

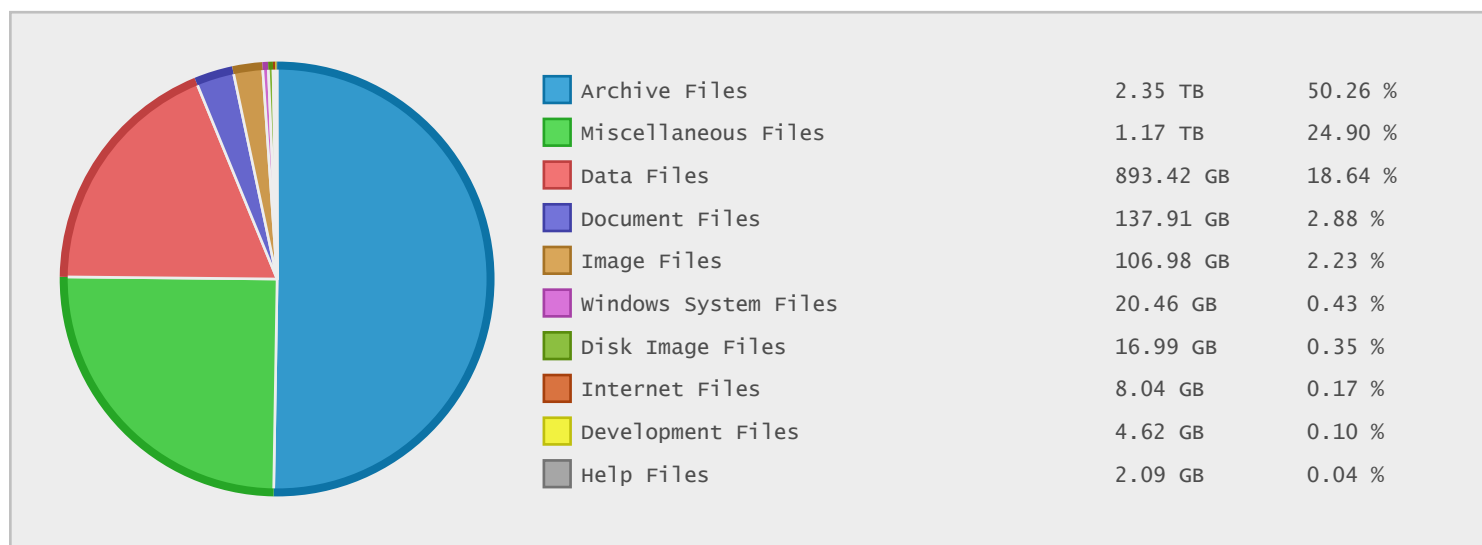
Massachusetts Institute of Technology - FTP Site Statistics

Property	Value
FTP Server	ftp-genome.wi.mit.edu
Description	Massachusetts Institute of Technology
Country	United States
Scan Date	30/Oct/2014
Total Dirs	20,171
Total Files	1,218,376
Total Data	4.68 TB

