Deriving Pacific Master’s Records Using the USMS Database

Bob Upshaw

29Dec2019

**Overview:**

The Pacific Masters LMSC maintains its own database of pool records set by Pacific Masters swimmers. A view of this database exists here:

<http://pacificmasters.org/pacm/records>

Maintaining the database is labor intensive, since it requires someone to manually review and approve times achieved in all USMS recognized pool swim meets. To make this a more manageable task it is proposed that a computer program use the USMS database of “fastest times” to look for Pacific Masters times that are faster than the corresponding times in the Pacific Masters Records database. The USMS “fastest times” database can be viewed here:

<http://www.usms.org/comp/meets/lmsc_fastest_times.php>

This document will specify what the program does allowing a software designer to build the program and meet a subset of the requirements that are stated in the Pac Masters Record Automation document here:

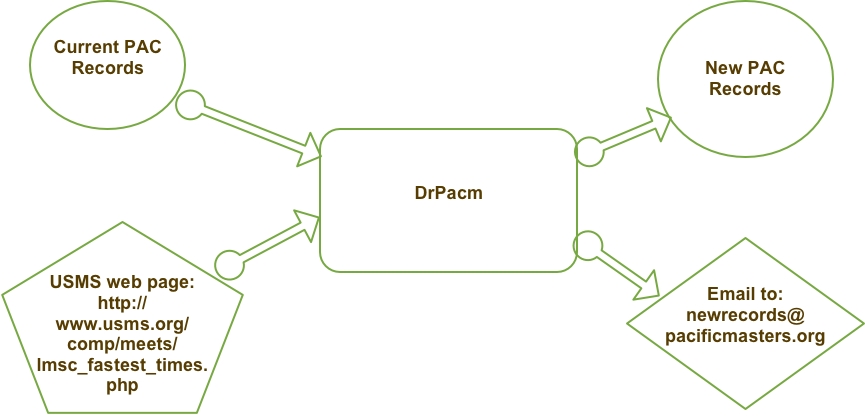
<https://drive.google.com/open?id=1ldhvOnChpiV17ZiTP0O9T4dG0aLehJUq>

**Glossary:**

* Course: SCY (Short Course Yards), SCM (Short Course Meters), and LCM (Long Course Meters). A pool record must be qualified as belonging to exactly one course.
* Pacific Masters Records database: the master collection of Pacific Masters pool records. This may be stored as a HyTek database, Excel spreadsheet, set of MySql tables, or something else. The important point is that we have an API allowing us access to this database.

**Specification:**

The program, hereafter known as “DrPacm” (Derive records for Pacific Masters) will be a PHP script that implements the following high-level flow:



Inputs:

DrPacm will require the following two input streams:

1. Current PAC Records: This will be a list of current Pacific Masters Records for a specific course. Details below.
2. The data presented by <http://www.usms.org/comp/meets/lmsc_fastest_times.php>: Ideallly these data would be available via a web service supported by USMS, but until then DrPacm will have to retrieve the data a different way (e.g. screen scraping.) The data will *probably* be limited to the Pacific Masters fastest times and be qualified by course. See Risks below.

Outputs:

1. New PAC Records: this will be a list of swims for a specific course where
   1. The distance is greater than 25 yards or meters, AND
   2. The time of the swim is greater than 0.00, AND
   3. The time, date (within 5 days), gender, age group, distance, and stroke of the swim does not match any existing PAC Record of any status (ftime), AND
   4. One of the following is true:
   5. The event corresponding to the New PAC Record (gender, age group, distance, and stroke) does not exist in the Pacific Masters Records database AND the swim occurred in the age group 90-94 or greater, OR
   6. The time for the New PAC Record is Faster than the corresponding verified current record in the Pacific Masters Records database, OR
   7. The time for the New PAC Record swim is equal to and more than 5 days after the corresponding verified current record in the Pacific Masters Records database.
2. Email: An email will be sent to the mail list “[newrecords@pacificmasters.org](mailto:newrecords@pacificmasters.org)” whenever one or more of the following situations occur:
   1. Whenever the generated list of New PAC Records contains one or more new records: the email will contain details of the new records.
   2. Whenever an “exception” is raised: the email will contain details of the exception.

