIC and IRD/ViPR for BV-BRC. 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 monthly, and in combination on the BRC Gateway website once this is publicly available. In addition, annual summaries will be included in the Annual Progress Reports.

It is important to note that usage 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.
- **Registered users that run a service**
  - *Definition* – Total number of unique registered users that run an analysis service (requiring login) during the month.
  - *Measurement mechanism* – Service logs.
  - *Measure* - Total unique registered users per month.

Table 1. BV-BRC Website Usage Metrics[1]

Metric	PATRIC	IRD	ViPR	BV-BRC	All Combined
Total visits	1,436	1,270	1,359	154,393	157,321
Total unique visitors	963	444	407	25,634	26,558
Total pageviews	5,522	33,836	20,955	535,671	595,979
Avg. pages / visit	3.84	26.64	15.41	3.46	3.78
Avg. visits / visitor	1.49	2.86	3.33	6.02	5.92
Avg. visit duration (seconds)	299	657	622	720	721

Bandwidth (GB)	0.81	1.01	32.90	40.63	73.32
Registered users that run a service[2,3]	1,261	8	8	1,261	1,269

Notes:

1. *A link to the BV-BRC summary AWStats page is available from the BV-BRC About page (https://www.bv-brc.org/about)*
2. *Note: This measure This will only be a fraction of the total usage by registered users because they may be doing other types of work on the site, either logged in or not.*
3. *PATRIC and BV-BRC Production are the same because both resources use the same computational services infrastructure. Similarly, IRD and ViPR use the same computational infrastructure, so those numbers are the same as well.*
4. *The AWStats settings have been adjusted to filter out data API calls which can double-count some website traffic. The table and the charts below have been updated to reflect these changes.*

