Date Obfuscator

The Date Obfuscator is a utility to turn dates (PHI) into intervals for public consumption. It is used by both the clinical and biospecimen XML generation processes.

For clarity, date values in this document are in the format YYYY-MM-DD. The value "XX" is used to represent missing data.

Assumptions

Events are assumed to take place at midnight on a given day. For example, the interval between 2013-08-20 and 2013-09-05 is 16 days because the events occur at 12:00 AM. There are 12 full days in August and 4 full days in September between the events.

Invalid dates

Invalid dates are days fall into two categories: non-calendar dates and dates which are insufficient to generate XML.

- Non-Calendar Dates: These are days that do not exist such as 4/31/2012. Dates in this category will throw an exception and will stop XML generation.
- Insufficient Dates: Dates without a month or year cannot be used to generate an interval. For example 2012-XX-XX and 2012-03-XX. Dates in either category will return `null` interval with a "Not Available" procurement status.

Interval Types

The obfuscator produces two different type of intervals:

- **DaysTo:** The number of calendar days between two events.
- **AgeAt**: The age of the patient at a given date, in years.

DaysTo Rules

Maximum interval cap

Given dates "A" and "B", date "C" is the date that occurs 90 years before date "B". If Date C is after Date A, then the interval returned is the difference in days between Date A and Date C. This has the effect of capping intervals at 90 years. (However, note that this is not the same as simply returning -365*90 = -32850.)

Example:

- Date A: 1900-01-01
- Date B: 2007-01-01
- Therefore Date C = 2007-01-01 minus 90 years = 1907-01-01

Date C is after Date A. The interval returned is -32872, the number of days between 2007-01-01 and 1907-01-01. The interval returned is not -32850.

A variation of this can occur when two intervals combined would reveal the patient is >= 90 years old. In this case, the second interval should be floored to not reveal the patient's true age.

Example:

- Date of birth: 1925-01-01
- Date of IPD: 2010-01-01
- Date of X: 2016-01-01

The calculated days to birth is 31046 days which is less than 90 years. The calculated days to X is 2191 days which is less than 90 years. However, the *combination* of these intervals is 33237 which is >=than 90 years. Therefore, the days to X is floored to 1826 (90 years - 31046 days = 1826 days).

Missing day value

If a day value is missing for either date it is set to 15.

Example:

- Date A: 2007-11-XX

- Date B: 1951-11-05

Date A's day value is replaced with 15. The interval returned is -20464, the number of days between 2007-11-15 and 1951-11-05.

Example:

- Date A: 2011-07-19
- Date B: 2012-09-XX

Date B's day value is replaced with 15. The interval returned is 107, the number of days between 2011-07-19 and 2012-09-15.

