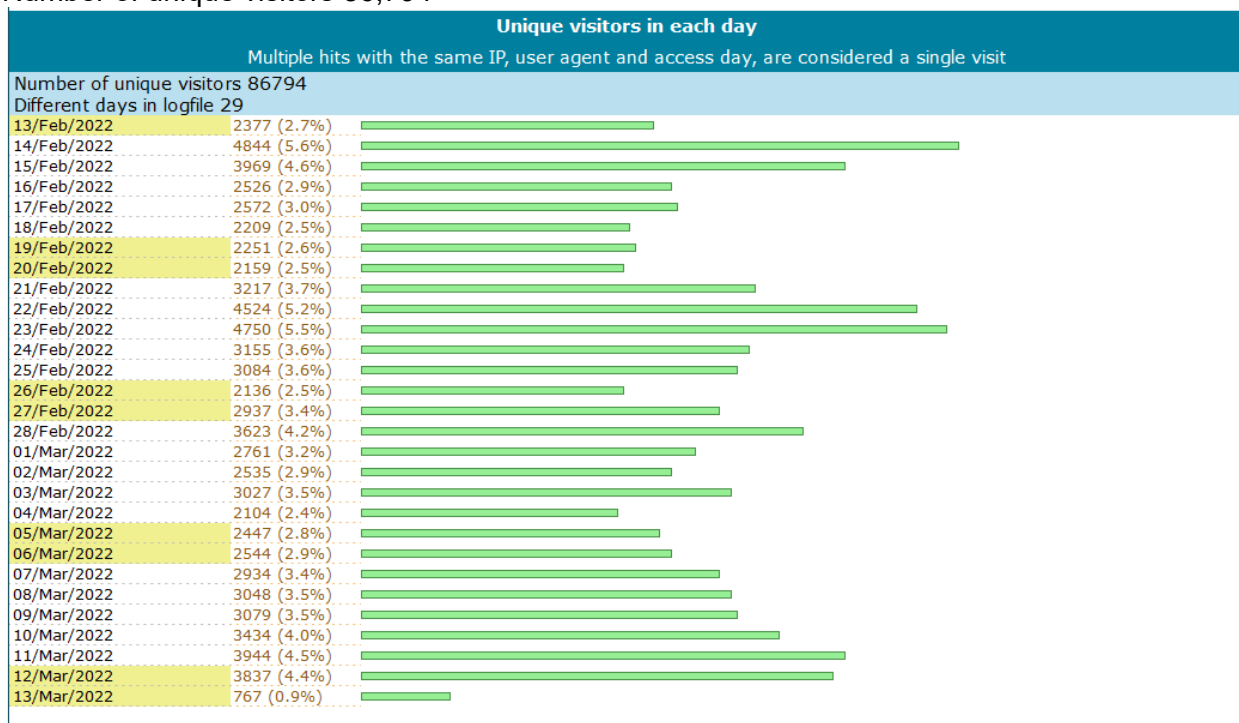# Access and Use Statistics and FAIR metrics on Metabolomics Workbench (MW)

Approach: We used both custom analysis of apache logs and Visitors tool (http://www.hping.org/visitors/) on combined access_log and ssl_access_log. vsftpd log was analyzed separately.

Access and use statistics are based on analysis of the log over about one month. (Feb 13 – Mar 13, 2022).

