BV-BRC

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

Monthly Usage Metrics Report

Performance Period: November 1, 2021 – November 30, 2021

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:
December 10, 2021

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 Prod.	All Combined
Total visits	254,687	12,535	22,393	2,714	286,701
Total unique visitors	19,240	7,120	13,798	1,473	36,913
Total pageviews	3,708,371	1,027,219	391,760	6,251	5,138,302
Avg. pages / visit	14.56	81.94	17.49	2.30	17.92
Avg. visits / visitor	13.23	1.76	1.62	1.84	7.76
Avg. visit duration (seconds)	1,389	664	457	187	1,299

Bandwidth (GB)	197.05	97.19	153.67	2.39	450.51
Registered users that run a service[2,3]	899	90	90	899	989

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.*

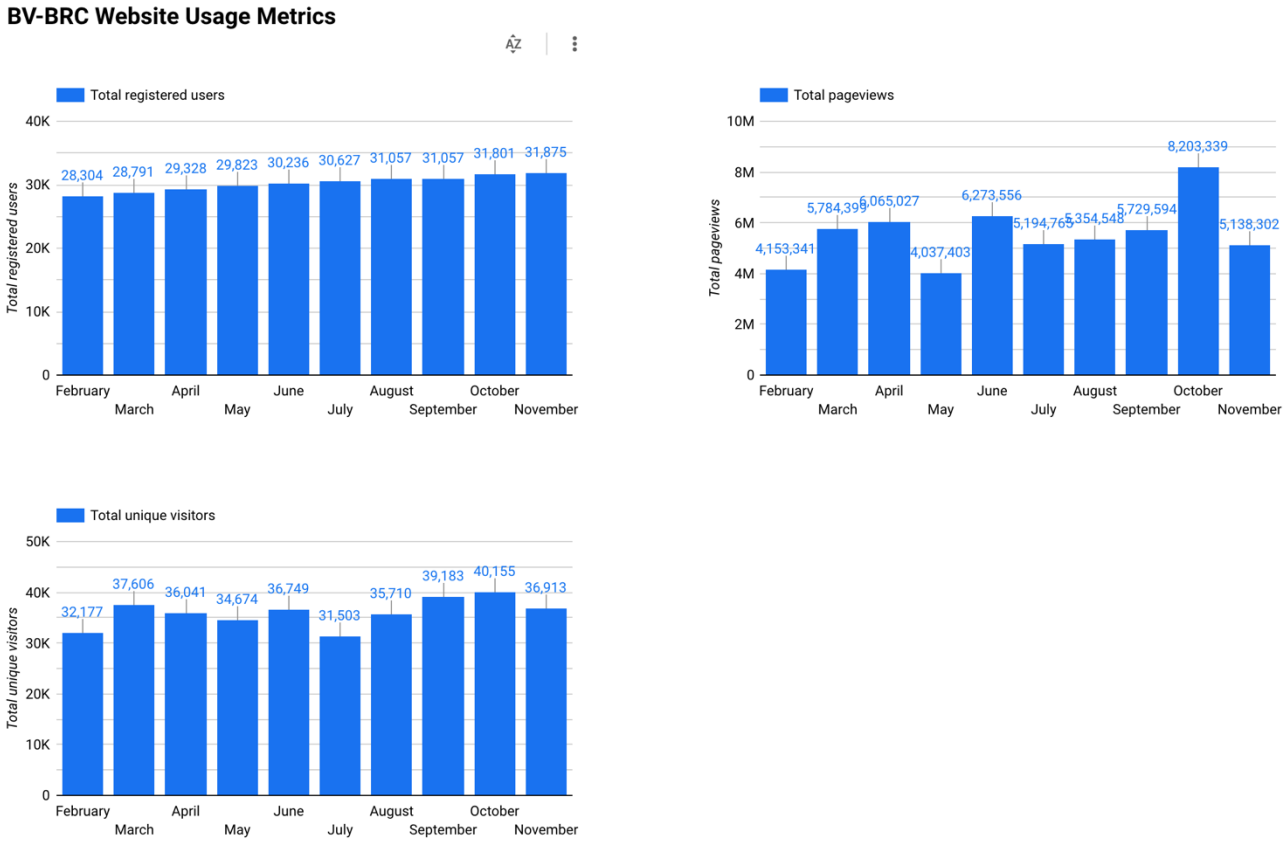

Figure 1. Selected BV-BRC website usage metrics.

