

Moving Momentarily

A simple scraper for Metrorail disruption reports.

Andrew Schmadel ([@schmod](#))

Disruption Reports

Disruption Reports

“ A daily report of anything unusual that happened on Metrorail yesterday. ”



search



Home

Rail

Bus

Accessibility

Getting Around

Fares

Rider Tools

About Metro

Rider Tools

Trip Planner

Alerts & Advisories

ELstat

Elevator/Escalator Status

Next Train Arrivals

BusETA Bus Arrivals

Service Nearby

MetroAlerts

Savings Calculator

Developer Resources

Alerts & Advisories



Rail

Bus

Metrorail Service Status

Sign up for MetroAlerts

Alerts

RED LINE

Red Line: Weekdays trains every 6 mins btwn Shady Grove & NoMa and every 10 mins btwn Ft Totten & Glenmont. For more info visit wmata.com
10/29/2016 08:38 AM

Red Line: Limited shuttles provided btwn NoMa & Ft Totten. If traveling btwn Gallery Pl & Ft Totten, use Green/Yellow as an alternate.
10/29/2016 08:37 AM

Red Line: Buses replace trains btwn NoMa & Ft Totten due to SafeTrack. As a result, Brookland & Rhode Island Ave stations are closed.



DAILY SERVICE REPORT

[All Daily Service Reports](#)

[Archives](#)



SUBSCRIBE TO
METRO NEWS
RELEASES

Daily Service Report for Monday, December 12, 2016

6:07 a.m. A Branch Avenue-bound Green Line train at Waterfront was delayed 4 minutes due to a door problem.

6:50 a.m. A Silver Spring-bound Red Line train at Rhode Island Avenue was delayed 5 minutes due to a signal problem.

7:18 a.m. A Vienna-bound Orange Line train at Potomac Avenue was offloaded due to a door problem. Passengers experienced a 20-minute delay.

8:13 a.m. A Branch Avenue-bound Red Line train at U Street was delayed 20 minutes due to a signal problem.

9:01 a.m. A Glenmont-bound Red Line train at Twinbrook was offloaded due to an equipment problem. Several trains were single tracked around the incident train and several trains were offloaded and turned back for schedule adherence/improved train spacing. Passengers experienced delays up to 30 minutes.

Service Report for December 12, 2016

6:07 a.m. A Branch Avenue-bound Green Line train at Waterfront was delayed 4 minutes due to a door problem.

6:50 a.m. A Silver Spring-bound Red Line train at Rhode Island Avenue was delayed 5 minutes due to a signal problem.

7:18 a.m. A Vienna-bound Orange Line train at Potomac Avenue was offloaded due to a door problem. Passengers experienced a 20-minute delay.

8:13 a.m. A Branch Avenue-bound Red Line train at U Street was delayed 20 minutes due to a signal problem.

Alerts

- Up-to-the minute disruption information.
- RSS Feed Available
- No archives.

Disruption Reports

- Human-edited report of all disruptions on a given day.
- No API or feed available.
- Archives available for all dates after April 30, 2012

Enough of that.

Let's talk about spreadsheets.

<i>fx</i>	=SUM(A1:5)	
	A	
1		1
2		2
3		3
4		4
5		5
6	A7	15 ×
7	=SUM(A1:5)	