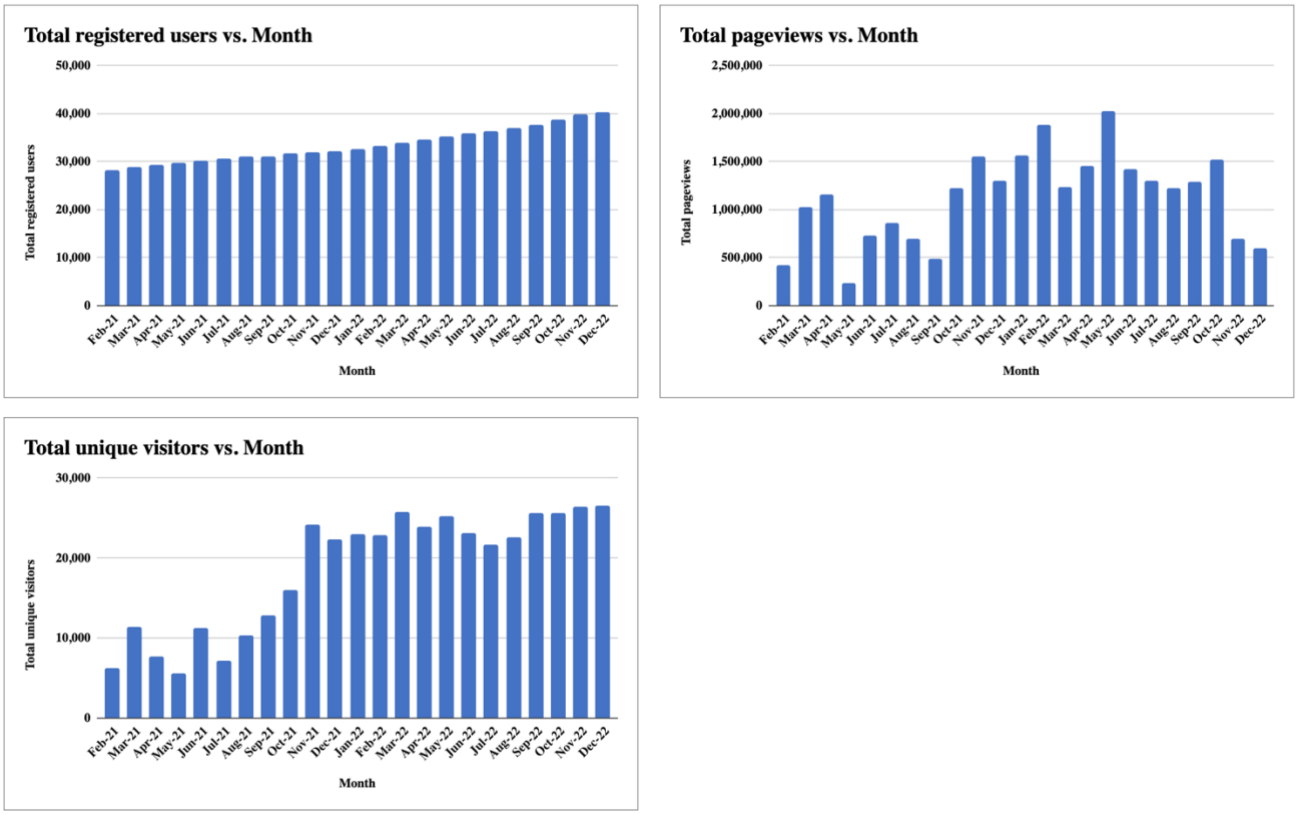

Figure 1. Selected BV-BRC website usage metrics.

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 pageviews by

taxa, which will be measured by querying the website usage statistics in Google Analytics by taxa name.

Table 2. BV-BRC Website Usage by Taxa

Taxa	Taxon ID	Domain	Species[1]	Genomes[1]	Page Views[1,2]
Acinetobacter	469	Bacteria	734	14,728	148
Bacillus	1386	Bacteria	1,028	8,763	1386
Bartonella	773	Bacteria	84	352	26
Borreliella	64895	Bacteria	21	5,822	29
Brucella	234	Bacteria	96	1,277	193
Burkholderia	32008	Bacteria	331	5,534	22
Campylobacter	194	Bacteria	307	9,146	24
Chlamydia	810	Bacteria	23	653	14
Clostridium	1485	Bacteria	498	5,029	59
Coxiella	776	Bacteria	14	208	21
Ehrlichia	943	Bacteria	7	46	51
Escherichia	561	Bacteria	194	50,810	383
Francisella	262	Bacteria	32	1,162	9
Helicobacter	209	Bacteria	93	3,344	28
Listeria	1637	Bacteria	46	6,280	13
Mycobacterium	1763	Bacteria	353	33,820	83
Pseudomonas	286	Bacteria	2,198	18,435	83
Rickettsia	780	Bacteria	75	416	15
Salmonella	590	Bacteria	401	33,343	73
Shigella	620	Bacteria	113	5,488	14
Staphylococcus	1279	Bacteria	591	27,094	193
Streptococcus	1301	Bacteria	464	39,763	78
Vibrio	662	Bacteria	469	7,467	39

Yersinia	629	Bacteria	34	1,719	14
Adenoviridae	10508	Virus	579	24,433	NA
Asfarviridae	137992	Virus	4	9,143	NA
Bunyaviridae	1980410	Virus	1,597	55,931	110
Caliciviridae	11974	Virus	245	63,332	41
Coronaviridae	11118	Virus	1,011	6,753,285	360
Filoviridae	11266	Virus	21	4,460	59
Flaviviridae	11050	Virus	511	368,836	251
Hepadnaviridae	10404	Virus	42	126,983	NA
Hepeviridae	291484	Virus	78	23,587	22
Herpesviridae	10292	Virus	783	65,070	12
Orthomyxoviridae	11308	Virus	153	999,951	5,118
Paramyxoviridae	11158	Virus	678	64,625	19
Parvoviridae	10780	Virus	745	30,154	NA
Picornaviridae	12058	Virus	1,169	162,640	110
Pneumoviridae	11244	Virus	16	49,665	12
Polyomaviridae	151341	Virus	277	13,341	NA
Poxviridae		Virus	283	15,917	97
Reoviridae	2732541	Virus	431	139,801	418
Rhabdoviridae	11270	Virus	714	39,490	10
Togaviridae	11018	Virus	69	12,787	60
SARS-CoV-2 (BV-BRC)	2697049	Virus	1	6,709,783	739

Notes:

1. *Since IRD and ViPR are now redirecting to BV-BRC we are no longer ingesting genomes into those databases. So, the Species and Genomes counts are from BV-BRC for the viral data.*
2. *Virus Page views data is from legacy Google Analytics. Where an N/A appears, ViPR does not have those species. They are only in BV-BRC.*

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 us. 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. VIPR/IRD pages views are combined (added together) based on data type.