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	Domain	Species	Genomes	Page Views
Acinetobacter	Bacteria	690	10,897	1,055
Bacillus	Bacteria	871	6,707	5,270
Bartonella	Bacteria	76	212	985
Borreliella	Bacteria	16	764	11
Brucella	Bacteria	86	1,170	1,926
Burkholderia	Bacteria	315	4,756	489
Campylobacter	Bacteria	268	6,580	1,524
Chlamydia	Bacteria	21	583	436
Clostridium	Bacteria	435	3,172	2,356
Coxiella	Bacteria	12	108	234
Ehrlichia	Bacteria	7	40	481
Escherichia	Bacteria	192	35,288	5,082
Francisella	Bacteria	29	1,066	114
Helicobacter	Bacteria	88	2,440	883
Listeria	Bacteria	43	5,425	367
Mycobacterium	Bacteria	312	30,661	2,540
Pseudomonas	Bacteria	1,842	14,399	4,666
Rickettsia	Bacteria	51	189	873
Salmonella	Bacteria	307	27,728	1,397
Shigella	Bacteria	111	4,143	771
Staphylococcus	Bacteria	578	22,679	2,184
Streptococcus	Bacteria	424	36,010	2,097
Vibrio	Bacteria	385	5,879	1,370
Yersinia	Bacteria	28	1,452	190
Bunyavirales	Virus	611	16,648	1,418
Caliciviridae	Virus	241	63,556	655

Coronaviridae	Virus	1,187	2,483,632	7,140
Filoviridae	Virus	26	4,297	865
Flaviviridae	Virus	465	369,130	7,099
Hepeviridae	Virus	47	20,552	303
Herpesviridae	Virus	860	64,360	4,304
Influenza	Virus	4	5,151	36,588
Paramyxoviridae	Virus	741	85,727	785
Picornaviridae	Virus	1,143	148,717	1,202
Pneumoviridae	Virus	17	45,440	2,098
Poxviridae	Virus	300	11,441	1,136
Reoviridae	Virus	404	136,224	3,319
Rhabdoviridae	Virus	686	37,273	256
SARS-CoV-2	Virus	1	2,699,796	2,316
Togaviridae	Virus	67	14,588	962
SARS-CoV-2 (BV-BRC)	Virus	1	2,492,537	861

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. BRC Website Usage by Data Type (BV-BRC)

Data Type	BRC Domain	Page Views
Taxonomy	PATRIC	26,245
Genome	PATRIC	78,339
Genome sequence	PATRIC	2,380
Feature (Genes/Proteins)	PATRIC	33,830
Specialty gene	PATRIC	7,625
Protein families	PATRIC	3,598
Pathway	PATRIC	8,464
Subsystems	PATRIC	3,805

Transcriptomics	PATRIC	1,499
Interactions	PATRIC	1,139
Phylogeny	PATRIC	2,246
Antibiotic	PATRIC	33
Workspace (User Data)	PATRIC	77,691
Genome	IRD/ViPR	23,032
Gene/Protein	IRD/ViPR	10,224
Strain	IRD/ViPR	4,401
Immune epitopes	IRD/ViPR	917
Ortholog groups	IRD/ViPR	143
Antiviral drugs	IRD/ViPR	265
Host factors	IRD/ViPR	186
Protein structures	IRD/ViPR	291
Protein domains and motifs	IRD/ViPR	88
Plasmids	IRD/ViPR	35
SFVT	IRD/ViPR	98
Surveillance	IRD/ViPR	478
Serology	IRD/ViPR	30
Phenotypes	IRD/ViPR	24
PCR Primers	IRD/ViPR	356
SARS-CoV-2 Variant Tracker	BV-BRC	861