ARRAY_CONSTRAIN TREND SUMXMY2 SUMX2PY2 SUMX2MY2 SUMPRODUCT MMULT MINVERSE MDETERM LOGEST LINEST GROWTH FREQUENCY
TRANSPOSE DMIN DVARP DSUM DSTDEVP DSTDEV DPRODUCT DVAR DMAX DGET DCOUNTA DCOUNT DAVERAGE DAYS360 EDATE EOMONTH HOUR
MINUTE MONTH NETWORKDAYS NETWORKDAYS.INTL NOW SECOND TIME TIMEVALUE TODAY WEEKDAY WEEKNUM WORKDAY WORKDAY.INTL YEAR
YEARFRAC DATE DATEVALUE DAY DATEDIF HEX2DEC BIN2DEC BIN2HEX BIN2OCT OCT2HEX OCT2DEC OCT2BIN HEX2OCT DEC2BIN HEX2BIN DELTA
DEC2OCT DEC2HEX UNIQUE FILTER SORT DB PPMT PRICE PRICEDISC PRICEMAT PV RATE RECEIVED SLN SYD TBILLEQ TBILLPRICE TBILLYIELD XIRR
XNPV YIELD YIELDDISC COUPPCD COUPNUM COUPNCD COUPDAYSNC COUPDAYS COUPDAYBS ACCRINTM ACCRINT CUMIPMT DOLLARDE DOLLARFR
DURATION EFFECT DDB FV FVSCHEDULE INTRATE IPMT IRR MDURATION MIRR NOMINAL NPER DISC NPV PMT CUMPRINC IMPORTDATA SPARKLINE
DETECTLANGUAGE GOOGLEFINANCE GOOGLETRANSLATE IMAGE ARRAYFORMULA IMPORTFEED IMPORTHTML IMPORTRANGE IMPORTXML QUERY NA
TYPE ERROR.TYPE ISBLANK ISEMAIL ISERR ISERROR ISFORMULA ISLOGICAL ISNA ISNONTEXT ISNUMBER ISREF ISURL CELL N ISTECH TRUE SWITCH OR
NOT IFS IFERROR IF FALSE AND GETPIVOTDATA COLUMNS ROWS CHOOSE ADDRESS VLOOKUP ROW OFFSET MATCH LOOKUP INDIRECT INDEX
HYPERLINK HLOOKUP COLUMN ISEVEN ISODD LCM LN LOG LOG10 MOD MROUND MULTINOMIAL ODD PI POWER PRODUCT QUOTIENT RADIANS RAND
RANDBETWEEN ROUND ROUNDDOWN ROUNDUP SERIESSUM SIGN SIN SINH SQRT SQRTPI SUBTOTAL SUM SUMIF SUMIFS SUMSQ TAN TANH TRUNC
ABS ACOS ACOSH ASIN ASINH ATAN ATAN2 ATANH CEILING COMBIN COS COSH COUNTBLANK COUNTIF COUNTIFS COUNTUNIQUE DEGREES ERFC
EVEN EXP FACT FACTDOUBLE FLOOR GAMMALN GCD INT MULTIPLY POW CONCAT DIVIDE EQ GT GTE LT LTE MINUS ADD NE UMINUS UNARY_PERCENT
UPLUS TO_TEXT TO_PURE_NUMBER TO_PERCENT TO_DOLLARS TO_DATE CONVERT GAMMA.DIST GAMMADIST GEOMEAN HARMEAN HYPGEOMDIST
INTERCEPT KURT LARGE LOGINV LOGNORMDIST MAX MAXA MAXIFS MEDIAN MIN MINA MINIFS MODE NEGBINOMDIST NORMDIST NORMINV
NORMSDIST NORMSINV PEARSON PERCENTILE PERCENTRANK PERCENTRANK.EXC PERCENTRANK.INC PERMUT POISSON PROB QUARTILE RANK
RANK.AVG RANK.EQ RSQ SKEW SLOPE SMALL STANDARDIZE STDEV STDEVA STDEVP STDEVPA STEYX T.INV T.INV.2T TDIST TINV TRIMMEAN CHITEST
VAR VARA VARP VARPA WEIBULL ZTEST CHISQ.INV.RT CONFIDENCE CORREL COUNT COUNTA COVAR CRITBINOM DEVSQ EXPONDIST F.DIST F.DIST.RT
F.INV F.INV.RT F.TEST FDIST FINV FISHER FISHERINV FORECAST FTEST AVEDEV AVERAGE AVERAGEA AVERAGEIF AVERAGEIFS BINOMDIST CHIDIST
CHIINV CHISQ.DIST CHISQ.DIST.RT CHISQ.INV TTEST UPPER TRIM TEXTJOIN TEXT T SUBSTITUTE SPLIT SEARCHB SEARCH ROMAN RIGHT REPT
REPLACE REGEXREPLACE REGEXMATCH REGEXEXTRACT VALUE MID LOWER LEN LEFT JOIN FIXED FINDB FIND EXACT DOLLAR CONCATENATE CODE
CLEAN CHAR ARABIC PROPER"

Three functions stand out: `IMPORTFEED`, `IMPORTHTML`, and `IMPORTXML`.

