

APOLLO Date Handling

De-identification of DICOM dates

The resulting DICOM dates are meaningless yet preserve the relative temporal distance between studies for a patient

De-identification of dates uses the DICOM standard “Retain Longitudinal With Modified Dates Option” which allows dates to be retained as long as they are modified from the original date. Date and Date-Time fields in TCIA DICOM image headers are de-identified by normalizing to a base date of January 1, 1975 and then shifted by the number of days between the original Study Date and an "anchor date". The anchor date for APOLLO is the Date of Diagnosis. The choice of '1975' was arbitrary, but it allows one to ensure that the dates in de-identified DICOM files have been properly de-identified as anything not around that year would be suspect.

$\text{TCIA Study Date} = 01/01/1975 + (\text{Original Study Date} - \text{Date of Diagnosis}).$

For example, if the original Study Date was 03/29/2018 and the Date of Diagnosis was 03/27/2018 then the Days from Diagnosis would be +2 and the TCIA Study Date would become 01/03/1975.

This technique de-identifies the dates while preserving the longitudinal relationship between dates. Therefore, a researcher won't know the precise date the scan occurred, but if a follow up scan was performed 120 days later, that same 120 day difference between scans of a subject will exist in the TCIA images. Dates that occur in DICOM tags other than Date or Date-Time fields are removed. An example of this would be a date entered into the Series Description field. If the date is associated with a library for Code Meaning then that date is preserved as the date would be required to look up the meaning in the correct version of the library. To show that the dates have been modified, the term “MODIFIED” is written into DICOM tag (0028,0303) “LongitudinalTemporalInformationModified”.

Original dates will be first normalized to 01 January, 1975 and then offset relative to the date of diagnosis. The CTP code for shifting the StudyDate is shown below:

```
<e en="T" t="00080020" n="StudyDate"> @dateinterval(StudyDate,diagnosisdate,PatientID,@NORMDATE)</e>
```

Insertion of computed "Days from Diagnosis" value

The inserted "Days from Diagnosis" value can be compared with similar values in the APOLLO clinical data to understand the clinical context of the imaging study

The number of days the study occurred relative to the date of diagnosis is calculated by the CTP software (using the diagnosis date in the CTP lookup table at the submission site) and automatically stored in the DICOM tag (0012,0052) Longitudinal Temporal Offset from Event with the associated tag (0012,0053) Longitudinal Temporal Event Type set to "Days from Diagnosis". The days from diagnosis links the imaging data to the clinical data for a given subject. The CTP code for this is:

```
<e en="T" t="00120052" n="LongitudinalTemporalOffsetfromEvent">@always()@dateinterval(StudyDate,ddate,
PatientID)</e>
```

```
<e en="T" t="00120053" n="LongitudinalTemporalEventType">@always()@param(@LTET)</e> (where LTET is
defined as DIAGNOSIS)
```

Insertion of "Diagnosis Year"

It is important for cancer researchers to know the timeframe for which the cancer was diagnosed to relate the prescribed cancer treatment or staging to what was available at that time.

In order to relate the treatments that were available at the time of the diagnosis, the year that the primary diagnosis was made is recorded in a CTP owned group 13 private tag as follows.

```
<e en="T" t="00131051" n="DiagnosisYear">@always()@lookup(PatientID,diagnosisdate)</e>
```

In a separate stage of the pipeline the diagnosisdate is truncated to be just the year that the diagnosis was made.

```
<e en="T" t="00131051" n="DiagnosisYear">@truncate(DiagnosisYear,-4)</e>
```

The approximate StudyYear can be calculated by adding the days from diagnosis in tag LongitudinalTemporalOffsetfromEvent to the DiagnosisYear.

In order to use a normalized date function the private tags must also be de-identified at the site using a CTP script that encapsulates the TCIA Safe Private Tag Knowledge Base. With this approach, only the Safe Private Tags contained within the TCIA Private Tag Knowledge Base and encoded into the CTP script at the time the CTP script was created will be retained. If there are Private Tags that are known to be important but not part of the current Safe tags of the TCIA Private Tag Knowledge Base, then it is up to the submitting site to submit a Private Tag Dictionary of those tags to TCIA for consideration.

The normalized date workflow described above requires that diagnosis date be present and this workflow does not handle the example where there no diagnosis date is present.