

**DATA EXTRACT GUIDE**

Ownership

Client Name	VITA
Data Contact	Carrie McDermott/Rudy Faas

Revision History

Version	Date	Author/ Editor	Summary Major Changes																																													
v 1.0	April 9, 2012	Joe Calleroz	Initial baseline document created																																													
v 2.0	July 16, 2012	Joe Calleroz/ Steve Radecki	Final attributes added; added revision history; added clarity re: level qualifiers in file characteristics and assumptions																																													
v 3.0	October 26, 2012	Steve Radecki	Added verbiage regarding removal of Shell, impostor and Trusts, etc. from source systems  Documented that users starting with 999 in SSN will be removed by the data load process																																													
v 4.0	November 7, 2012	Steve Radecki	Changed the following "total existing values" for the following fields. It does not affect the extract files, only Initiate settings.  <table border="0"> <tr> <td>Gender</td> <td>1</td> <td>1</td> </tr> <tr> <td>Gender Level Qualifier</td> <td>10</td> <td>1</td> </tr> <tr> <td> </td> <td></td> <td></td> </tr> <tr> <td>Birth date</td> <td>1</td> <td>1</td> </tr> <tr> <td>Birth date Level Qualifier</td> <td>10</td> <td>1</td> </tr> <tr> <td> </td> <td></td> <td></td> </tr> <tr> <td>Medicaid ID</td> <td>1</td> <td>1</td> </tr> <tr> <td>Medicaid ID Level Qualifier</td> <td>10</td> <td>1</td> </tr> <tr> <td> </td> <td></td> <td></td> </tr> <tr> <td>Death indicator</td> <td>1</td> <td>1</td> </tr> <tr> <td>Death indicator Level Qualifier</td> <td>10</td> <td>1</td> </tr> <tr> <td> </td> <td></td> <td></td> </tr> <tr> <td>Out of State Indicator</td> <td>1</td> <td>1</td> </tr> <tr> <td>Out of State Indicator Level Qualifier</td> <td>10</td> <td>1</td> </tr> <tr> <td>Level Qualifier</td> <td></td> <td></td> </tr> </table>	Gender	1	1	Gender Level Qualifier	10	1	 			Birth date	1	1	Birth date Level Qualifier	10	1	 			Medicaid ID	1	1	Medicaid ID Level Qualifier	10	1	 			Death indicator	1	1	Death indicator Level Qualifier	10	1	 			Out of State Indicator	1	1	Out of State Indicator Level Qualifier	10	1	Level Qualifier		
Gender	1	1																																														
Gender Level Qualifier	10	1																																														
Birth date	1	1																																														
Birth date Level Qualifier	10	1																																														
Medicaid ID	1	1																																														
Medicaid ID Level Qualifier	10	1																																														
Death indicator	1	1																																														
Death indicator Level Qualifier	10	1																																														
Out of State Indicator	1	1																																														
Out of State Indicator Level Qualifier	10	1																																														
Level Qualifier																																																
v 5.0	December 7, 2012	Carrie McDermott	Clarification additions. No structural changes were made to the document. Cosmetic updates only.																																													
v 6.0	January 17, 2013	Carrie McDermott	Last Update Date – Added 'Time' stamp and made a required field  Added a new indicator in first position – 'Delete Indicator' is a required field																																													
v 7.0	July 7, 2013	Carrie McDermott	Ownership box  Out of State Indicator– acceptable values "Y" or "N" Death indicator – acceptable values "Y" or "N"  Line #88 – Last Update User - changed reference number to #87  A <b>*Note</b> was added to the "Sample Data" area  DMAS removed as Source of Data  Removed IBM Security section and IBM Customization section.																																													
v 8.0	August 21, 2013	Carrie McDermott	A new "Assumption" was added – No double quotes in data allowed.																																													

v9.0	March 13, 2014	Carrie McDermott	<p>The following items have been included in this guide:</p> <ul style="list-style-type: none"> <li>- Acceptable date/time values</li> <li>- Matching fields are: <ul style="list-style-type: none"> <li>Name</li> <li>Address</li> <li>Phone</li> <li>DOB</li> <li>SSN</li> <li>DL#</li> <li>Medicaid ID</li> <li>Gender</li> <li>Email</li> </ul> </li> <li>- Matching caveat</li> </ul> <p><b>*Note</b> – Matching will depend heavily on how well populated the matching fields are. If matching is not taking place to the extent that suits the Data Stewards, adjustments will need to be made.</p> <ul style="list-style-type: none"> <li>- LASTUPDATEUSER will now be a required field.</li> </ul>
------	----------------	------------------	---