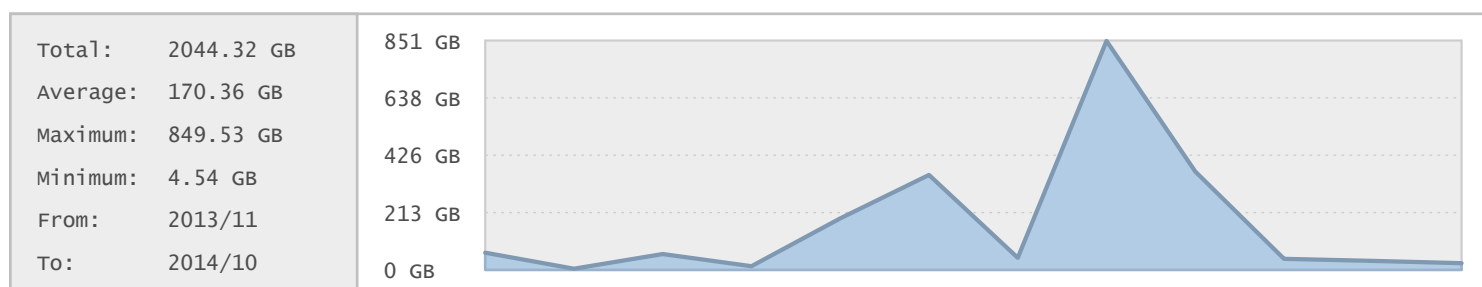
Top 20 Directories Sorted by Disk Space

Name	Dirs	Files	Data	Data (%)
/	20171	1218376	4.68 TB	100.00 %
/distribution	18268	949866	2.88 TB	61.41 %
/outgoing	1864	267902	931.88 GB	19.44 %
/incoming	36	606	918.09 GB	19.15 %
/outgoing/lincRNA	1064	35001	761.88 GB	15.89 %
/outgoing/lincRNA/CLIP	1059	34937	711.79 GB	14.85 %
/outgoing/lincRNA/CLIP/HiSeq_01_15_14	1058	34937	711.79 GB	14.85 %
/distribution/papers	273	5034	489.47 GB	10.21 %
/incoming/chengz	1	3	448.28 GB	9.35 %
/incoming/chengz/Luhan	0	3	448.28 GB	9.35 %
/distribution/assemblies	148	1552	415.35 GB	8.66 %
/distribution/proteomics	536	128954	398.65 GB	8.32 %
/distribution/assemblies/mammals	113	1330	371.42 GB	7.75 %
/distribution/seq	12698	764367	358.93 GB	7.49 %
/distribution/seq/msc	12667	763370	345.91 GB	7.22 %
/distribution/proteomics/public_datasets	433	124627	298.04 GB	6.22 %
/incoming/Dirk_RISK_FASTQ	0	359	293.37 GB	6.12 %
/distribution/genepattern	3435	21100	279.27 GB	5.83 %
/distribution/seq/msc/anopheles	27	331	237.00 GB	4.94 %
/distribution/genepattern/rna_seq	40	255	225.39 GB	4.70 %

Top 10 File Categories Sorted By Disk Space

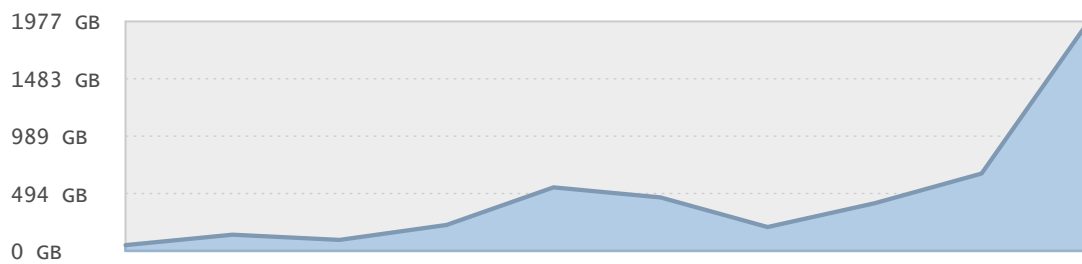


Last 12 Months Modified Disk Space History



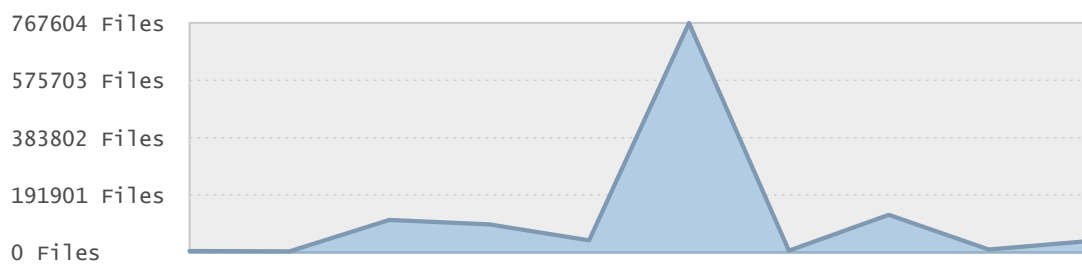
Last 10 Years Modified Disk Space History

