

ADDENDUM B

Phase II – Updated Document Indexing Recommendation and Updated Data Field Extraction List

Table of Contents

Revision History	ii
1 Purpose.....	1
2 Core Document Indexes.....	1
2.1 File Number	1
2.2 Claimant Name	2
2.3 Date of Receipt	3
2.4 Document Type	3
3 Data Field Extraction.....	4
3.1 Contentions	4

1 Purpose

This document is in response to the Performance Work Statement (PWS) Sections 4.3.2 and 4.3.3 which state:

- Section 4.3.2 - the Contractor shall re-evaluate the forms and documents, and identify the fields to be automatically extracted for document indexing. At a minimum, each document must be indexed by claim number, veteran full name, document type and category, and date of receipt. The Contractor must update the Document Indexing Recommendation within 45 calendar days of receipt of updated forms or receipt of additional forms supplied as GFI.
- Section 4.3.3 – the Contractor shall re-evaluate the necessary VETSNET data fields (which shall be provided) and claims forms and documents, and identify the fields to be automatically extracted for database update. The Contractor shall update the Data Field Extraction List within 45 calendar days of receipt of updated forms or receipt of additional forms supplied as GFI.

To meet these requirements, the Phase I Document Indexing and data field extraction scope was discussed with VBA staff and confirmed to be accurate as of the date of this document. The VA is currently evaluating claims processes and forms for additional document indexing and data field extraction requirements. This document will be updated following any changes to the document indexing and/or data field extraction requirements.

2 Core Document Indexes

All documents will be indexed with a minimum of four core metadata or index values: *VA File Number*, *Claimant Name*, *Date of Receipt*, and *Document Type*. *Claimant Name* is comprised of *First Name*, *Middle Initial*, and *Last Name* fields. The *File Number* and *Claimant Name* will be extracted from the claims file binder or document coversheets. *Date of Receipt* and *Document Type* will be determined based on document content.

2.1 File Number

A single numeric file number value will be used for all claims documents. A high degree of variability exists to reference claim numbers on individual documents. Often different numbers, number formats, and reference labels will appear depending on the type and age of the document. Reference number variability is the greatest for older veterans. The scanning operations will utilize the file number as it appears on the outside of claims file binders or document coversheet. In most cases this will be the veteran SSN or internally VA generated “C Number”. NARA will deliver a single file number (9 digit SSN or 8 digit C Number). The single file number will be assigned to all documents within the binder. Document level reference numbers such as SSNs, C Numbers, etc. will not be extracted from the individual documents.

Field Name: FILE_NBR

Description: A 9 or 8 digit reference number which identifies the claimant and benefit claim.

Purpose: Route document metadata and images to the appropriate FileNet folder.

Size: 8-9 numbers

Format: VARCHAR2

2.2 Claimant Name

A single *Claimant Name* (i.e. Veteran's name) will be used for all veteran claims file documents. First, middle, and last name will be provided separately. The scanning operations will utilize the *Claimant Name* as it appears on the outside of claims file binders or document coversheet. Middle names will be specified as blank or only as the initial when the full middle name is not present. NARA will deliver all three values as 30 character fields. The single *Claimant Name* values will be assigned to all documents within the binder. Document level *Claimant Names* will not be extracted from the individual documents.

Field Name: FIRST_NM

Description: First name as it appears on the outside of claims files binders or document coversheets.

Purpose: Route document metadata and images to the appropriate FileNet folder.

Size: 30 characters

Format: VARCHAR2

Field Name: MIDDLE_NM

Description: First name as it appears on the outside of claims files binders or document coversheets.

Purpose: Route document metadata and images to the appropriate FileNet folder.

Size: 30 characters

Format: VARCHAR2

Field Name: LAST_NM

Description: First name as it appears on the outside of claims files binders or document coversheets.

Purpose: Route document metadata and images to the appropriate FileNet folder.