See the section titled “[Email](#mpjzxrhacql3)” below.

Known Issues:

1. If two (or more) people of the same gender and age group swim the same stroke and distance in the same course on the same date and tie (their swims are the same duration) for a fastest time, and all appear to DrPacm as new fastest times at the same time, all will qualify as a new unverified PAC record. However, if for some reason only one of the swims appears to DrPacm as a new fastest time, and then the other appears at a different time to DrPacm, the latter fastest time will be silently ignored, since it will violate rule 1c above.
2. If a swimmer swims a new fastest time in a sanctioned meet whose pool does NOT support an electronic timing system (e.g. Rinconada) then neither PMS nor USMS will not recognize that time as a record even though USMS recognizes it as a “fastest time”. DrPacm can only report this fastest time as a candidate new record and no other record for that event will be found by DrPacm until someone swims it faster than this “non-record fastest time”.

**Current PAC Records:**

The list of Current PAC Records will be delivered upon execution of the PHP function named “**ind\_records\_extract**”. For more details see the document named “Maintenance, Storage, and Export of Pacific Masters Pool Records” by Caroline Lambert, the section titled “**Individual records extraction for AGSOTY and DrPacM**”. At the time of this writing this document can be found here:

<https://docs.google.com/document/d/1KsYdXoXzJz-suD39Mh1YAU92psll3HLqxkH3LS7xwlc/edit?ts=5a3c5184>When this function is executed it is passed the desired course which is one of: SCY, SCM, or LCM. The result of the call will be an array of associative arrays where each associative array contains at least the following values:

* **G** : is gender. must be a single letter and one of F or M. Case is insignificant.
* **AgeGroup** : the age group of the swimmer when the record was set. Must be in the form ‘xx-yy’ where ‘xx’ is the youngest age in the age group, and ‘yy’ is the oldest age in the age group. E.g. 18-24 or 50-54. ‘xx’ and ‘yy’ must be 2 or more digits.
* **Distance** : distance of the event. Must be one of: 50, 100, 200, 400, 500, 800, 1000, 1500, and 1650. Note that some distances are legal for only some courses and strokes.
* **Stroke** : the stroke of the event. Must be one of: Freestyle, Butterfly, Backstroke, Breaststroke, or Individual Medley. Note that “Medley” is reserved for another purpose. The spelling of the stroke is case insensitive, thus “Freestyle” and “freestyle” are both recognized as the same.
* **date** : the date of the swim in the form yyyy-mm-dd where yyyy is exactly 4 digits, and mm and dd are 1 or 2 digits.
* **duration** : the duration of the swim, in the form “h:m:s.t’ where:
  + ‘h’ is hours part of the duration. Can be 0 or more digits. If 0 digits the ‘h:’ must be missing, making the duration be in the form ‘m:s.t’. ‘0’ or ‘00’ are valid.
  + ‘m’ is the minutes part of the duration. Must be present. Must be 1 or 2 digits.
  + ‘s’ is the seconds part of the duration. Must be present. Must be 1 or 2 digits.
  + ‘t’ is the fractional part of a second which is part of the duration. Must be 1 or more digits. If only one digit it represents tenths of a second, if 2 digits it represents hundredths, if 3 digits then thousandths, etc.
* **course** : one of SCY, SCM, or LCM. It must be the same as the ‘XXX’ passed as part of the HTTP query used to receive these data. Case is insignificant.
* **ftime** : the status of the record (see the document named “Maintenance, Storage, and Export of Pacific Masters Pool Records” by Caroline Lambert for details.) All records regardless of ftime are used by DrPacm to avoid duplicate unverified records.
* **splashId** : a unique identification of the splash that set the record, as the last part of the URL for the splash on the USMS website, if available.

**The data presented by http://www.usms.org/comp/meets/lmsc\_fastest\_times.php:**

DrPacm will process these data as appropriate. Since these data will only be available in a form dictated by USMS, and we have no control of that form, DrPacm will be customized to process these data as necessary. If USMS changes the form of these data then DrPacm will need to be fixed to handle the changes. See Risks below.