Overall access to the MW website

Number of unique visitors 86,794

Unique visitors in each day	
Multiple hits with the same IP, user agent and access day, are considered a single visit	
Number of unique visitors 86794	
Different days in logfile 29	
13/Feb/2022	2377 (2.7%)
14/Feb/2022	4844 (5.6%)
15/Feb/2022	3969 (4.6%)
16/Feb/2022	2526 (2.9%)
17/Feb/2022	2572 (3.0%)
18/Feb/2022	2209 (2.5%)
19/Feb/2022	2251 (2.6%)
20/Feb/2022	2159 (2.5%)
21/Feb/2022	3217 (3.7%)
22/Feb/2022	4524 (5.2%)
23/Feb/2022	4750 (5.5%)
24/Feb/2022	3155 (3.6%)
25/Feb/2022	3084 (3.6%)
26/Feb/2022	2136 (2.5%)
27/Feb/2022	2937 (3.4%)
28/Feb/2022	3623 (4.2%)
01/Mar/2022	2761 (3.2%)
02/Mar/2022	2535 (2.9%)
03/Mar/2022	3027 (3.5%)
04/Mar/2022	2104 (2.4%)
05/Mar/2022	2447 (2.8%)
06/Mar/2022	2544 (2.9%)
07/Mar/2022	2934 (3.4%)
08/Mar/2022	3048 (3.5%)
09/Mar/2022	3079 (3.5%)
10/Mar/2022	3434 (4.0%)
11/Mar/2022	3944 (4.5%)
12/Mar/2022	3837 (4.4%)
13/Mar/2022	767 (0.9%)

Resource types

CATEGORY	Feb 13th to March 13th (28 days)	hits per day
Studies	1196326	42726
Structures	1160259	41438
Analysis	764103	27289
REST	420859	15031
RefMet	191062	6824
MS	59576	2128
Protocols	59050	2109
Genes/proteins	46760	1670
Search	30425	1087

Homepage		23616	843
About		12958	463
Unclassified		1021	36

Top page hits:

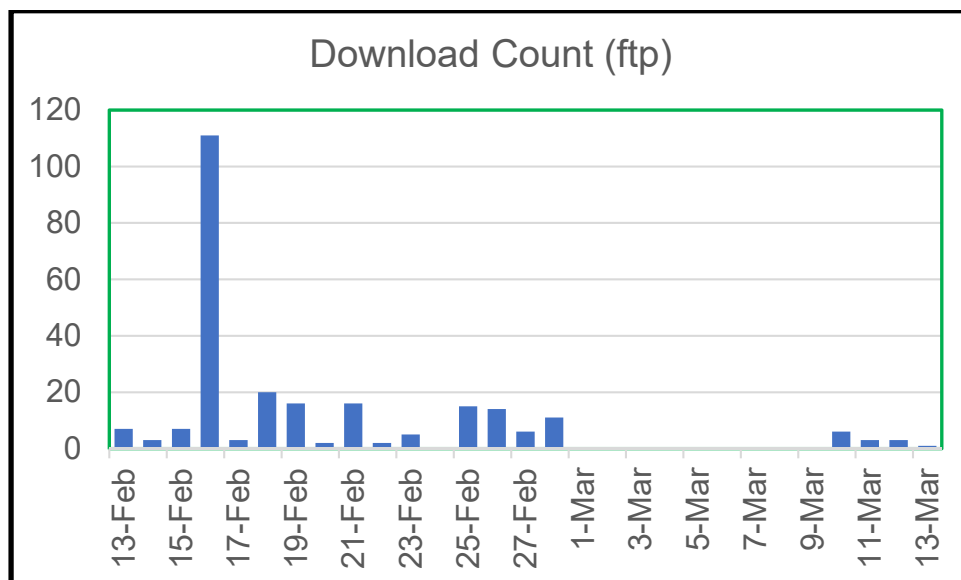| Type of page	End-point	Count
Studies	/data/DRCCMetadata.php	914403
Statistical analysis	/data/metabolite_id_show.php	499372
Statistical analysis	/data/show_metabolite_data_by_factors.php	369753
REFMET	/rest/compound	248426
Statistical analysis	/data/ttest1.php	170709
Structure data	/data/StructureData.php	167719
REFMET	/rest/refmet	132267
Statistical analysis	/data/ttest1_grid.php	130274
REFMET	/databases/refmet/refmet.php	116999
Statistical analysis	/data/anova1.php	111341
Statistical analysis	/data/zscore_calc.php	97147
Pathways	/data/show_h_pathway_metabolites_reactome.php	81689
Study search	/data/show_studies_by_pubchem.php	55645
Pathways	/data/show_h_pathway_metabolites2.php	45631
Proteome	/databases/proteome/MGP_table.php	44257
Massbank	/data/massbank.php	43359
Studies	/rest/study	35293
Study summary	/data/DRCCStudySummary.php	31337
Pathways	/data/show_h_pathway_metabolites.php	29338
Protocols	/protocols/protocoldetails.php	28466