Size: 30 characters

Format: VARCHAR2

2.3 Date of Receipt

Date of Receipt will be determined by valid date stamps appearing on documents, or extracted from documents according to procedures defined in the VBMS Scanning Business Rules document (refer to Appendix C). A single date will be provided for each document. The VBMS Scanning Business Rules generally define procedures differently for internally generated VA documents and documents received from Veterans or third parties. *Date of Receipt* for externally generated documents (those submitted by veterans or third parties) is determined by the

earliest of the valid date received stamp applied to a form, document, or document grouping. The earliest date is identified to provide the time advantage to veterans and applicants. Multiple date stamps may be applied if documents are received and then routed among different VA/VBA organizations. VBA organizations are required to follow specific specifications when ordering and procuring date stamp equipment. Considerable variance does occur in the date stamp look and placement. Only “Official” VA Date Stamps (as defined in the Scanning Business Rules) are recognized for determining Date of Receipt. These organizations include VARO, VSC, VA, Veterans Administration, RMC, and BDD intake sites. Determination of *Date of Receipt* will be validated manually to ensure accuracy relative to VBA requirements to capture earliest date. Only a small percentage of documents are not expected to have a date stamp or other date indicated. VBA procedures are to apply date from adjacent documents in folder or batch of documents, whichever is applicable. Scanning Business Rules and system checks will be evaluated and implemented as appropriate to prevent erroneous, invalid, or future dates used as Date of Receipt.

Field Name: DATE_RCVD

Description: Date of receipt identifies when a VA document was created or when a document was first received by the VA. Date of Receipt is critical for business workflow and reporting.

Purpose: Reference date for documents.

Size: 8 characters

Format: DATE8

2.4 Document Type

Document Type ID codes will correspond to short “Form Number” style code. Document Type codes correspond to the codes listed in the Phase II Deliverable – Updated Document Classification and Extraction Listing.

Field Name: DOC_TYPE

Description: Determines lowest level of document type.

Purpose: Used for search and retrieval, reporting, and associating higher level document attributes.

Size: 12 characters

Format: VARCHAR2

3 Data Field Extraction

Document specific data field extraction was reviewed and assessed during the Phase I Document Taxonomy work effort. Form 21-526 was the only document type that was configured in the VBMS Scanning System for data field extraction. Contention data fields are currently being extracted for Form 21-526 and configuration work is being completed to extract contention data fields from Forms 21-526c and 21-0819. Four additional document types were discussed in Phase I (Forms 21-22, 21-4138, 21-4142, and DD-214) and a “Wish List” of data fields were doc-

umented for each document type. NARA is waiting for futher direction on data field extraction priorities and formats in Phase II.

3.1 Contentions

Contention values are gathered primarily from versions of initial claims forms (Forms VA 21-526, 21-526c, 21-0819). The number of contentions is dependent upon the version of the 526 with a maximum of 9 contentions allowed. Contention values are to be extracted and / or entered exactly as entered on the form and not corrected for grammar or spelling.

Field Name: CONTENTION_1

Description: Value represents what the veteran entered in the 1st contention field

Purpose: Used to document the 1st contention

Size: 255

Format: VARCHAR2

Field Name: CONTENTION_2

Description: Value represents what the veteran entered in the 2nd contention field

Purpose: Used to document the 2nd contention

Size: 255

Format: VARCHAR2

Field Name: CONTENTION_3

Description: Value represents what the veteran entered in the 3rd contention field

Purpose: Used to document the 3rd contention

Size: 255

Format: VARCHAR2

Field Name: CONTENTION_4

Description: Value represents what the veteran entered in the 4th contention field

Purpose: Used to document the 4th contention

Size: 255

Format: VARCHAR2

Field Name: CONTENTION_5

Description: Value represents what the veteran entered in the 5th contention field

Purpose: Used to document the 5th contention

Size: 255

Format: VARCHAR2

Field Name: CONTENTION_6

Description: Value represents what the veteran entered in the 6th contention field

Purpose: Used to document the 6th contention

Size: 255

Format: VARCHAR2

Field Name: CONTENTION_7

Description: Value represents what the veteran entered in the 7th contention field

Purpose: Used to document the 7th contention

Size: 255

Format: VARCHAR2

Field Name: CONTENTION_8

Description: Value represents what the veteran entered in the 8th contention field

Purpose: Used to document the 8th contention

Size: 255

Format: VARCHAR2

Field Name: CONTENTION_9

Description: Value represents what the veteran entered in the 9th contention field

Purpose: Used to document the 9th contention

Size: 255

Format: VARCHAR2