Reclaim the Records
BIRLS Death File Release

March 18, 2022

Table of Contents

1   Summary

To comply with a request from the Office of General Counsel (OGC) related to the March 24, 2020, court order in Case 1:18-cv-08449, *Reclaim the Records and Brooke Schreier Ganz verse the Department of Veterans Affairs (VA)*, VA was directed to release the remainder of the BIRLS (Beneficiary Identification Records Locator Subsystem) Death File (not previously released) to Reclaim the Records by April 1, 2022.

To meet the terms of the court order, the Veterans Benefits Administration's (VBA's) Office of Business Integration (OBI) leveraged a team comprised of Data Scientists, Data Analysts and Informaticists to analyze and remove records from the BIRLS Death File (BDF) which may represent living persons to mitigate the risk of releasing Personally Identifiable Information (PII) to the public. This action was taken to address the known risk that an extremely large dataset (>18.5 million persons) would inevitably contain a small percentage of data quality errors, thus, VBA utilized a sound methodology to identify "Evidence of Life".

The BDF contains a total of 18,538,702 records, of which 1,350 (0.037%) were redacted due to "Evidence of Life" as determined by the following three parameters: 1) filing for a claim with VA, 2) reversal of a Date of Death notification and 3) any call or correspondence with VA.

2   History

Under the Freedom of Information Act (FOIA), Reclaim the Records requested VA release the BDF. It was noted in the lawsuit the BDF was already available online via Ancestry.com and Fold3.com. On January 9, 2018, VBA's FOIA Officer, (b)(6)   , denied the request because the records "contained information that falls within the disclosure protections of FOIA Exemption 6." On April 26, 2018, the Veterans Health Administration (VHA) responded that as the proper data custodian, they would process the FOIA. On July 26, 2018, VHA denied the FOIA request and stated the voluntary release to Ancestry.com was "erroneous" and resulted in a "data breach of Veteran's personal information." The court found the FOIA statute does not give an agency license to broadly withhold non-exempt records because the agency has errantly comingled them with exempt records. The court ordered VA to comply with the release of the BDF by April 1, 2022, a two-year timeline to meet the deadline.

On June 10th, 2020, a new BDF was extracted and provided to VBA's OBI Data Quality (DQ) Team by the Office of Performance Analysis and Integrity (PA&I). The directive was to provide records only for individuals that had died prior to the ruling with no "Evidence of Life". Since that date, OBI's DQ Team has been improving the methodology used to eliminate the possibility of improper PII release by establishing thresholds to redact/remove records that did not meet criteria to yield a high confidence the Veteran was deceased.

3 Methodology

The BDF was used as the base dataset, as directed.

The initial **Methodology #1** from June 2020 to January 2021 utilized a "proof of death" approach by comparing Dates of Death in the BDF and the Corporate Database (CorpDB). Python computer code was developed to query the BDF which produced initial "Determined as Deliverable" records at ~96%. During the evaluation of this methodology, additional rules were implemented to account for social security number (SSN) mismatches which instilled the caveat that the Date of Death must be present in both systems. This refined **Methodology #2** using a "proof of death" approach dropped the yield of "Determined as Deliverable" to 45%. There were over 9M records that didn't have SSN matches to the CorpDB. This occurs due to CorpDB records only being created if a claim is processed, and many of the Veterans in the BDF had been deceased prior to the launch of that system, therefore, would not have a record in CorpDB.

In March 2021, the methodology underwent further revision and **Methodology #3** accounted for Veterans of the World Wars not being in CorpDB and the records were "Determined as Deliverable" if the BDF Date of Death was mature (greater than one year past March 2020). As the script was updated and analyzed, the Master Person Index (MPI), VA's authoritative identify service, was incorporated into **Methodology #4**. In this version, the python computer script searched for Dates of Birth and Dates of Death and compared them between the BDF, CorpDB, and MPI. For any given record, if either of the two sources did not have a corresponding record, then that system was ignored for that record; however, if there was a record present, then Dates of Birth and Dates of Death had to agree for the record to be "Determined as Deliverable".