### Sources of Data

The following are the customer data sources and descriptions:

Data Source (Vendor, Source, Version)	Description	Client Expected Record Count (Not to Exceed)
DMV	Department of Motor Vehicles	

### Structure of Extract

Question	Description	Response
What is your primary identifier?	This is the number used to uniquely identify a person. For example, it may be the MRN, Corporate ID, or SSN. <i>Briefly describe- note characters.</i>	
Does the primary ID have a meaningful prefix, suffix, or any other characteristic within the identifier?	Describe any values in the identifier that represent a distinct population, such as a facility, type of customer, or region.	
How is the primary ID assigned?	Indicate if the primary ID is assigned serially (sequentially), pulled from a pool or block, or created algorithmically. <i>Describe if block or algorithm.</i>	
Can you extract one record per primary identifier?	When you perform your data extract, are you able to pull one record per unique identifier?	
Do you assign a secondary identifier?	For example, a Corporate ID where your local source ID would be the primary, or reverse (for integration, reporting, et cetera)? <i>Describe.</i>	
Do you have multiple facilities (i) sharing the same unique ID or (ii) assigning a unique ID by facility?	In either one of these scenarios, a Corporate ID is being assigned and may or may not be known to the user, but the Corporate ID is linking the same-source, different facility identifiers?	
Do you have multiple facilities assigning a unique ID by facility?	In this scenario, an Enterprise ID is being assigned	
Do multiple sources contain a shared identifier?	If you have a system-wide identifier to link customers across multiple systems, IBM can assess the accuracy of these "pointers".	
Do you have records without a primary identifier?	If you wish to include these in your evaluation, a primary identifier needs to be assigned prior to submitting the file, or IBM can assign an identifier during processing. <i>Briefly describe.</i>	
Can the source file contain "merged" records?	Customer acknowledges that merged records will be removed prior to sending the file to IBM.	

## File Characteristics

The information used to relate records is comprised of identifying information contained within source files. Formats and attribute content should conform to the specifications outlined in this guide. The listed required attributes are necessary for member matching. The richness of the attributes and the degree to which they are populated factors into the strength of our linking and matching. Additional attributes aid in validating records and provides meaningful context to the data.

- Please refer to the following rules re: population of level qualifiers:

- If there is a value in a field relating to an attribute – always send a corresponding value in level qualifier
- If there is NOT a value in a field relating to an attribute – do NOT send a corresponding level qualifier
- For those level qualifiers that relate to multiple fields (i.e. name and address), if all fields are NULL, do NOT send a level qualifier (if any one of the fields in group are populated, the associated level qualifier should contain a value)
- In all instances, the custom composite view will always render the attribute with the highest level qualifier (irrespective of whether or not the attribute is populated) - A value will always supersede a blank

- Changes were made to the source systems to remove extraneous data such as shell records, trusts, and impostors:

- The shell records were identified with data with prefixes and 11-digit numbers
  - T281
  - T286
  - T197
- Impostors had the word 'impostor' in the last name field
- Trusts had the word 'trust' in the last name'

- Members sent from source systems beginning w/ '999' in the social security system will be removed from the extracts prior to load into the Initiate Hub, by the loading process

The following is the agreed upon format of the source data for all your sources:

For any data fields that have repeating values, the data should be included in a separate file. For example, if you have more than one Previous Name for a Patient/Customer record, you should include the Previous Name in a different file with the Source|Enterprise ID|Previous Name fields.