Table 3. BV-BRC Website Usage by Data Type

Data Type	BRC Domain[1]	Page Views
Taxonomy	BV-BRC	5,649
Genome	BV-BRC	15,149
Genome sequence	BV-BRC	3
Feature (Genes/Proteins)	BV-BRC	9,691
Specialty (gene)	BV-BRC	768
Protein families	BV-BRC	457
Pathway	BV-BRC	1,763
Subsystem	BV-BRC	336
Transcriptomics	BV-BRC	9
Interactions	BV-BRC	5
Phylogeny	BV-BRC	83
Antibiotic	BV-BRC	63
Workspace (User Data)	BV-BRC	4,635
Genome	IRD/ViPR	2,477
Gene/Protein	IRD/ViPR	423
Strain	IRD/ViPR	291
Immune epitopes	IRD/ViPR	145
Ortholog groups	IRD/ViPR	0
Antiviral drugs	IRD/ViPR	33
Host factors	IRD/ViPR	2
Protein structures	IRD/ViPR	20
Protein domains and motifs	IRD/ViPR	0
Plasmids	IRD/ViPR	1
SFVT	IRD/ViPR	7
Surveillance	IRD/ViPR	37
Serology	IRD/ViPR	0
Phenotypes	IRD/ViPR	13

PCR Primers	IRD/ViPR	28
Variant (SARS Variant Tracker)	BV-BRC	1,062

Notes:

1. *Switched from PATRIC to BV-BRC since redirects in place.*

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. VIPR/IRD tools/services are combined (added together) that are common in both systems.

- **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. BRC Tools/Services Usage Metrics

Tool/Service	BRC Domain	Submitted	Completed
Codon Tree	BV-BRC/PATRIC	551	532
Comparative Systems	BV-BRC	424	385
Comprehensive Genome Analysis	BV-BRC/PATRIC	3,223	3,009
Differential Expression	BV-BRC/PATRIC	18	14
FastqUtils	BV-BRC/PATRIC	796	742
Gene Tree	BV-BRC	122	90
Genome Alignment	BV-BRC/PATRIC	239	221
Genome Annotation	BV-BRC/PATRIC	5,744	5,387
Genome Assembly	BV-BRC/PATRIC	9,915	9,239
Genome Comparison	BV-BRC/PATRIC	249	220

Homology	BV-BRC	1,942	1,854
MSA	BV-BRC	448	335
MetaCATs	BV-BRC	111	107
Metagenome Binning	BV-BRC/PATRIC	485	427
Metagenomic Read Mapping	BV-BRC/PATRIC	219	214
Primer Design (new)	BV-BRC	123	118
RNA-Seq Analysis	BV-BRC/PATRIC	264	123
Subspecies Classification	BV-BRC	353	320
Taxonomic Classification	BV-BRC/PATRIC	889	815
Tn-Seq Analysis	BV-BRC/PATRIC	83	74
Variation Analysis	BV-BRC/PATRIC	435	399
Alignment Viewer	IRD/ViPR	7	6
Antiviral-Resistance-Risk	IRD/ViPR	4	4
BLAST	IRD/ViPR	19	19
Enrichment	IRD/ViPR	0	0
Genotype-Recombination	IRD/ViPR	0	0
H1-Clade Classifier	IRD only	31	29
H1N1-classifier	IRD only	3	3
H5N1-classifier	IRD only	23	23
Ha Numbering	IRD only	148	137
MGC	IRD/ViPR	0	0
MSA	IRD/ViPR	80	76
Mutation-analysis	IRD/ViPR	10	10
Primer3	IRD/ViPR	5	5
Read-seq	IRD/ViPR	8	8
Rva Genotyper	IRD/ViPR	152	144
Short-seqsearch	IRD/ViPR	0	0
SNP-analysis	IRD/ViPR	30	30
Surveillance-data-mapping	IRD/ViPR	1	1
Tbl-formatter	IRD/ViPR	0	0