This methodology continued until October 2021 when several runs of query made it evident errors in each of the datasets queried were being compounded impacting the final yield. OBI's DQ team re-evaluated the dataflows and flags (inputs/outputs) and determined the best approach that negated compounding the inherent individual system data errors and lag or lack of data between systems to validate information, would be to scan the BDF for "Evidence of Life" as directed in the court order, as opposed to "Corroborating Information of Death". Furthermore, it was noted "notifications of death" (via the Social Security Administration, state coroners, National Cemetery Administration, etc.) reported to the MPI were propagated to CorpDB anyway, so its inclusion was redundant.

In the final **Methodology #5,** utilized from October 2021 to present**,** MPI was removed as a data source to eliminate compounded data errors from individual datasets. The foundation for this methodology is that when a Veteran has a halt in benefits due to incorrectly being labeled as "deceased", they would be required to contact the VA to update their status to "living", which would be done in the CorpDB not MPI. This aligned with the court order for identifying "Evidence of Life", assuming the BDF was correct, ignoring any further investigation into records where CorpDB agreed with the Date of Death, and redacting only those records who demonstrated the following "signs of life":

a. Filing for a claim with VA
b. Reversal of a Date of Death notification
c. Any call or correspondence with VA

The individual methodologies and search criteria are incorporated into **Table 1,** and the more specific criteria that were explored as listed in **Table 2**.

4 Results

OBI's DQ team, after careful evaluation of the BDF using several methodologies, determined the best approach to avoid compounding dataset errors and targeting the objective of not releasing PII on living Veterans, adopted **Methodology #5** which scanned for "Evidence of Life" as per the court order.
The BDF contains a total of 18,538,702, of which 1,350 (0.037%) were redacted due to "Evidence of Life" as determined by the three parameters for "Evidence of Life" noted above in methodology.

Methodology Comparison					
Table 1	Methodology #1	Methodology #2	Methodology #3	Methodology #4	Methodology #5 "Final"
Time Frame	6/20-01/21	02/21-03/21	03/21-06/21	06/21 - 10/21	10/21 - present
BIRLS Death File (BDF)	✓	✓	✓	✓	✓
Master Person Index (MPI)			✓		
Corporate Database (CorpDB)	✓	✓	✓	✓	✓
Criteria 1	✓				
Criteria 2		✓			
Criteria 3 "3rd Data Source-MPI"			✓		
Criteria 4 "Proof of Deceased"				✓	
Criteria 5 "Evidence of Life"					✓
Number of Records Redacted	0.22%/42K 3.43%/70K Undetermined	~277K/1.5% 53.4%/9.9M Undetermined	~1M / 6.0%	~180K/ 1.0%	1,350 / 0.004%

Table 2	Methodology Criteria
Criteria 1	**Deliverable Criteria Methodology 1:** 1) Presumed Deceased (age > 122yrs) 2) BIRLS DOD matched to CorpDB
Criteria 2	**Deliverable Criteria Methodology 2:** 1) No SSN Match Between BIRLS & CorpDB, but DOD in BIRLS Is Mature (DOD < March 2020) 2) SSN Match between BIRLS & CorpDB, DOB must match in CorpDB
Criteria 3	**Deliverable Criteria Methodology 3:** 1) MPI added as additional system for comparing death indication data 2) Date of Birth and Date of Death (within one month) had to match in all three systems in order to be released. 3) If a record did not exist in a given system, that system was ignored for that specific record.
Criteria 4	**Deliverable Criteria Methodology 4: " Proof of Death"** MPI determined to be redundant and untrustworthy for death information, so it was removed as a system of comparison. 1) BDF records compared only to CorpDB as a corroborating data source. 2) If CorpDB did not have a corresponding record, BDF was assumed to be accurate, and record marked for released. 3) If CorpDB did have a corresponding record with a matching date of death, record was marked for release. 4) If CorpDB did have a corresponding record, but without a matching date of death, record was flagged to be redacted.
Criteria 5	**Deliverable Criteria Methodology 5: "Evidence of Life"** It was noted that the vast majority of records in CorpDB without a recorded date of death were "shell" records, containing little to no information. Any contact with the VA by a living Veteran would trigger activity on that Veterans record, so posture was moved from "Evidence of Death" to "Evidence of Life", defined as 1) Filing for a claim with VA 2) Reversal of a Date of Death notification 3) Any call or correspondence with VA