	A	B	C	D	E	F	G	H
1	RED	http://www.wmata.com	Wed, 14 Dec 201	Red Line: Single tracking btwn Van Ness & Dupont Circle due to scheduled track work. Expect delays through tonight's closing.				
2	SILVER	http://www.wmata.com	Tue, 13 Dec 201	Silver Line: Through December 21, trains will operate every 20 mins btwn Wiehle & Largo due to SafeTrack. For more info, visit wmata.com				
3	ORANGE	http://www.wmata.com	Mon, 28 Nov 201	Orange Line: Thru December 21, trains operate every 20 mins btwn New Carrollton & Vienna due to SafeTrack. For more info, visit wmata.com				
4	ORANGE	http://www.wmata.com	Mon, 28 Nov 201	Orange Line: Through December 21, trains single track btwn E Falls Church & W Falls Church due to SafeTrack. Expect crowded trains & delays				
5				This is an important message from MetroAccess. MetroAccess subscription trips for the Thanksgiving holiday, Thursday, November 24 and Friday, November 25, have been canceled. MetroAccess customers who wish to travel on these dates may book trips utilizing the MetroAccess online reservation system at www.WMATA.com or contact the MetroAccess Reservations Center at 301-562-5360 to speak with a reservationist. Subscription trips for Saturday, November 26, will operate as normal. If you do not plan to travel				

A	B	C	D	E	F	G
<code>=TRANPOSE(IMPORTXML("https://www.wmata.com/service/daily-report/daily-service-report.cfm?id=20161213013733", "//div[@id='cs_control_2897']"))</code>						

ALGORITHMS BY COMPLEXITY

MORE COMPLEX →

LEFTPAD QUICKSORT

GIT
MERGE

SELF-
DRIVING
CAR

GOOGLE
SEARCH
BACKEND

SPRAWLING EXCEL SPREADSHEET
BUILT UP OVER 20 YEARS BY A
CHURCH GROUP IN NEBRASKA TO
COORDINATE THEIR SCHEDULING

Google Apps Script!

- A tiny JavaScript engine that sits on top of Google Apps
- No NPM, No Build Tools, No ES6.
- Web-based IDE
- Comprehensive standard library.
- Scripts can run via triggers, or a timer.
- Comprehensive [documentation](#).

```
function getReport(id){
  var url = "http://www.wmata.com/rail/service_reports/viewPage_update.cfm?Rep
  var html = UrlFetchApp.fetch(url).getContentText();

  var date = new Date(content.match(/<h1>Daily Service Report for (.*)</h1>/))
  var incidents = [];

  var regexp = /<[pP]>([^\<]*)/g;
  var description, time, line, direction, cause;
  while ((matches = regexp.exec(content)) !== null){
    var incident = (matches[1].replace(/<[pP]>/, ' ').trim());
    if(incident != ' '){

      incidents.push({
        description: incident.replace(/^([0-1]?\d:[0-6]\d [ap]\.m\./, '') || null
        time: match(/^[0-1]?\d:[0-6]\d [ap]\.m\./i,incident,0),
        line: match(/((Red)|(Blue)|(Green)|(Yellow)|(Orange)|(Silver)) Line/i,
        direction: match(/A (.*)-bound/i,incident,1) || match(/(both direction
        delay: parseDelay(incident),
        cause: parseCause(incident)
      });
    }
  }
}
```

Lessons Learned

- Keep It Simple
 - Regular Expressions ended up providing a surprisingly durable scraper!
 - Structure of page didn't change *at all* between 2013 and 2016.
 - The disruption reports are worded consistently, and predictably.
- "Set it and forget it"
 - Google provided the hosting, and (mostly) kept scraping updates every night.
 - No dependencies outside of Google Apps Script.

Lessons Learned

- Closed-source, open-data
 - No emotional burden of managing an open-source project.
 - Didn't want lots of people running scraper and burdening WMATA's website.
 - ✗ Didn't provide consumers of the data with enough insight into how I parsed the data.
 - ✗ Dataset was used by several journalists. None contacted me beforehand. A few did not provide attribution.

Lessons Learned

- On November 4, 2016, a Glenmont-bound Red Line train at Silver Spring was delayed 15 minutes due to a deer in the tunnel.

Thanks!

Andrew Schmadel

schmod.com