Total:	4772.14 GB
Average:	477.21 GB
Maximum:	1975.83 GB
Minimum:	49.54 GB
From:	2005
To:	2014

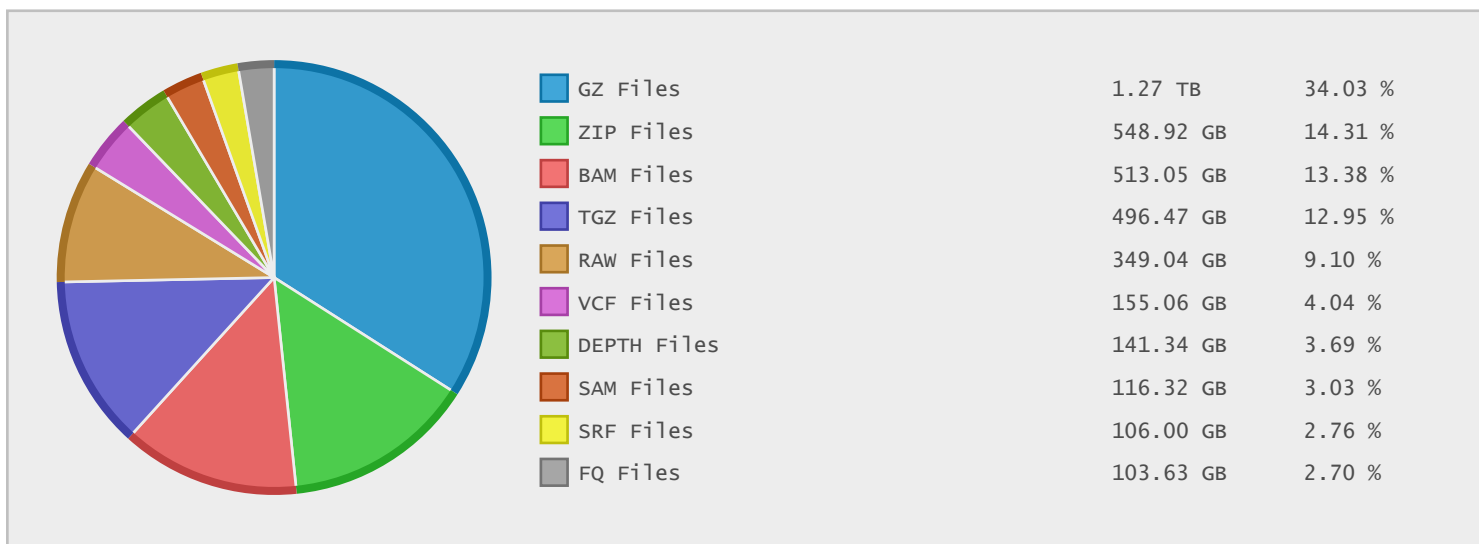


Last 10 Years Modified File Count History

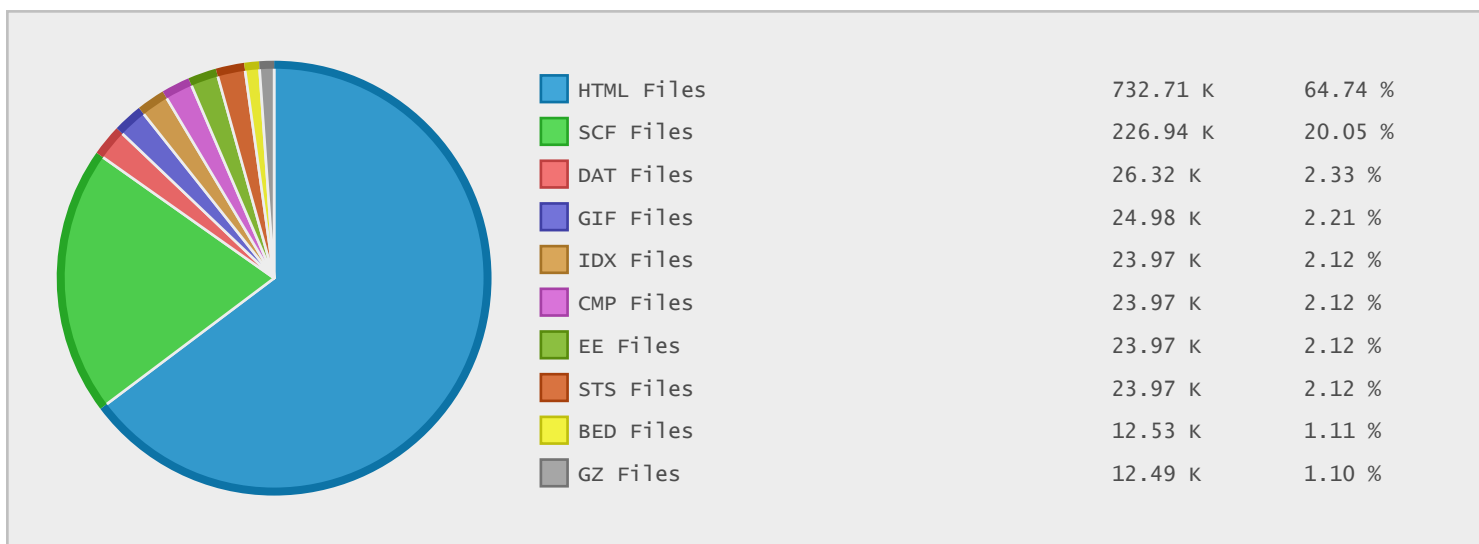
Total:	1196335 ...
Average:	119634 Files
Maximum:	767603 Files
Minimum:	3303 Files
From:	2005
To:	2014



Top 10 File Extensions Sorted By Disk Space



Top 10 File Extensions Sorted By Number of Files



Top 10 File Categories Sorted by Disk Space

Category	Files	Data	Data (%)
Archive Files	18924	2.35 TB	50.24 %
Miscellaneous Files	118004	1.17 TB	24.89 %
Data Files	66357	893.42 GB	18.64 %
Document Files	11822	137.91 GB	2.88 %
Image Files	28524	106.98 GB	2.23 %
Windows System Files	228061	20.46 GB	0.43 %
Disk Image Files	21	16.99 GB	0.35 %
Internet Files	741767	8.04 GB	0.17 %
Development Files	3943	4.62 GB	0.10 %
Help Files	219	2.09 GB	0.04 %

Top 20 File Types Sorted by Disk Space