Tree	IRD/ViPR	55	47
VIGOR Annotator	IRD/ViPR	0	0
SARS-2 Genome Assembly and Annotation	BV-BRC	80	59

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.
- **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
    - BV-BRC:
      - PATRIC BRC: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=%28PATRIC+AND+Wattam%29+OR+%E2%80%9Cpatricbrc%22+OR+%22pathosystems+resource+integration+center%22

- RAST/RASTtk: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=%28RAST+AND+overbeek%29+OR+%22rast.nmpdr.org%22
- IRD/ViPR: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=%22viprbrc%22+OR+%22virus+pathogen+resource%22+OR+%E2%80%9Cfludb%22+OR+%22influenza+research+database%22

- *Measure* - Cumulative number of citations, cumulative.

Table 5. Citations to BRC Publications and Resources

	Number of Citations (YTD)	Number of Citations (Cumulative)
Citations to BV-BRC publications	2,160	17,431
Citations to BV-BRC resources	2,568	19,120

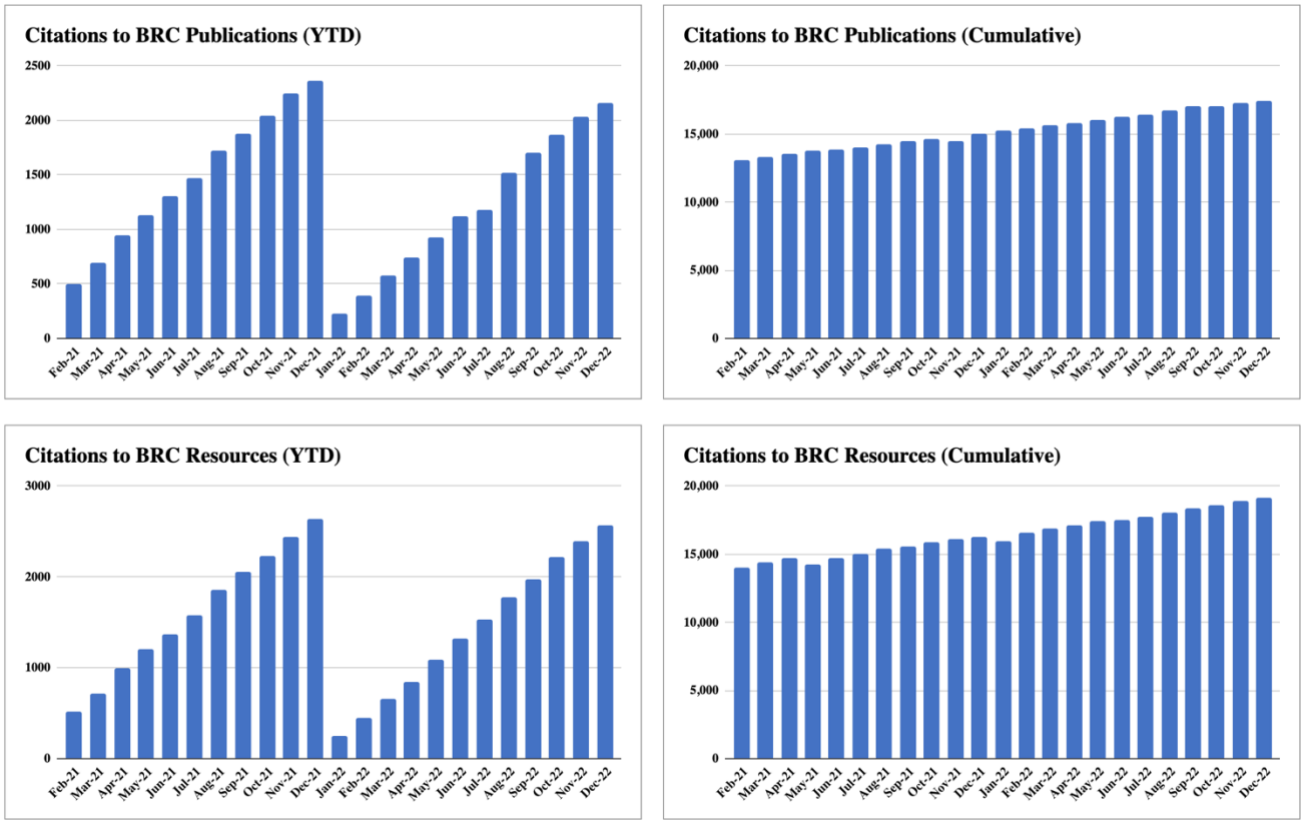