Position	Field	Description	Active Values	Total Existing Values	Max Length	Format
1.	Delete Indicator*	Identifies a record to be deleted	Y – Yes N - No	2	1	Text
2.	Source*	Identifies source of data	N/A	N/A	40	Text
3.	Unique ID*	Unique ID to source	N/A	N/A	60	Text
4.	First Name	Customer first name	1	10	30	Text
5.	Middle Name	Customer middle name	1	10	30	Text
6.	Last Name	Customer last name	1	10	75	Text
7.	Suffix	Customer Suffix (Jr., Sr.), This can be included with the Last Name also, if that's how the client has their data. [Optional]	1	10	10	Text
8.	Person Name Validation Level Qualifier	Person Name Validation Level Qualifier	1	10	3	Number
9.	Gender	Gender	1	1	10	Text
10.	Person Gender Validation Level Qualifier	Person Gender Validation Level Qualifier	1	1	3	Number
11.	Birth Date	Birth Date	1	1	19	YYYYMMDD
12.	Person Birth Date Validation Level Qualifier	Person Birth Date Validation Level Qualifier	1	1	3	Number
13.	Physical Address (Address Type Code)	Physical Address Type Code	1	10	3	Text
14.	Physical Address (Stline1)	Physical Address Line 1	1	10	75	Text
15.	Physical Address (Stline2)	Physical Address Line 2	1	10	75	Text

16.	Physical Address (City)	Physical Address City	1	10	30	Text
17.	Physical Address (State)	Physical Address State	1	10	15	Text
18.	Physical Address (Zipcode)	Physical Postal code (including extension)	1	10	15	Text
19.	Physical Address (Country)	Physical Address Country	1	10	50	Text
20.	Physical Address (State-Locality Code)	Physical Address State-Locality Code	1	10	5	Text
21.	Physical Address Validation Level Qualifier	Physical Address Validation Level Qualifier	1	10	3	Number
22.	Mailing Address (Address Type Code)	Mailing Address Type Code	1	10	3	Text
23.	Mailing Address (Stline1)	Mailing Address Line 1	1	10	75	Text
24.	Mailing Address (Stline2)	Mailing Address Line 2	1	10	75	Text
25.	Mailing Address (City)	Mailing Address City	1	10	30	Text
26.	Mailing Address (State)	Mailing Address State	1	10	15	Text
27.	Mailing Address (Zipcode)	Mailing Address Postal code (including extension)	1	10	15	Text
28.	Mailing Address (Country)	Mailing Address Country	1	10	50	Text
29.	Mailing Address (State-Locality Code)	Mailing State-Locality Code	1	10	5	Text
30.	Mailing Address Validation Level Qualifier	Mailing Address Validation Level Qualifier	1	10	3	Number
31.	Other Address (Address Type Code)	Other Address Type Code	1	10	3	Text
32.	Other Address (Stline1)	Other Address Line 1	1	10	75	Text
33.	Other Address (Stline2)	Other Address Line 2	1	10	75	Text
34.	Other Address (City)	Other Address City	1	10	30	Text
35.	Other Address (State)	Other Address State	1	10	15	Text
36.	Other Address (Zipcode)	Other Address Postal code (including extension)	1	10	15	Text
37.	Other Address (Country)	Other Address Country	1	10	50	Text
38.	Other Address (State-Locality Code)	Other Address State-Locality Code	1	10	5	Text
39.	Other Address Validation Level Qualifier	Other Address Validation Level Qualifier	1	10	3	Number
40.	Alias First Name	Alias first name	1	10	30	Text
41.	Alias Middle Name	Alias middle name	1	10	30	Text
42.	Alias Last Name	Alias last name	1	10	75	Text