File Type	Files	Data	Data (%)
GZIP Compressed Archive File	12492	1.27 TB	27.22 %
Miscellaneous Files	118004	1.17 TB	24.89 %
ZIP Compressed Archive File	3971	548.92 GB	11.45 %
Genome Analysis Data File	3121	513.05 GB	10.70 %
GZIP Compressed TAR Archive	1148	496.47 GB	10.36 %
RAW Data File	1182	349.04 GB	7.28 %
Lotus Word Pro Document	203	116.32 GB	2.43 %
Sony RAW Image File	11	106.00 GB	2.21 %
TAR Archive File	25	41.62 GB	0.87 %
Windows Explorer Command	226944	20.39 GB	0.43 %
Binary Data File	26375	17.42 GB	0.36 %
BZIP Compressed Archive File	868	16.75 GB	0.35 %
Open Virtualization Format File	2	15.76 GB	0.33 %
Text Document	3915	14.47 GB	0.30 %
Table File	26	5.79 GB	0.12 %
Java Archive Applet File	8355	4.25 GB	0.09 %
HTML Document	732713	3.78 GB	0.08 %
Common Data Format File	31	3.70 GB	0.08 %
CSV Data File	351	2.93 GB	0.06 %
Adobe Plugin File	1288	2.58 GB	0.05 %