OS and browsers
(unknown may include REST-based visits or calls from other automated scripts)
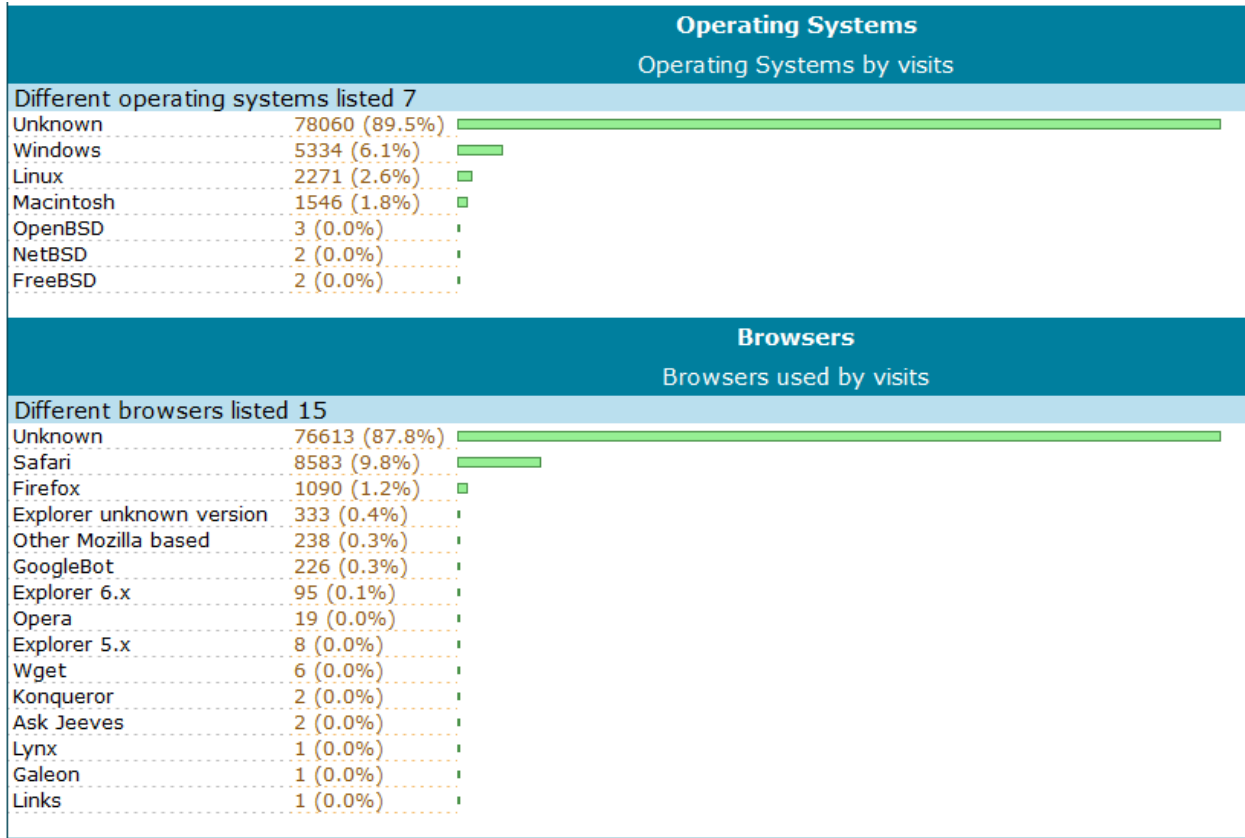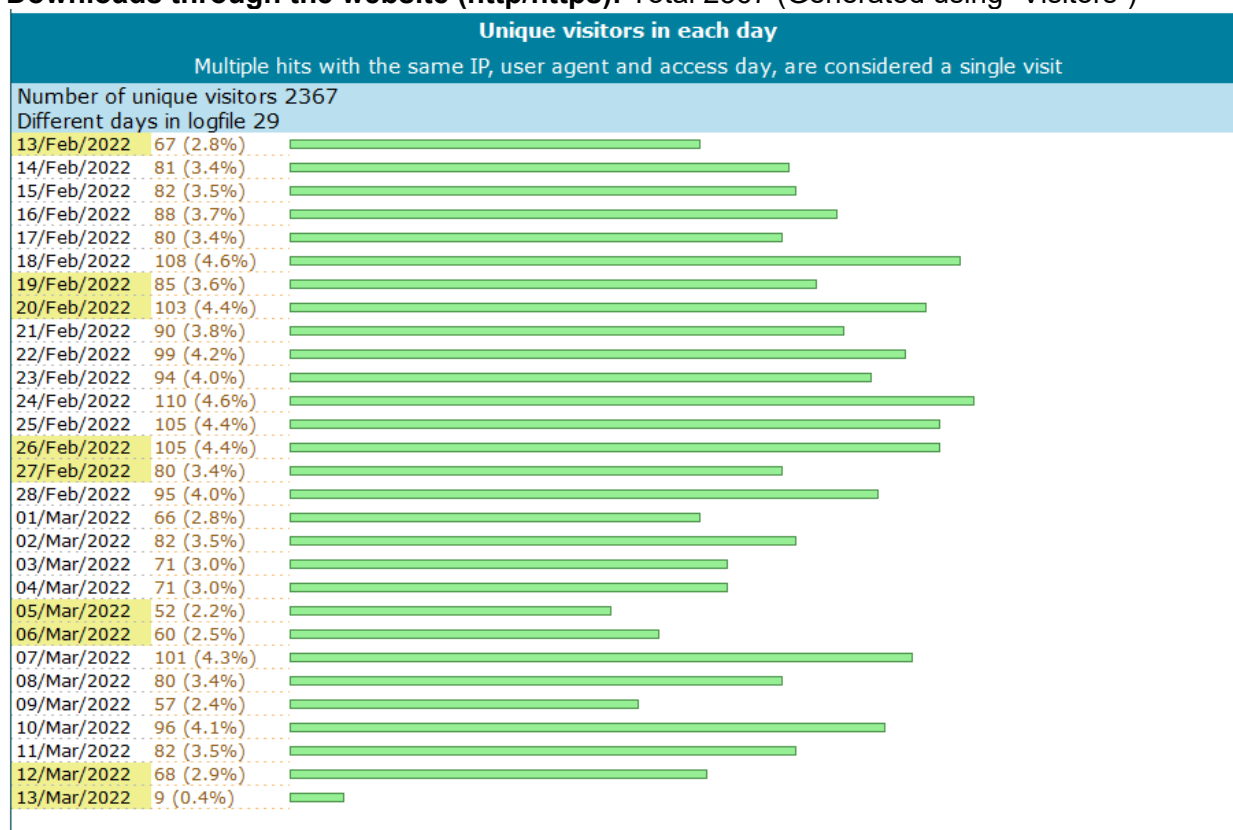
Download statistics

