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 (July 2023, unless indicated otherwise). Access by web crawlers, bots, etc., has been excluded from the beginning of the analysis (see IP Exclusion List 1 at the end).

Overall access to the MW website

Number of unique visitors: 222,455

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 222455
Different days in logfile 31

01/Jul/2023	6315 (2.8%)	
02/Jul/2023	7619 (3.4%)	
03/Jul/2023	8480 (3.8%)	
04/Jul/2023	7834 (3.5%)	
05/Jul/2023	6913 (3.1%)	
06/Jul/2023	6547 (2.9%)	
07/Jul/2023	6468 (2.9%)	
08/Jul/2023	5716 (2.6%)	
09/Jul/2023	7266 (3.3%)	
10/Jul/2023	8201 (3.7%)	
11/Jul/2023	10521 (4.7%)	
12/Jul/2023	7621 (3.4%)	
13/Jul/2023	7381 (3.3%)	
14/Jul/2023	6347 (2.9%)	
15/Jul/2023	6224 (2.8%)	
16/Jul/2023	6246 (2.8%)	
17/Jul/2023	7560 (3.4%)	
18/Jul/2023	8119 (3.6%)	
19/Jul/2023	6768 (3.0%)	
20/Jul/2023	8008 (3.6%)	
21/Jul/2023	8175 (3.7%)	
22/Jul/2023	6746 (3.0%)	
23/Jul/2023	7383 (3.3%)	
24/Jul/2023	7262 (3.3%)	
25/Jul/2023	8560 (3.8%)	
26/Jul/2023	6200 (2.8%)	
27/Jul/2023	7218 (3.2%)	
28/Jul/2023	7493 (3.4%)	
29/Jul/2023	5601 (2.5%)	
30/Jul/2023	6032 (2.7%)	
31/Jul/2023	5631 (2.5%)	

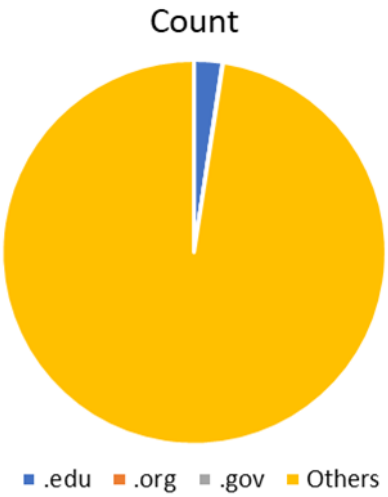
Country-based access statistics

From the IP addresses, hostname and their geographical location (country) is identified and summarized.

Top 30 counties:

Country_short	Country_long	Count
US	United States of America	3897663
FR	France	1066706
GB	United Kingdom of Great Britain and Northern Ireland	427693
SG	Singapore	384423
GR	Greece	204293
IN	India	155737
FI	Finland	114532
DE	Germany	112511
CA	Canada	105078
-	-	91400
IT	Italy	27836
CN	China	21471
UA	Ukraine	20344
SE	Sweden	14866
ES	Spain	11604
CH	Switzerland	10971
BR	Brazil	10364
JP	Japan	9976
VN	Viet Nam	6394
KR	Korea (Republic of)	4099
HK	Hong Kong	3894
HU	Hungary	3856
NO	Norway	3840
NL	Netherlands	3600
RO	Romania	3239
TR	Turkey	2946
RU	Russian Federation	2822
BE	Belgium	2638
TW	Taiwan (Province of China)	1503
PL	Poland	1338

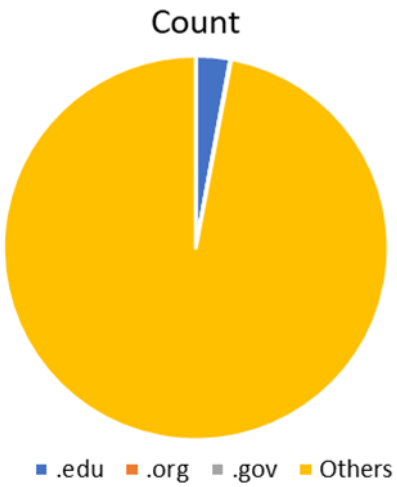
Category	Count	Percent
.edu	160144	2.375
.org	2884	0.043
.gov	1330	0.020
Others	6578797	97.563