Top 20 File Extensions Sorted by Disk Space

Extension	Files	Data	Data (%)
GZ Files	12492	1.27 TB	27.22 %
ZIP Files	3970	548.92 GB	11.45 %
BAM Files	3121	513.05 GB	10.70 %
TGZ Files	1148	496.47 GB	10.36 %
RAW Files	1182	349.04 GB	7.28 %
VCF Files	34	155.06 GB	3.23 %
DEPTH Files	17	141.34 GB	2.95 %
SAM Files	203	116.32 GB	2.43 %
SRF Files	11	106.00 GB	2.21 %
FQ Files	196	103.63 GB	2.16 %
FASTA Files	214	102.33 GB	2.13 %
BED Files	12526	89.60 GB	1.87 %
BSUB Files	3760	76.10 GB	1.59 %
FASTQ Files	276	55.68 GB	1.16 %
FA Files	829	55.34 GB	1.15 %
10MERS Files	2581	52.53 GB	1.10 %
MAF Files	180	51.18 GB	1.07 %
TAR Files	25	41.62 GB	0.87 %
READS Files	175	40.68 GB	0.85 %
EFASTA Files	11	27.89 GB	0.58 %

Top 50 Directories Sorted by Disk Space

Name	Dirs	Files	Data	Data (%)
/	20171	1218376	4.68 TB	100.00 %
/distribution	18268	949866	2.88 TB	61.41 %
/outgoing	1864	267902	931.88 GB	19.44 %
/incoming	36	606	918.09 GB	19.15 %
/outgoing/lincRNA	1064	35001	761.88 GB	15.89 %
/outgoing/lincRNA/CLIP	1059	34937	711.79 GB	14.85 %
/outgoing/lincRNA/CLIP/HiSeq_01_15_14	1058	34937	711.79 GB	14.85 %
/distribution/papers	273	5034	489.47 GB	10.21 %
/incoming/chengz	1	3	448.28 GB	9.35 %
/incoming/chengz/Luhan	0	3	448.28 GB	9.35 %
/distribution/assemblies	148	1552	415.35 GB	8.66 %
/distribution/proteomics	536	128954	398.65 GB	8.32 %
/distribution/assemblies/mammals	113	1330	371.42 GB	7.75 %
/distribution/seq	12698	764367	358.93 GB	7.49 %
/distribution/seq/msc	12667	763370	345.91 GB	7.22 %
/distribution/proteomics/public_datasets	433	124627	298.04 GB	6.22 %
/incoming/Dirk_RISK_FASTQ	0	359	293.37 GB	6.12 %
/distribution/genepattern	3435	21100	279.27 GB	5.83 %
/distribution/seq/msc/anopheles	27	331	237.00 GB	4.94 %
/distribution/genepattern/rna_seq	40	255	225.39 GB	4.70 %
/distribution/crd	273	3801	178.20 GB	3.72 %
/distribution/cancer	163	264	165.31 GB	3.45 %
/outgoing/lincRNA/CLIP/HiSeq_01_15_14/...	428	12244	162.98 GB	3.40 %
/outgoing/lincRNA/CLIP/HiSeq_01_15_14/...	488	14036	159.75 GB	3.33 %
/distribution/genepattern/rna_seq/pre_...	4	167	159.04 GB	3.32 %
/distribution/gsa	18	628	154.87 GB	3.23 %
/outgoing/lincRNA/CLIP/HiSeq_01_15_14/Pam	41	1318	154.04 GB	3.21 %
/distribution/cel_files	6	1831	150.68 GB	3.14 %
/outgoing/lincRNA/CLIP/HiSeq_01_15_14/...	1	34	141.36 GB	2.95 %
/distribution/papers/chipseq	22	497	129.49 GB	2.70 %
/outgoing/lincRNA/CLIP/HiSeq_01_15_14/...	0	31	115.32 GB	2.41 %
/distribution/papers/illumina_bias	1	14	114.06 GB	2.38 %
/distribution/papers/illumina_bias/Air...	0	14	114.06 GB	2.38 %
/distribution/seq/msc/pub	12636	763015	108.04 GB	2.25 %
/distribution/crd/dropbox	5	61	107.26 GB	2.24 %
/distribution/seq/msc/anopheles/walton	0	6	104.24 GB	2.17 %
/distribution/svtoolkit	41	163	103.80 GB	2.17 %
/distribution/gsa/1000GenomesLowCovRecal	4	182	103.79 GB	2.16 %
/distribution/cancer/gcnmp	66	119	100.08 GB	2.09 %
/distribution/papers/assembly	21	347	87.65 GB	1.83 %
/outgoing/lincRNA/CLIP/HiSeq_01_15_14/data	19	1991	87.03 GB	1.82 %
/distribution/svtoolkit/misc	30	102	85.91 GB	1.79 %
/distribution/svtoolkit/misc/1kg	23	89	84.07 GB	1.75 %
/distribution/proteomics/public_datase...	18	316	81.10 GB	1.69 %
/distribution/seq/msc/anopheles/plasmodium	9	171	77.08 GB	1.61 %
/outgoing/lincRNA/CLIP/HiSeq_01_15_14/...	4	2520	76.38 GB	1.59 %
/distribution/annotation	162	12294	75.23 GB	1.57 %
/distribution/proteomics/public_datase...	14	289	73.03 GB	1.52 %
/distribution/genepattern/rna_seq/pre_...	0	72	70.51 GB	1.47 %
/distribution/papers/assembly/MacCallu...	12	293	70.35 GB	1.47 %