Amount of data downloaded: about 22 TB

FTP-based downloads: Total 251

Download Count (ftp)

Downloads through the website (http/https): Total 2367 (Generated using "Visitors")

Unique visitors in each day		
Multiple hits with the same IP, user agent and access day, are considered a single visit		
Number of unique visitors 2367		
Different days in logfile 29		
13/Feb/2022	67 (2.8%)	
14/Feb/2022	81 (3.4%)	
15/Feb/2022	82 (3.5%)	
16/Feb/2022	88 (3.7%)	
17/Feb/2022	80 (3.4%)	
18/Feb/2022	108 (4.6%)	
19/Feb/2022	85 (3.6%)	
20/Feb/2022	103 (4.4%)	
21/Feb/2022	90 (3.8%)	
22/Feb/2022	99 (4.2%)	
23/Feb/2022	94 (4.0%)	
24/Feb/2022	110 (4.6%)	
25/Feb/2022	105 (4.4%)	
26/Feb/2022	105 (4.4%)	
27/Feb/2022	80 (3.4%)	
28/Feb/2022	95 (4.0%)	
01/Mar/2022	66 (2.8%)	
02/Mar/2022	82 (3.5%)	
03/Mar/2022	71 (3.0%)	
04/Mar/2022	71 (3.0%)	
05/Mar/2022	52 (2.2%)	
06/Mar/2022	60 (2.5%)	
07/Mar/2022	101 (4.3%)	
08/Mar/2022	80 (3.4%)	
09/Mar/2022	57 (2.4%)	
10/Mar/2022	96 (4.1%)	
11/Mar/2022	82 (3.5%)	
12/Mar/2022	68 (2.9%)	
13/Mar/2022	9 (0.4%)	

Top downloads requested:

Requested pages		
Page requests ordered by hits		
Different pages requested 1375		
1)	859	/studydownload/ST001430_AN002391_Results.txt
2)	207	/studydownload/ST000923_AN001513_Results.txt
3)	206	/studydownload/ST000758_AN001190_Results.txt
4)	154	/studydownload/ST000879_AN001429_Results.txt
5)	135	/studydownload/ST000879_AN001432_Results.txt
6)	81	/studydownload/ST001443_AN002414_Results.txt
7)	74	/studydownload/ST001430_AN002392_Results.txt
8)	53	/studydownload/ST001421.zip
9)	49	/studydownload/ST001394.7z
10)	41	/studydownload/ST002044_AN003325_Results.txt
11)	36	/studydownload/ST000063.zip
12)	27	/studydownload/ST000306.zip
13)	20	/studydownload/ST000763_AN001202_Results.txt
14)	20	/studydownload/ST001000_AN001879_Results.txt
15)	17	/studydownload/ST001000_AN001880_Results.txt
16)	14	/studydownload/ST000918_AN001504_Results.txt
17)	14	/studydownload/ST001764_HILIC_K562_MTX.zip
18)	12	/studydownload/ST000050.zip
19)	12	/studydownload/ST000315.zip
20)	12	/studydownload/ST001386_Normalized_UnIdentified_Primary_Metabolites.txt

Access statistics of our GitHub page for Jupyter notebooks repository (covers 1/1/2022 – 03/30/2022)
https://github.com/metabolomicsworkbench

- How many git clones

Unique: 5 [Information about the method GitHub uses to records access is not clear.], Total: 7

- Unique hits by file path on GitHub

Unique	path
56	/metabolomicsworkbench/jupyter-notebooks
2	/metabolomicsworkbench/jupyter-notebooks/blob/6db27e08aa59166f9b756723dc6b7f74e117afdb/MWPerformCluster
2	/metabolomicsworkbench/jupyter-notebooks/blob/master/environment.yml
6	/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformClusteredHeatMapAnalysis.ipynb
10	/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformDataNormalization.ipynb
10	/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformRandomForestAnalysis.ipynb
8	/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformRelativeLogAbundanceAnalysis.ipynb
18	/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformVolcanoPlotAnalysis.ipynb
13	/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPlotNamedMetabolitesResultsExample.ipynb
3	/metabolomicsworkbench/jupyter-notebooks/blob/master/MWRestAPIGeneDataExample.ipynb
2	/metabolomicsworkbench/jupyter-notebooks/blob/master/MWRestAPINamedMetabolitesResultsExample.ipynb