After excluding additional IP addresses (in R script) with ‘bot’, ‘spider’, etc., in their hostnames (see IP Exclusion List 2 at the end): top 30 countries:

Country_short	Country_long	Count
US	United States of America	3766712
GB	United Kingdom of Great Britain and Northern Ireland	425960
FR	France	218708
SG	Singapore	217179
GR	Greece	204293
IN	India	155737
CA	Canada	105078
FI	Finland	104495
DE	Germany	98891
-	-	91400
IT	Italy	27822
CN	China	21428
UA	Ukraine	20344
SE	Sweden	14866
ES	Spain	11604
CH	Switzerland	10971
BR	Brazil	10363
JP	Japan	9976
VN	Viet Nam	6394
KR	Korea (Republic of)	4099
HK	Hong Kong	3894
HU	Hungary	3856
NO	Norway	3840
NL	Netherlands	3600
RO	Romania	3239
TR	Turkey	2946
BE	Belgium	2638
TW	Taiwan (Province of China)	1503
RU	Russian Federation	1501
PL	Poland	1338

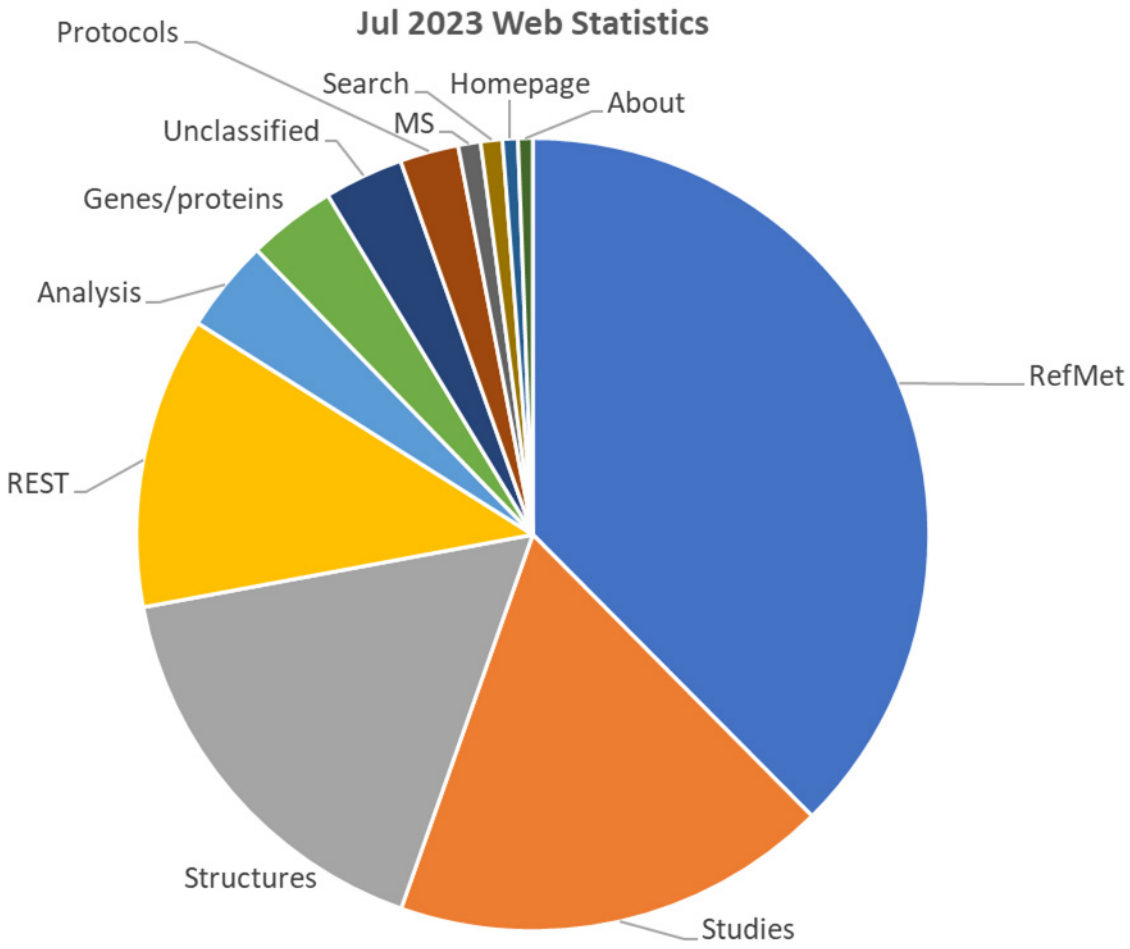
Category	Count	Percent
.edu	160144	2.875
.org	2883	0.052
.gov	1330	0.024
Others	5405835	97.049