Do not allow nonsensical negative intervals

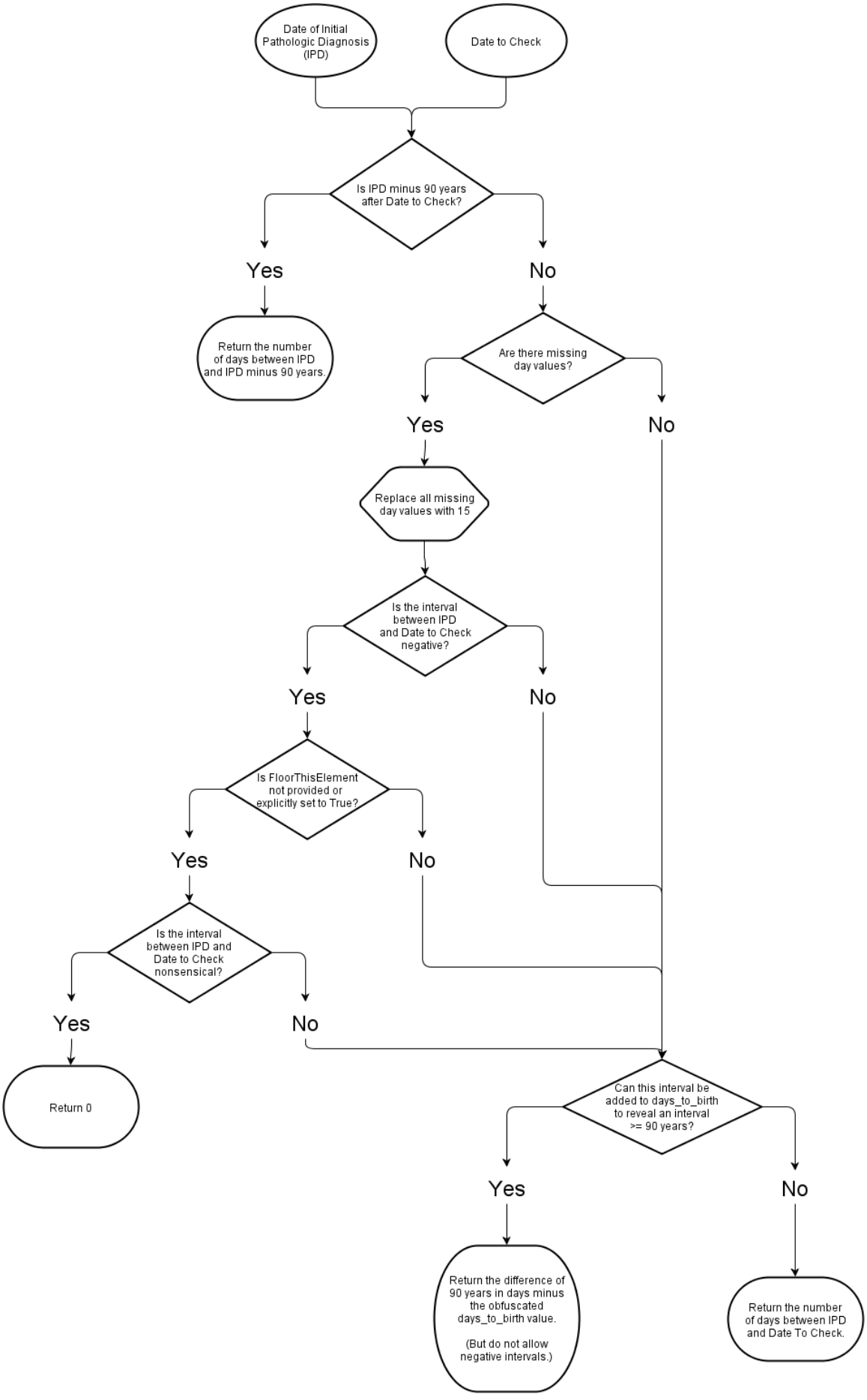
Due to the rule "Missing day value", In the event that a day value is unknown, it is possible to return a negative interval that does not make logical sense. The list of non-negative elements is maintained within the properties of the nch-utilities project. These intervals should be floored to zero. All other elements can legitimately remain negative.

Example:

- Date of initial pathologic diagnosis: 2007-03-20
- Date of treatment start: 2007-03-XX

Date of treatment start's days value is replaced with 15. The interval returned is 0, not -5, the difference between 2007-03-20 and 2007-03-15.

Note: There are some instances when these negative intervals do make clinical sense. To allow for these edge cases, the API provides a switch to turn off this rule.


Date of Initial Pathologic Diagnosis (IPD)
Date to Check
Is IPD minus 90 years after Date to Check?
Yes
No
Return the number of days between IPD and IPD minus 90 years.
Are there missing day values?
Yes
No
Replace all missing day values with 15
Is the interval between IPD and Date to Check negative?
Yes
No
Is FloorThisElement not provided or explicitly set to True?
Yes
No
Is the interval between IPD and Date to Check nonsensical?
Yes
No
Return 0
Can this interval be added to days_to_birth to reveal an interval >= 90 years?
Yes
No
Return the difference of 90 years in days minus the obfuscated days_to_birth value.
(But do not allow negative intervals.)
Return the number of days between IPD and Date To Check.

AgeAt Rules

Maximum age value

Do not allow ages greater than 90. If the age at value is greater than 90, return 90.

Precision

In addition to returning a value and procurement status, the obfuscator will also return a precision attribute that describes how precise the calculation is.

- **day:** The interval is accurate to the day. Thus an interval of 100 with precision "day" means the two events actually occurred 100 days apart.
- **month:** The interval is accurate only to the month. This occurs when a date is incomplete and is missing the day value. An interval of 100 with precision "month" means that the events occurred somewhere in the range of 84 and 100 days. (Because we fill the missing day with 15, the interval value of 100 is exactly correct **only if** the missing date is 15. The maximum variation is if the event occurred on the 31st of the month. 31-15 = 16 days in error.)
- **year:** The year precision is not used. See DCC Wiki: Clinical Date Obfuscation for details.