Processing Reports

- Reporting Format
- Sample file
- Parsing reports
  - Fetching the last report
  - User contributed code samples

Reporting Format

After processing your file, processing reports are automatically generated and uploaded to your organisations report area on dataupload.wmda.info. The reports have the following format:

Line	Content
1-5	Meta processing information
6-9	Totals

10 and further	Details per records:
	
	Field separator : |
	
	`Column 1 : R (rejection) or W (warning)`
	`Column 2 : POOL (ION)`
	`Column 3 : GRID`
	`Column 4 : Error messages.`
	
	Please note that HLA errors and warnings may contain multiple messages on the same line. Individual messages are separated by #

last 2 lines	Processing statistics (performance)

Sample file

```
File: ION-XXXX-D.xml.gpg 2018-05-21 08:04:35
Pool(s): XXXX
Content Type: D
Start processing: 2018-05-21 08:05:00
Schema version: 2.1
Total records processed: 3427
Total records with warnings: 280
Total records rejected: 51
Total valid records: 3376


W | XXXX | 947522905891 | N/A | HLA:   #Error at locus DRB1*: invalid HLA code DRB1*8:04. Locus blanked.
R | XXXX | 947529306154 | N/A | BIRTH_DATE (Record Rejected) Field type must be [date]. (Record Rejected) Date
is not provided in the correct format. The format must be [YYYY-MM-DD].
W | XXXX | 947532906148 | N/A | HLA:  #Warning at locus A*: deprecated HLA code A*23:AHENZ
W | XXXX | 947534106167 | N/A | HLA:   #Error at locus B*: invalid HLA code B*7:XX. Locus blanked. #Error at
locus B*: invalid HLA code B*8:XX. Locus blanked.
R | XXXX | 947534106167 | N/A | (Record Rejected) Incomplete typing found: HLA-A, HLA-B are required for DONOR
and CBU. HLA-DRB1 is required for CBU.
W | XXXX | 947535306189 | N/A | HLA:   #Error at locus A*: invalid HLA code A*3:XX. Locus blanked.
R | XXXX | 947535306189 | N/A | (Record Rejected) Incomplete typing found: HLA-A, HLA-B are required for DONOR
and CBU. HLA-DRB1 is required for CBU.
R | XXXX | 947537206209 | N/A | BIRTH_DATE (Record Rejected) Field type must be [date]. (Record Rejected) Date
is not provided in the correct format. The format must be [YYYY-MM-DD].
W | XXXX | 947537506212 | N/A | HLA:   #Error at locus DRB1*: invalid HLA code DRB1*8:XX. Locus blanked.
W | XXXX | 733230748329 | N/A | HLA:   #Error at locus DRB1*: invalid HLA code DRB1*1101. Locus blanked. #Error
at locus DRB1*: invalid HLA code DRB1*1111. Locus blanked.
W | XXXX | 734555524080 | N/A | HLA:   #Error at locus C*: invalid HLA code C*0102. Locus blanked. #Error at
locus C*: invalid HLA code C*0401/15. Locus blanked.
W | XXXX | 733260938482 | N/A | HLA:   #Error at locus C*: invalid HLA code C*0401. Locus blanked. #Error at
locus C*: invalid HLA code C*0501. Locus blanked.
............
Processing finished at: 2018-05-21 08:06:32
Total processing time: 1 minutes.
```

Parsing reports

You may use standard text processing tools like SED, AWK and TR to process the report file, eg to collect statistics: please note that these are only examples

Sample processing directives (suggestions)

```
# Example 1 : Count warnings and errors per locus :
# Replace all # by newline, get the lines with locus errors, then pick the first part before: , sort it, and
count unique values

tr '#' '\n' < testreport.txt |grep 'at locu' | cut -d':' -f1 | sort | uniq -c
     32 Error at locus A*
     24 Error at locus B*
    112 Error at locus C*
     16 Error at locus DQB1*
     57 Error at locus DRB1*
     76 Warning at locus A*
     80 Warning at locus B*

# Example 2 : create a list of all ID's that have HLA warnings
tr ':' '|' <testreport.txt | cut -d'|' -f1,3,5 | grep -i 'HLA'

# Example 3 : Give ID from records that have a rejection due to invalid SEX field
grep -i 'sex.*rejected' testreport.txt | cut -d'|' -f3
 737849095

# Example 4 : get deprecated HLA stats
tr '#' '\n' < testreport.txt |grep 'depr' | cut -d':' -f2 | sort | uniq -c

     67  deprecated HLA code A*01
      2  deprecated HLA code A*02
    113  deprecated HLA code A*03

#Example 5 : Count all warnings
tr '|' '\n' < testreport.txt | sort | uniq -c | grep Warn
```

Fetching the last report

In order to automatically process your reports, you can use the API to fetch the last report.

Documentation is provided here : How to automate data upload/download with the REST API

Basically you ask the server to provide you a list of reports, that you may parse to find the last one.

User contributed code samples

If you have a code sample that you use, you are kindly invited to share that with the WMDA community.

Getting and parsing report (PHP sample by Marco Vitale)

Fetch last report

```
<?php

// Retrieve the last couple of reports generated (donors and cords)

$ion="7450";
$remote_url = "https://dataupload.wmda.info/api/v2/fs/reports-ion".$ion."/?children=f";
include("wmda_credentials.php");

// Create a stream

$opts = array(
  'http'=>array(
    'method'=>"GET",
    'header' => "Authorization: Basic " . base64_encode("$username:$password")
  )
);

$context = stream_context_create($opts);

// Open the file using the HTTP headers set above

$file = file_get_contents($remote_url, false, $context);

$xml = simplexml_load_string($file);
$json = json_encode($xml);
$array = json_decode($json,TRUE);

$mod_c='';$mod_d='';$file_c='';$file_d='';
foreach ($array["tree"] as $ar_file)
   foreach ( $ar_file["@attributes"] as $key => $value )
      {
      if ($key=="filename"){if (strpos($value,$ion."-C")>0){$tipo="C";$file_c=$value;}else{$tipo="
D";$file_d=$value;}}
      else if ($key=="ajxp_modiftime")
         {
         if ($tipo=="C"){if ($value>$mod_c){$mod_c=$value;}}
         else           {if ($value>$mod_d){$mod_d=$value;}}
         }
      }

echo $file_c."\n";
echo $file_d."\n";
?>
```