Resource types (Major categories)

Category	Jul 2023 Web Statistics (31 days)	Hits per day
RefMet	2532756	81702
Studies	1202249	38782
Structures	1124456	36273
REST	802116	25875
Analysis	254930	8224
Genes/proteins	243182	7845
Unclassified	218114	7036

Protocols	161477	5209
MS	61622	1988
Search	59478	1919
Homepage	42358	1366
About	40320	1301



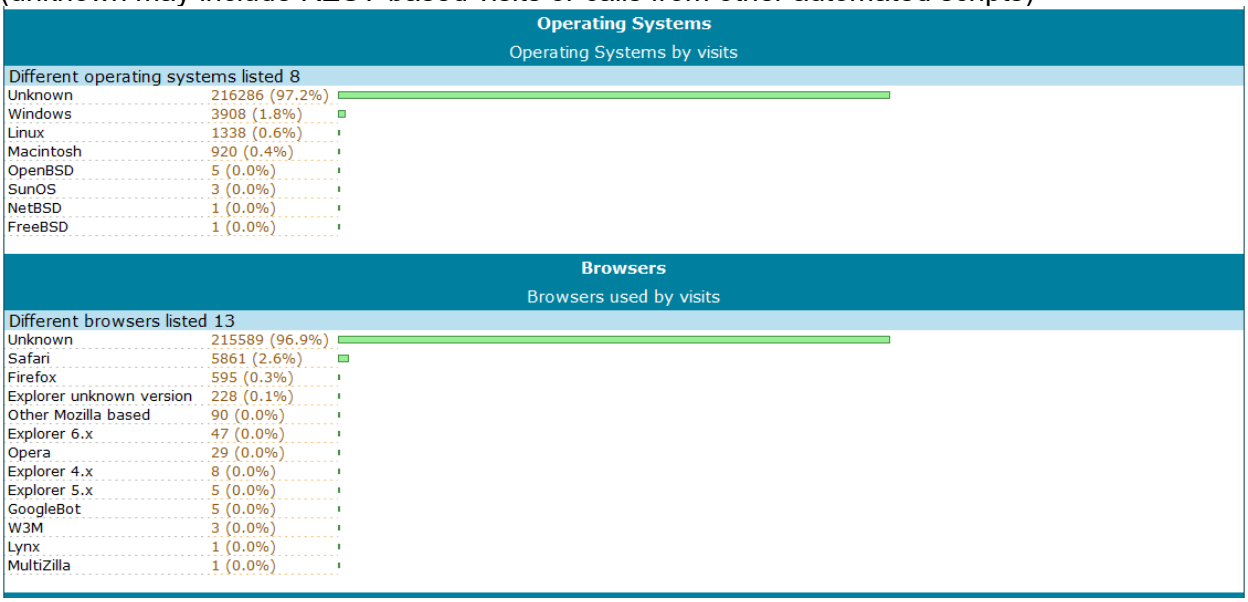
Top page hits over July 2023:

Type of page	End-point	Count
REFMET	/databases/refmet/name_to_refmet_new_min.php	1492779
REFMET	/databases/refmet/refmet.php	906576
Studies	/data/DRCCMetadata.php	734002
Studies	/rest/study	475219
REFMET	/rest/refmet	306953
General	/data/png_display_inline2.php	246499
Statistical analysis	/data/metabolite_id_show.php	231472
Structure data	/data/StructureData.php	209120
Statistical analysis	/data/show_metabolite_data_by_factors.php	196407
Metabolite-Gene	/databases/proteome/MGP_table.php	195394