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	PATRIC	442	441
Comprehensive Genome Analysis	PATRIC	1,579	1,442
Differential Expression	PATRIC	6	1
FastqUtils	PATRIC	2,403	1,909
Genome Alignment	PATRIC	145	133
Genome Annotation	PATRIC	20,626	20,448
Genome Assembly	PATRIC	5,664	5,561
Genome Comparison	PATRIC	166	158
Metagenome Binning	PATRIC	174	155
Metagenomic Read Mapping	PATRIC	116	115
RNASeq Analysis	PATRIC	1,791	1,506
Taxonomic Classification	PATRIC	495	495
TnSeq Analysis	PATRIC	10	3
Variation Analysis	PATRIC	287	261
Alignment Viewer	IRD/ViPR	26	26
Antiviral-Resistance-Risk	IRD/ViPR	7	7
BLAST	IRD/ViPR	352	341
Enrichment	IRD/ViPR	3	3
Genotype-Recombination	IRD/ViPR	8	6
H1-Clade Classifier	IRD only	89	89
H1N1-classifier	IRD only	12	12
H5N1-classifier	IRD only	96	95
Ha Numbering	IRD only	120	120

MGC	IRD/ViPR	55	43
MSA	IRD/ViPR	599	584
Mutation-analysis	IRD/ViPR	14	14
Primer3	IRD/ViPR	68	68
Read-seq	IRD/ViPR	52	51
Rva Genotyper	IRD/ViPR	1,137	1,127
Short-seqsearch	IRD/ViPR	37	33
SNP-analysis	IRD/ViPR	535	531
Surveillance-data-mapping	IRD/ViPR	43	43
Tbl-formatter	IRD/ViPR	7	2
Tree	IRD/ViPR	542	525
VIGOR Annotator	IRD/ViPR	54	54
SARS-2 Genome Assembly and Annotation	BV-BRC	35	29

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.

	Number of Citations (YTD)	**Number of Citations (Cumulative)**
Citations to BV-BRC publications	2,247	14,481
Citations to BV-BRC resources	2,442	16,110

1. *YTD Citations to BV-BRC publications was incorrectly overreported last month.*