Top 50 Files Sorted by File Size

File Name	Modified	File Size
/incoming/chengz/Luhan/Luhan_cas9_fastq.tgz	04/Apr/2014	289.58 GB
/incoming/Chengz/Luhan/Luhan_cas9.tgz	29/Mar/2014	136.92 GB
/distribution/papers/illumina_bias/A...ABXX.7.aligned.duplicates_marked.bam	09/Dec/2010	33.03 GB
/incoming/Csh/csh_guam_final.tgz	16/Aug/2014	27.49 GB
/distribution/papers/illumina_bias/A...ABXX.8.aligned.duplicates_marked.bam	09/Dec/2010	26.18 GB
/distribution/assemblies/mammals/29mammals/pi.ewig.tar	03/Jun/2011	23.80 GB
/distribution/seq/msc/anopheles/RNAseq/RNAseq_Broad_Jan2012.tar.gz	13/Feb/2012	23.60 GB
/distribution/gsa/forDavidCraig/NA06985.SOLID.SRP000031.2009_04.recal.bam	30/Jun/2009	22.89 GB
/incoming/chengz/Luhan/Luhan-1_S1_L001_R1_001.fastq.gz	02/Apr/2014	21.77 GB
/distribution/papers/cDNA/Levin2009/...081030_SL-XBF_0001_FC30CB2AAXX.6.srf	03/Jun/2009	19.81 GB
/distribution/gsa/1000GenomesLowCovRecal/LowCov-YRI.list.recal2.bam	29/Jun/2009	19.49 GB
/distribution/assemblies/mammals/microbat/myoluc2/MITO/unplaced.qual.gz	02/Aug/2010	18.31 GB
/outgoing/lincRNA/forNEB/mES.gsnap.bam	02/Oct/2013	18.25 GB
/outgoing/lincRNA/forNEB/total.gsnap.sorted.bam	02/Oct/2013	18.25 GB
/distribution/seq/msc/anopheles/walton/G20366.vcf	18/Jun/2013	17.80 GB
/distribution/seq/msc/anopheles/walton/G20367.vcf	18/Jun/2013	17.75 GB
/distribution/seq/msc/anopheles/walton/G20369.vcf	18/Jun/2013	17.53 GB
/distribution/seq/msc/anopheles/vcf/filtered.vcf	08/May/2013	17.50 GB
/distribution/seq/msc/anopheles/walton/G20368.vcf	18/Jun/2013	17.22 GB
/distribution/seq/msc/anopheles/walton/G20354.vcf	18/Jun/2013	17.00 GB
/distribution/seq/msc/anopheles/walton/G20355.vcf	18/Jun/2013	16.94 GB
/distribution/papers/cDNA/Levin2009/...081030_SL-XBF_0001_FC30CB2AAXX.7.srf	03/Jun/2009	16.81 GB
/distribution/gsa/1000GenomesLowCovRecal/LowCov-CEU.list.recal2.bam	29/Jun/2009	15.03 GB
/distribution/igv/TEST/ALL.chr22.pha...101123.snps_indels_svsn.genotypes.vcf	22/Nov/2013	14.25 GB
/distribution/papers/illumina_bias/Aird2010/205D8ABXX.8.aligned.bam	02/Nov/2010	14.16 GB
/distribution/papers/illumina_bias/Aird2010/205D8ABXX.7.aligned.bam	02/Nov/2010	13.93 GB