REFMET	/databases/refmet/core_browse.php	190406
REFMET	/databases/refmet/abbrev_generic.php	173899
Study summary	/data/DRCCStudySummary.php	141957
Protocols	/protocols/protocoldetails.php	88748
Files	/data/file_extract.php	71894
REFMET	/databases/refmet/refmet_details.php	62039
Files	/data/file_extract_7z.php	57405
Search	/search/sitesearch.php	48549
Protocols	/protocols/general.php	47029
Massbank	/data/massbank.php	46131
General	/homepage	42358
Metabolite-Gene	/databases/proteome/MGP_detail.php	35430
Studies	/data/study_textformat_view.php	29801
Pathways	/data/show_h_pathway_metabolites2.php	22515
Studies	/data/study_textformat_list.php	22381
Metabolites	/data/show_metabolites_by_study.php	20027
Study search	/data/show_studies_by_pubchem.php	16600
Search	/data/s3.php	15502
Statistical analysis	/data/stats_toolbox.php	14643
Protocols	/protocols/studyspecific.php	14460

OS and browsers

(unknown may include REST-based visits or calls from other automated scripts)



Operating Systems

Operating Systems by visits

Different operating systems listed 8	
Unknown	216286 (97.2%)
Windows	3908 (1.8%)
Linux	1338 (0.6%)
Macintosh	920 (0.4%)
OpenBSD	5 (0.0%)
SunOS	3 (0.0%)
NetBSD	1 (0.0%)
FreeBSD	1 (0.0%)

Browsers

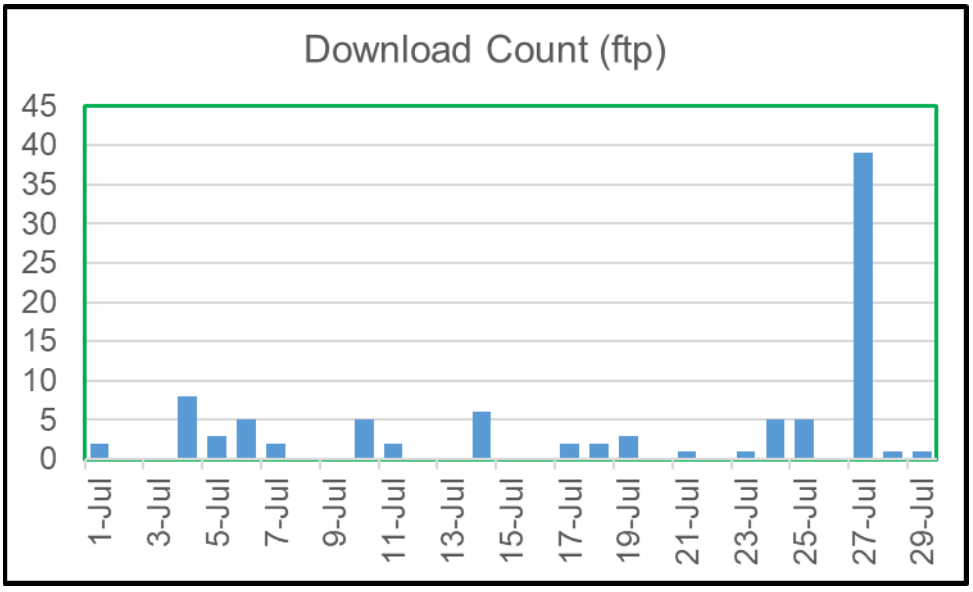
Browsers used by visits

Different browsers listed 13	
Unknown	215589 (96.9%)
Safari	5861 (2.6%)
Firefox	595 (0.3%)
Explorer unknown version	228 (0.1%)
Other Mozilla based	90 (0.0%)
Explorer 6.x	47 (0.0%)
Opera	29 (0.0%)
Explorer 4.x	8 (0.0%)
Explorer 5.x	5 (0.0%)
GoogleBot	5 (0.0%)
W3M	3 (0.0%)
Lynx	1 (0.0%)
MultiZilla	1 (0.0%)

Download statistics

Amount of total data downloaded (http + ftp): about 32 TB

FTP-based downloads: Total 93



Downloads through the website (http/https): Total 880 (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 880
Different days in logfile 31