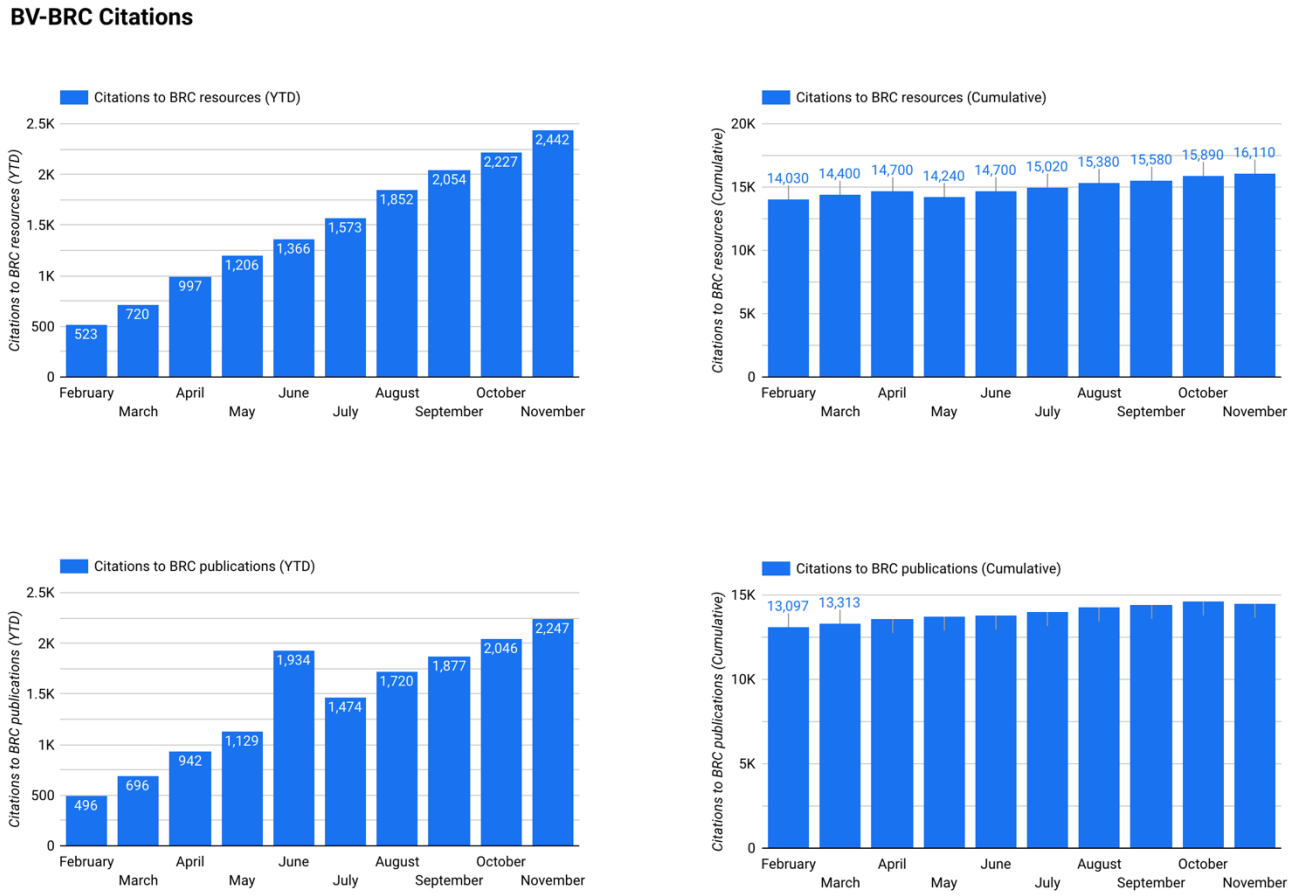

Figure 2. Citations to BV-BRC resources and publications. The June YTD citations to BRC publications was overstated due to an error in the query. This has been resolved for all subsequent months.

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**	**Total**
Total registered users	31,875[1]	11,732	31,875
Total storage used for user data (TB)	176.8	0.58	177.38
User requests: • Request for help • Report bug • Suggest improvement	49 100% 0% 0%	7 85% 14% 0%	56 98% 2% 0%
Webinar/workshop events	0	0	0
Total webinar/workshop participants	0	0	0
Total MOOC registrants (cumulative)	4,260	NA	4,260
Twitter followers	512	366	878
Facebook followers	242	1709	1,951
YouTube followers	297	180	477
YouTube views	997	141	1138
BRC Subreddit subscribers	NA	NA	76
BRC Subreddit viewers	NA	NA	350

1. *The number of total PATRIC registered users had an apparent large increase due to the merger of IRD/ViPR and PATRIC user databases. The Total (BV-BRC) is an accurate count of both resources combined.*

BV-BRC User Requests

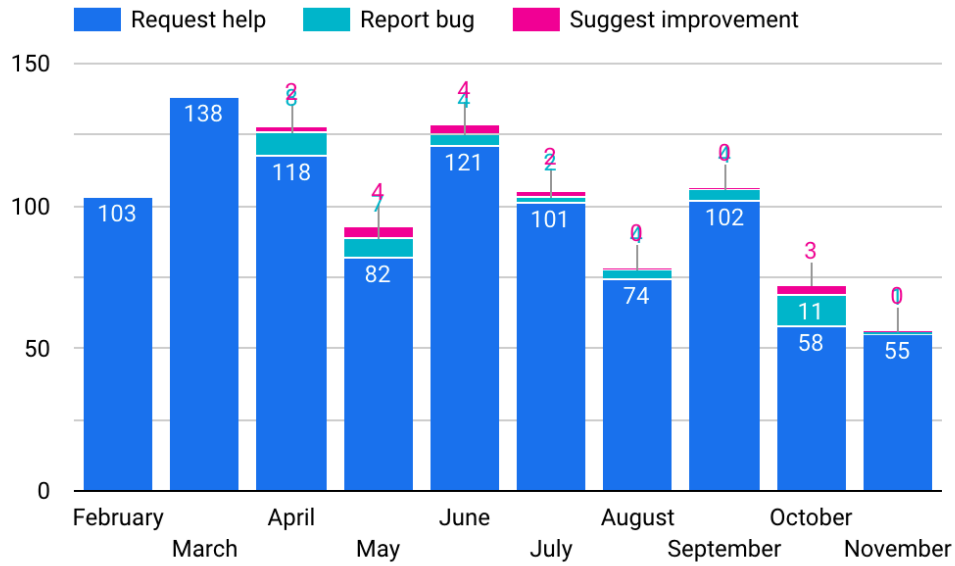

Figure 3. Requests by users, sorted by type.