/distribution/svtoolkit/misc/1kg/042...ni25_superarray_intensity_matrix.dat	22/Apr/2011	13.73 GB
/distribution/papers/illumina_bias/Aird2010/205D8ABXX.6.aligned.bam	02/Nov/2010	13.59 GB
/distribution/svtoolkit/misc/1kg/IRS...y_matrix_2506samples.20120621.dat.gz	14/Jan/2014	13.58 GB
/distribution/crd/dropbox/bob/a.fin/a.paths	11/Jun/2014	13.55 GB
/distribution/papers/illumina_bias/Aird2010/205D8ABXX.5.aligned.bam	02/Nov/2010	13.11 GB
/distribution/papers/hybrid_selectio...081016_SL-XAL_0002_FC30CB3AAXX.6.srf	15/Jan/2009	12.35 GB
/distribution/svtoolkit/public_metadata/1000G_phase3_20130502_mdv1.tar.gz	15/Jan/2014	12.24 GB
/incoming/Csh/csh_guam_ok.tgz	16/Aug/2014	11.18 GB
/distribution/crd/ALLPATHS/Release-2-0/sample-2.0.ncrassa.tar.gz	10/Apr/2009	10.84 GB
/distribution/papers/assembly/MacCal...assemblies/sample-2.0.ncrassa.tar.gz	13/Apr/2009	10.84 GB
/distribution/genepattern/rna_seq/StrandSpecRNASeqEval/dUTP.sam	19/Jun/2010	10.64 GB
/distribution/papers/hybrid_selectio...080122_SL-XAL_0001_FC205G5AAXX.5.srf	15/Jan/2009	10.45 GB
/distribution/assemblies/mammals/29mammals/omega.12mers.wig.gz	03/Jun/2011	10.13 GB
/distribution/svtoolkit/misc/1kg/IRS...y_matrix_2141samples.20120131.dat.gz	14/Jan/2014	9.30 GB
/distribution/svtoolkit/misc/1kg/Omn...y_matrix_2141samples.20120131.dat.gz	29/May/2013	9.30 GB
/distribution/svtoolkit/misc/1kg/IRS...atrix_2141samples_v2.20120131.dat.gz	04/May/2014	9.27 GB
/distribution/svtoolkit/misc/1kg/Omn...atrix_2141samples_v2.20120131.dat.gz	01/May/2014	9.27 GB
/distribution/annotation/rnaSeq/japonicus/merged_sorted_1.fq	10/Dec/2009	8.94 GB
/distribution/annotation/rnaSeq/japonicus/merged_sorted_2.fq	10/Dec/2009	8.94 GB
/distribution/cancer/mmgp/MMRC_Refer...y_number_profiles/mmrctxt.files.zip	29/May/2009	8.93 GB
/distribution/papers/hybrid_selectio...080212_SL-XAL_0001_FC20448AAXX.6.srf	15/Jan/2009	8.78 GB
/outgoing/lincRNA/CLIP/HiSeq_01_15_14/analysis/hnRNPT_3prime_plus.depth	27/Jun/2014	8.70 GB
/outgoing/lincRNA/CLIP/HiSeq_01_15_14/analysis/hnRNPT_5prime_plus.depth	27/Jun/2014	8.70 GB
/outgoing/lincRNA/CLIP/HiSeq_01_15_1...ce_junctions/Input_3prime_plus.depth	27/Jun/2014	8.70 GB