01/Jul/2023	41 (4.7%)	
02/Jul/2023	42 (4.8%)	
03/Jul/2023	38 (4.3%)	
04/Jul/2023	37 (4.2%)	
05/Jul/2023	42 (4.8%)	
06/Jul/2023	45 (5.1%)	
07/Jul/2023	32 (3.6%)	
08/Jul/2023	11 (1.2%)	
09/Jul/2023	23 (2.6%)	
10/Jul/2023	25 (2.8%)	
11/Jul/2023	35 (4.0%)	
12/Jul/2023	24 (2.7%)	
13/Jul/2023	26 (3.0%)	
14/Jul/2023	22 (2.5%)	
15/Jul/2023	18 (2.0%)	
16/Jul/2023	24 (2.7%)	
17/Jul/2023	37 (4.2%)	
18/Jul/2023	35 (4.0%)	
19/Jul/2023	26 (3.0%)	
20/Jul/2023	15 (1.7%)	
21/Jul/2023	33 (3.8%)	
22/Jul/2023	20 (2.3%)	
23/Jul/2023	36 (4.1%)	
24/Jul/2023	34 (3.9%)	
25/Jul/2023	27 (3.1%)	
26/Jul/2023	18 (2.0%)	
27/Jul/2023	31 (3.5%)	
28/Jul/2023	20 (2.3%)	
29/Jul/2023	8 (0.9%)	
30/Jul/2023	19 (2.2%)	
31/Jul/2023	36 (4.1%)	

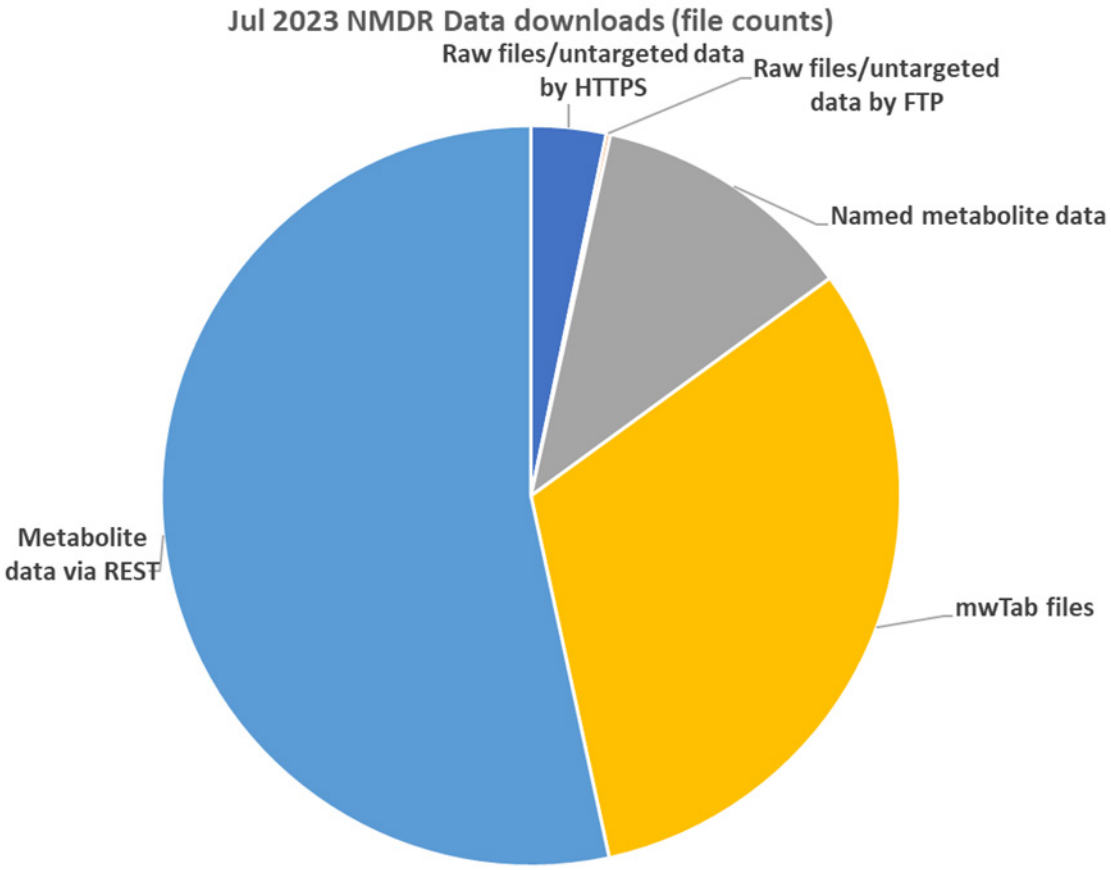
Top 20 downloads requested:

Requested pages		
Page requests ordered by hits		
Different pages requested 913		
1)	48	/studydownload/ST000923_AN001514_Results.txt
2)	29	/studydownload/ST000923_AN001513_Results.txt
3)	24	/studydownload/ST001192_AN001987_Results.txt
4)	23	/studydownload/ST000919_AN001506_Results.txt
5)	13	/studydownload/ST001335_AN002224_Results.txt
6)	12	/studydownload/ST001000_RawFile_SampleID_mapping.csv
7)	12	/studydownload/ST001335_AN002225_Results.txt
8)	11	/studydownload/ST000604_AN000925_Results.txt
9)	10	/studydownload/ST001000_AN001629_Results.txt
10)	10	/studydownload/ST001430_AN002392_Results.txt
11)	10	/studydownload/ST002233_HZV029_HILICpos.zip
12)	10	/studydownload/ST002722_Rawdata.7z
13)	10	/studydownload/ST000604_AN000924_Results.txt
14)	9	/studydownload/ST002764_POS_rawdata.zip
15)	9	/studydownload/ST002112_AN004118_Results.txt
16)	9	/studydownload/ST001914_OA_Fecal_Metabolomics_LCMS_RawData.7z
17)	9	/studydownload/ST000913_AN001483_Results.txt
18)	9	/studydownload/ST002202_AN003605_Results.txt
19)	9	/studydownload/ST000954_AN001564_Results.txt
20)	9	/studydownload/ST002112_AN004117_Results.txt

Mode of download (Jul 2023)

Unlike the filtering for other statistics, these are based on the entire log for July 2023 (without excluding bots, spiders, etc).

Type	Jul 2023 NMDR Data downloads (file counts)
Raw files/untargeted data by HTTPS	3362
Raw files/untargeted data by FTP	190
Named metabolite data	11846
mwTab files	32455
Metabolite data via REST	54833



Access statistics of our GitHub page for Jupyter notebooks repository (covers 3/2/2023 – 08/19/2023)

https://github.com/metabolomicsworkbench/jupyter-notebooks

	Unique	Total Count
Views	174	524
Clones	57	70
Referers		
github.com	8	32
Google	107	156
metabolomicsworkbench.org	99	496
Views		
/metabolomicsworkbench/jupyter-notebooks	63	99
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformClusteredHeatMapAnalysis.ipynb	20	41
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformDataNormalization.ipynb	81	140
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformLinearDiscriminantAnalysis.ipynb	3	5
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformPartialLeastSquaresDiscriminantAnalysis.ipynb	7	10
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformPrincipalComponentAnalysis.ipynb	19	25
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformRandomForestAnalysis.ipynb	9	16
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformRelativeLogAbundanceAnalysis.ipynb	7	11
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPerformVolcanoPlotAnalysis.ipynb	12	19
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWPlotNamedMetabolitesResultsExample.ipynb	32	47
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWRestAPIExactMassDataExample.ipynb	18	30
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWRestAPIGeneDataExample.ipynb	10	18
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWRestAPINamedMetabolitesResultsExample.ipynb	4	12
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWRestAPIProteinDataExample.ipynb	33	46
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWRestAPIRefMetDataExample.ipynb	11	90
/metabolomicsworkbench/jupyter-notebooks/blob/master/MWRestAPIStudyDataExample.ipynb	18	36