**New PAC Records:**

DrPacm will generate a set of zero or more swim records, where each swim record contains the following data:

* **G** : is gender. must be a single letter and one of F or M. It will be capitalized.
* **AgeGroup** : the age group of the swimmer when the record was set. Must be in the form ‘xx-yy’ where ‘xx’ is the youngest age in the age group, and ‘yy’ is the oldest age in the age group. E.g. 18-24 or 50-54. ‘xx’ and ‘yy’ must be 2 or more digits.
* **Distance** : distance of the event. Must be one of: 50, 100, 200, 400, 500, 800, 1000, 1500, and 1650. Note that some distances are legal for only some courses and strokes.
* **Stroke** : the stroke of the event. Must be one of: Freestyle, Butterfly, Backstroke, Breaststroke, or Individual Medley. Note that “Medley” is reserved for another purpose. The stroke will have the first letter of each word capitalized.
* **FullName** : swimmer’s full name in display format, e.g. “Bob B Upshaw” or “Caroline Lambert”. The format will mirror the format used by USMS to display the name on the [lmsc\_fastest\_times.php](http://www.usms.org/comp/meets/lmsc_fastest_times.php) page.
* **Club** : the initials of the club the swimmer belonged to when the record was set.
* **date** : the date of the swim in the form yyyy-mm-dd where yyyy is exactly 4 digits, and mm and dd are 1 or 2 digits.
* **duration** : the duration of the swim, in the form “h:m:s.t’ where:
  + ‘h’ is hours part of the duration. Can be 0 or more digits. If 0 digits the ‘h:’ must be missing, making the duration be in the form ‘m:s.t’. ‘0’ or ‘00’ are valid.
  + ‘m’ is the minutes part of the duration. Must be present. Must be 1 or 2 digits.
  + ‘s’ is the seconds part of the duration. Must be present. Must be 1 or 2 digits.
  + ‘t’ is the fractional part of a second which is part of the duration. Must be 1 or more digits. If only one digit it represents tenths of a second, if 2 digits it represents hundredths, if 3 digits then thousandths, etc.
* **Sid**: the USMS “swimmer Id” of the swimmer. This will be a sequence of 5 capitalized alphanumeric characters.
* **splashId** : an unique identification of the splash that set the record, or empty if not known. Must be of the form ‘DDDDDDD’:
  + ‘DDDDDDD’ - a sequence of 7 or less digits specified by USMS to identify the splash. Usually this splashId is appended to a URL to request details of the specific splash from which the record was set. Example URL:
    - <http://www.usms.org>/comp/meets/swim.php?s=2721522
* **course** : one of SCY, SCM, or LCM. It must be the same as the ‘XXX’ passed as part of the HTTP query used to receive these data. Case is insignificant.

The set of records can span multiple courses. If the set generated by DrPacm is empty (there were no new records found) then nothing will be generated.

If records are generated then the transport of those records is via a PHP function call. The details of this function call are contained in the document named “Maintenance, Storage, and Export of Pacific Masters Pool Records” by Caroline Lambert. See the section titled “**PHP function for adding records to the database**”. At the time of this writing this document can be found here:

<https://docs.google.com/document/d/1KsYdXoXzJz-suD39Mh1YAU92psll3HLqxkH3LS7xwlc/edit?ts=5a3c5184>

**Email:**

Email will be generated and sent to “[newrecords@pacificmasters.org](mailto:newrecords@pacificmasters.org)” using whatever email infrastructure exists on the server hosting DrPacm (eventually pacificmasters.org.) There will be two conditions that will cause email to be generated:

1. New records (or ties) are discovered and submitted for insertion into the PAC records database. The exact content of the email is TBD, but is likely something like this:

Subject: New Pacific Masters pool records

From: drpacm@pacificmasters.org

The following new Pacific Masters pool records were discovered on <date and time> and uploaded to the pacificmasters.org web server. They will require review by the Pacific Masters administrator before being final.

...record…

...record…

…..

1. An “exception” is discovered. Exceptions are situations discovered by DrPacm and warrant human review. This email will list all exceptions discovered during a single run. The exact content of this email is TBD, but is likely something like this:

Subject: Exceptions discovered during a DrPacm execution

From: drpacm@pacificmasters.org

DrPacm began on <date><time> and here's what happened:

...exception…

...exception…

…..