- Who is referring traffic to the GitHub site

Unique	referrer	count
17	github.com	31
29	Google	52
24	metabolomicsworkbench.org	52

- How many people are viewing the GitHub repository (unique and non-unique)

Unique: 56, Total: 107

The number of accesses to the Jupyter/Binder entry page at:
https://www.metabolomicsworkbench.org/data/AnalyzeUsingJupyterNotebooks.php

The number of hits to the entry page last year was pretty low- about 20 per week. This doesn't necessarily mean that users are clicking on the external Binder/Github links on this page, though.

FAIRShake metrics

Precalculated stats on MW: https://fairshake.cloud/project/85/assessments/

Project Assessments (6697)

Target	Rubric		Globally unique identifier	Persistent identifier	Machine-readable metadata	Standardized metadata	Resource identifier in metadata	Resource discovery through web search	Open, Free, Standardized Access protocol	Protocol to access restricted content	Persistence of resource and metadata
Fatb Induction Experiment (FatBIE)	FAIR metrics by fairmetrics.org		no (0.00)	no (0.00)	yes (1.00)	yes (1.00)	yes (1.00)	yes (1.00)	yes (1.00)		no (0.00)
Intestinal Samples II pre/post transplantation	FAIR metrics by fairmetrics.org		no (0.00)	no (0.00)	yes (1.00)	yes (1.00)	yes (1.00)	yes (1.00)	yes (1.00)		no (0.00)
Metabolomic analysis of mouse embryonic fibroblasts, embryonic stem cells, and induced pluripotent stem cells	FAIR metrics by fairmetrics.org		no (0.00)	no (0.00)	yes (1.00)	yes (1.00)	yes (1.00)	yes (1.00)	yes (1.00)		no (0.00)

Apparently, the FAIRshake tool that generated the above statistics is not fully mature and requires manual curation. MW uses document object identifiers (DOI) for the projects, which is included (referred to as Persistent Identifier) in the metadata submitted to the CFDE portal. Through the DOI, one can access all the publicly available data on the MW website.

https://fairshake.cloud/digital_object/6578/assessments/
Tags: NIHcommonfund

Digital Object Assessments (1)

Rubric	Project		The repository provides contact information for staff to enable users with questions or suggestions to interact with repository experts.	Tools that can be used to analyze each dataset are listed on the corresponding dataset pages.	The repository maintains licenses to manage data access and use.	The repository hosts data and metadata according to a set of defined criteria to ensure that the resources provided are consistent with the intent of the repository.	The repository provides documentation for each resource to permit its complete and accurate citation.	A description of the methods used to acquire the data is provided.	Version information is provided for each resource, where available.	The structure of the repository permits efficient discovery of data and metadata by end users.	The repository uses a standardized protocol to permit access by users.
The FAIRshake repository rubric	NIH Data Sharing Repositories		yes (1.00)	yes (1.00)	no (0.00)	yes (1.00)	yes (1.00)	yes (1.00)	yes (1.00)	yes (1.00)	yes (1.00)

The above tables lists "no" for license, but we have stated the terms of use.
In the future, we will decide the license term (CC BY-NC, CC BY-NC-SA or something allowing commercialization of derivative work) and update the "Terms of use" page.

METABOLOMICS
WORKBENCH

Terms of Use

The Metabolomics Workbench is provided by the NIH Common Fund's National Metabolomics Data Repository(NMDR) on an "as is" basis, without warranty or representation of any kind, express or implied. The content of the Metabolomics Workbench is protected by international copyright, trademark and other laws. You may download articles and web pages from this site for your personal, non-commercial use only, provided that you keep intact all authorship, copyright and other proprietary notices. If you use the Metabolomics Workbench, you accept these terms. The NMDR reserves the right to modify these terms at any time.