https://github.com/metabolomicsworkbench/MetENP

	Unique	Total Count
Views	150	756
Clones	41	47
Referers		
biorxiv.org	14	31
github.com	8	42
Google	78	265
metabolomicsworkbench.org	78	265
Views		
/metabolomicsworkbench/MetENP	169	294
/metabolomicsworkbench/MetENP/blob/main/DESCRIPTION	6	18
/metabolomicsworkbench/MetENP/blob/main/README.md	9	21
/metabolomicsworkbench/MetENP/blob/main/vignettes/Case%20study%201_%20Study%20ST000915_20220619.html	3	6
/metabolomicsworkbench/MetENP/blob/main/vignettes/Case%20study%201_%20Study%20ST000915.html	5	11
/metabolomicsworkbench/MetENP/blob/main/vignettes/MetENP_ST000084.ipynb	17	58
/metabolomicsworkbench/MetENP/blob/main/vignettes/MetENP_ST000897.ipynb	24	104
/metabolomicsworkbench/MetENP/blob/main/vignettes/MetENP_ST000915.ipynb	9	45
/metabolomicsworkbench/MetENP/blob/main/vignettes/MetENP_ST001256.ipynb	6	40
/metabolomicsworkbench/MetENP/blob/main/vignettes/MetENP_ST001742.ipynb	2	12
/metabolomicsworkbench/MetENP/blob/main/vignettes/MetENP_ST002085.ipynb	2	16
/metabolomicsworkbench/MetENP/blob/main/vignettes/MetENP_vignette_Jupyter_notebook.ipynb	2	2
/metabolomicsworkbench/MetENP/blob/main/vignettes/MetENP_vignette.rmd	4	10
/metabolomicsworkbench/MetENP/tree/main/vignettes	28	67

https://github.com/metabolomicsworkbench/MetENPAppyter

	Unique	Total Count
Views	5	8
Clones	18	25
Referers		
Google	5	12
Views		
/metabolomicsworkbench/MetENPAppyter	6	10
/metabolomicsworkbench/MetENPAppyter/blob/main/MetENP_R/MetENP/R/convert_refmet.R	3	6

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 is 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)

Assessment			Metrics								
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)

Assessment			Metrics								
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.

IP Exclusion List 1

As php code:

```
$blocked_ip=array(
'132.249', '3.91', '5.9', '34.238', '35.174', '40.77', '46.229', '46.4', '54.209', '66.249', '69.3', '78.46',
'91.137', '92.22', '95.216', '95.91', '106.120', '136.243', '144.76', '148.251', '157.55', '162.21',
'173.234', '178.255', '178.63', '180.76', '185.25', '192.151', '207.46', '213.174'
);
$blocked_ip3 = array(
'100.26.127', '103.131.71', '104.245.145', '110.249.201', '110.249.202', '110.93.150',
'111.225.148', '111.225.149', '114.111.32', '114.119.132', '114.119.133', '114.119.135',
'114.119.137', '114.119.140', '114.119.145', '114.119.149', '114.119.152', '114.119.153',
'114.119.154', '114.119.155', '116.179.32', '116.179.37', '118.184.177', '123.125.109',
'123.183.224', '125.209.235', '128.127.105', '13.66.139', '135.181.137', '135.181.138',
'135.181.62', '141.8.142', '144.217.135', '147.92.153', '149.154.161', '149.155.131', '149.56.150',
'149.56.160', '154.51.131', '154.54.249', '157.90.177', '17.121.112', '17.121.113', '17.121.114',
'17.121.115', '184.75.211', '185.101.32', '185.138.241', '185.191.171', '185.54.230', '191.96.106',
'194.187.171', '194.9.191', '195.201.106', '198.134.108', '198.134.109', '198.251.73',
'198.98.183', '199.47.82', '204.15.110', '207.241.229', '207.241.233', '211.249.46', '213.180.203',
'216.244.66', '220.181.108', '31.3.152', '31.3.153', '37.46.121', '49.7.20', '49.7.21', '5.102.173',
'5.133.192', '5.255.253', '5.45.207', '5.62.16', '5.62.20', '5.62.41', '5.62.43', '5.62.56', '5.62.57',
'5.62.58', '5.62.59', '5.62.60', '5.62.61', '5.62.62', '5.62.63', '50.21.188', '54.161.41', '58.250.125',
'61.135.159', '65.108.103', '65.108.6', '65.108.99', '65.21.231', '69.160.160', '74.208.2',
'76.164.224', '77.234.44', '77.75.73', '77.75.76', '77.75.77', '77.75.78', '77.75.79', '77.88.5',
'79.142.76', '81.17.57', '85.208.98', '85.31.186', '87.250.224', '91.143.80', '91.219.212',
'91.242.162', '93.158.161', '95.108.213', '95.142.121', '95.142.127', '95.87.154'
);
```

IP Exclusion List 2

As php code:

$blocked_ip is as above in List 1.