Figure 2. Citations to BV-BRC resources and publications.

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.
- **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.

	PATRIC	**IRD/ViPR**	**BV-BRC**	**Total**
Total registered users	40,411	12,260	40,411	40,411
Total storage used for user data (TB)	265.9	0.52	265.9	266.42
User requests: • Request for help • Report bug • Suggest improvement	0 0 (0%) 0 (0%) 0 (0%)	0 0 (0%) 0 (0%) 0 (0%)	119 119 (100%) 0 (0%) 0 (0%)	119 119 (100%) 0 (0%) 0 (0%)
Webinar/workshop events	0	0	2	2

Total webinar/workshop participants	0	0	151	151
Total MOOC registrants (cumulative)	9,205	NA	NA	9,205
Twitter followers	654	435	235[1]	1,324
Facebook followers	257	2,055	1.1K	3,412
YouTube subscribers	427	196	229	852
YouTube views	414	38	747	1,199
BRC Subreddit members	NA	NA	NA	93
BRC Subreddit views	NA	NA	NA	87

Notes:

1. *BV-BRC Twitter followers was overreported in previous reports.*

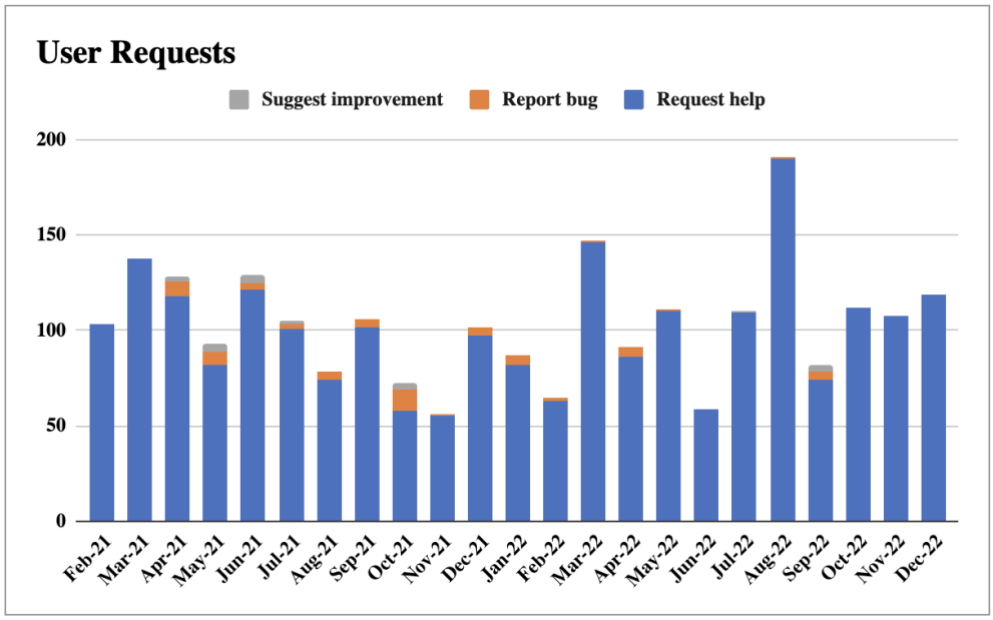

Figure 3. Requests by users, sorted by type.