All exceptions will list details that will allow further investigation. Possible exceptions:

* (Internally known as situation [1]) A possible new record (or newly tied record) was found that is over 1 year old. THIS EXCEPTION DID NOT PREVENT THE NEW RECORD FROM BEING INCLUDED INTO THE PAC RECORDS DATABASE. If this possible new record is not truly a new record then human intervention is required.
* A Fastest Time or a PMS Record was found with a swim time of 0.00. This seems to happen with some NS (No Start) swims. Obviously such a “splash” should not be a fastest time or a record!
* A Fastest Time matching an existing PMS Record (meaning that the gender, age group, distance, stroke, date, duration, and course of the Fastest Time is the same as those fields for the PMS Record), but the splashId of the Fastest Time does NOT match the (non-empty) splashId of the PMS Record then the details of this mismatch will be displayed. This will require human intervention to fix.

**Risks:**

1. DrPacm will depend on USMS data which, at the moment of this writing, is only available on a web page. If that page goes away then DrPacm will fail to work. If USMS stops maintaining that page then DrPacm may receive incorrect data. If the format of that web page changes then DrPacm will have to be modified to correctly parse the changed page.
2. This specification is subject to changes if USMS changes the format for their data. For example, if they use a different scheme to identify a meet then we will probably have to adapt to the same scheme.

**Appendix 1**: **Automation of record verification**

Currently a record is inserted into the PAC database by DrPacm is in an unverified state. Someone has to review the record and “verify” or “nullify” it. Automating this process is work that we have NOT undertaken. This appendix documents some of the steps we could consider automating.

* Pool Certification: a record is not a record unless the pool it was set in was certified. That means that it was measured and found to be the correct length (+/- some margin - unknown what that is.) It may also involve other factors (referee signatures, touch pads, etc.) Here are the steps taken for part of that certification:

1. Use the splash id from the fastest time to view the “Event Results Database Splits” page.
2. On that page click on the “Meet:” link. That will show you the “Event/Meet Details” page.
3. On that page you’ll *hopefully* find a “Location:” or a “Meet Web Info:” field.
   1. Location: one or more of the words in the location will *hopefully* be the pool or facility name.
   2. Web Info: drilling down will *hopefully* tell you the location of the meet.
4. If you have a location then search for that location in the “poollengthdb.xls” file (see below.) If you find one or more rows for this location narrow down your search to the most relevant row (most recent, best match to the full name, etc.)
5. If you’ve narrowed it down to one row of the spreadsheet then look at the “Bulkhead” column and the “No Pad” and “1 Pad” column. If there is a bulkhead then you need to confirm that the pool was measured at the beginning and end of the meet. (How do we do that?) At the moment we’re not exactly sure what to use the “No Pad” and “1 Pad” columns for, but if they are “ok” then the measurement was probably OK.
6. If you cannot find any rows in the spreadsheet matching the location of the meet then it’s possible that the pool was not certified (thus records can not be verified.)

Finding the “poollengthdb.xls” file:

1. Go to usms.org.
2. In the Search box (upper-right corner) enter “Pool Certification List”.
3. Scroll down the results until you find “USMS Measured Pools” with the most recent date. Download that spreadsheet.

**Appendix 2**: **Installing and Executing DrPacm**

At the time of this writing this is how we install and execute DrPacm.

DrPacm runs on our production server but not from the web server tree. Here is how you can install DrPacm if it doesn’t already exist on the production server:

* (log into the production server as ‘pacmasters’
* cd ~/Automation
* git clone <https://github.com/bobup/DrPacm>

If DrPacm already exists and just needs to be updated do this:

* (log into the production server as ‘pacmasters’
* cd ~/Automation/DrPacm
* git pull

At this point the DrPacm/ directory exists and is ready for use. To run DrPacm by hand use the RunDrPacm.bash script.

To automate execution of DrPacm update the crontab entry to say this:

### Run DrPacm on a schedule:

# 4:01am EST every day of the week:

1 4 \* \* 1-7 cd /usr/home/pacmasters/Automation/DrPacm; ./RunDrPacm.bash -p -r -y -l -s -c"These are the results of an AUTOMATIC run of DrPacm on pacmasters@www1.g20.pair.com"