$blocked_ip3 = array('100.26.127', '103.131.71', '104.245.145', '110.249.201', '110.249.202',
'110.93.150', '111.225.148', '111.225.149', '114.111.32', '114.119.128', '114.119.129',
'114.119.130', '114.119.131', '114.119.132', '114.119.133', '114.119.134', '114.119.135',
'114.119.136', '114.119.137', '114.119.138', '114.119.139', '114.119.140', '114.119.141',
'114.119.142', '114.119.143', '114.119.144', '114.119.145', '114.119.146', '114.119.147',
'114.119.148', '114.119.149', '114.119.150', '114.119.151', '114.119.152', '114.119.153',
'114.119.154', '114.119.155', '114.119.156', '114.119.157', '114.119.158', '114.119.159',
'114.119.160', '114.119.161', '114.119.162', '114.119.163', '114.119.165', '114.119.166',
'114.119.167', '116.179.32', '116.179.37', '118.184.177', '123.125.109', '123.126.68',
'123.183.224', '125.209.235', '128.127.105', '129.206.45', '13.66.139', '135.125.216',
'135.181.137', '135.181.138', '135.181.140', '135.181.62', '135.181.74', '141.8.142',
'144.217.135', '147.162.3', '147.162.36', '147.92.153', '149.154.161', '149.155.131', '149.56.150',
'149.56.160', '154.51.131', '154.54.249', '157.90.177', '157.90.182', '157.90.209', '159.138.102',
'162.19.100', '162.19.101', '168.119.64', '168.119.65', '168.119.68', '17.121.112', '17.121.113',
'17.121.114', '17.121.115', '17.22.237', '17.22.245', '17.22.253', '17.241.219', '17.241.227',
'17.241.75', '17.246.15', '17.246.19', '17.246.23', '179.190.203', '184.75.211', '185.101.32',
'185.138.241', '185.191.171', '185.54.230', '189.124.192', '191.96.106', '194.187.169',
'194.187.171', '194.9.191', '195.154.122', '195.154.123', '195.154.126', '195.191.219',
'195.201.106', '198.134.108', '198.134.109', '198.251.73', '198.98.183', '199.47.82', '20.15.133',
'204.15.110', '207.241.229', '207.241.231', '207.241.233', '207.241.235', '207.241.236',
'211.249.46', '213.180.203', '216.244.66', '220.181.108', '23.22.35', '3.224.220', '31.3.152',
'31.3.153', '35.196.132', '35.227.62', '35.237.4', '37.46.121', '49.7.20', '49.7.21', '5.102.173',
'5.133.192', '5.255.231', '5.255.253', '5.45.207', '5.62.16', '5.62.20', '5.62.34', '5.62.41', '5.62.43',
'5.62.47', '5.62.49', '5.62.56', '5.62.57', '5.62.58', '5.62.59', '5.62.60', '5.62.61', '5.62.62', '5.62.63',
'50.21.188', '51.210.144', '52.167.144', '52.70.240', '54.161.41', '54.236.1', '54.236.121',
'54.36.148', '54.36.149', '58.250.125', '61.135.159', '65.108.103', '65.108.41', '65.108.6',
'65.108.64', '65.108.99', '65.21.231', '65.21.35', '69.160.160', '74.208.2', '76.164.224',
'77.234.43', '77.234.44', '77.75.72', '77.75.73', '77.75.76', '77.75.77', '77.75.78', '77.75.79',
'77.88.5', '79.142.76', '81.17.57', '85.208.96', '85.208.98', '85.31.186', '87.250.224', '91.143.80',
'91.219.212', '91.242.162', '93.158.161', '95.108.213', '95.142.112', '95.142.121', '95.142.127',
'95.217.72', '95.87.154');