43.	Alias Suffix	Alias Suffix (Jr., Sr.), This can be included with the Last Name also, if that's how the client has their data. [Optional]	1	10	10	Text
44.	Alias Validation Level Qualifier	Alias Validation Level Qualifier	1	10	3	Number
45.	Social Security Number assigning authority	Assigning authority code for Social Security Number (e.g. SSA)	1	1	40	Text
46.	Social Security Number	Social Security Number as assigned by assigning authority	1	1	60	Text
47.	Social Security Number Validation Level Qualifier	Social Security Number Validation Level Qualifier	1	1	3	Number
48.	Drivers License Number assigning authority	Assigning authority code for Drivers License Number (e.g. VA DMV)	1	1	40	Text
49.	Drivers License Number	Drivers License Number as assigned by assigning authority	1	1	60	Text
50.	Drivers License Number Validation Level Qualifier	Drivers License Number Validation Level Qualifier	1	1	3	Number
51.	Medicaid ID	Medicaid ID	1	1	60	Text
52.	Medicaid ID Validation Level Qualifier	Medicaid ID Validation Level Qualifier	1	10	3	Number
53.	Primary (Phone Country Code)	Primary Phone Country Code	1	10	5	Text
54.	Primary (Phone Area Code)	Primary Phone Area Code	1	10	3	Text
55.	Primary (Phone Number)	Primary Phone Number	1	10	20	Text
56.	Primary (Phone Extension)	Primary Phone Extension	1	10	15	Text
57.	Primary (Phone Number Validation Level Qualifier)	Primary Phone Number Validation Level Qualifier	1	10	3	Number
58.	Alternate (Phone Country Code)	Alternate Phone Country Code	1	10	5	Text
59.	Alternate (Phone Area Code)	Alternate Phone Area Code	1	10	3	Text
60.	Alternate (Phone Number)	Alternate Phone Number	1	10	20	Text
61.	Alternate (Phone Extension)	Alternate Phone Extension	1	10	15	Text
62.	Alternate (Phone Number Validation Level Qualifier)	Alternate Phone Number Validation Level Qualifier	1	10	3	Number
63.	Person Citizenship (Code)	Person Citizenship Code	1	10	3	Text
64.	Person Citizenship (Validation Level Qualifier)	Validation Level Qualifier	1	10	3	Number
65.	Personal Email	Personal Email Address	1	10	75	Text
66.	Personal Email (Validation Level Qualifier)	Personal Email Validation Level Qualifier	1	10	3	Number
67.	Work Email	Work Email Address	1	10	75	Text
68.	Work Email (Validation Level Qualifier)	Work Email Validation Level Qualifier	1	10	3	Number
69.	Other Email	Other Email Address	1	10	75	Text
70.	Other Email (Validation Level Qualifier)	Other Email Validation Level Qualifier	1	10	3	Number

71.	Person Ethnicity (Identifying Code)	Person Ethnicity Identifying Code	1	10	3	Text
72.	Person Ethnicity (Validation Level Qualifier)	Person Ethnicity Validation Level Qualifier	1	10	3	Number
73.	Person Race (Identifying Code)	Person Race Identifying Code	1	10	3	Text
74.	Person Race (Validation Level Qualifier)	Person Race Validation Level Qualifier	1	10	3	Number
75.	Person Physical Location	Person Physical Location	1	10	250	Text
76.	Person Physical Location (Validation Level Qualifier)	Person Physical Location Validation Level Qualifier	1	10	3	Number
77.	Person Place of Birth (State)	Person Place of Birth State	1	10	15	Text
78.	Person Place of Birth (Country)	Person Place of Birth Country	1	10	50	Text
79.	Person Place of Birth (Validation Level Qualifier)	Person Place of Birth Validation Level Qualifier	1	10	3	Number
80.	Mother Birth Name	Mother Birth Name	1	10	75	Text
81.	Mother Birth Name (Validation Level Qualifier)	Mother Birth Name Validation Level Qualifier	1	10	3	Number
82.	ConnectVirginia HIE Participant Status	ConnectVirginia HIE Participant Status - Placeholder for now	1	10	3	Text
83.	Date of Death	Date of Death	1	1	19	YYYYMMDD
84.	Date of Death (Validation Level Qualifier)	Date of Death Validation Level Qualifier	1	1	3	Number
85.	Death Indicator	Death Indicator	1	1	3	Y or N
86.	Death Indicator Validation Level Qualifier	Death Indicator Validation Level Qualifier	1	1	3	Number
<b>87.</b>	<b>Last Update Date and Time*</b>	Last Update Date and Time	1	1	19	YYYYMMDD HHMMSS
<b>88.</b>	<b>Last Update User</b>	Last Update User (associated with the item #87)	1	1	100	Text
89.	Out of State Indicator	Out of State Indicator	1	1	3	Y or N
90.	Validation Level Qualifier	Out of State Indicator Validation Level Qualifier	1	1	3	Number

\* Excludes non-surviving/merged records

#### Assumptions

- **\*Bold fields are required**
- Source ID is unique within the source
- Standard data load files for the IBM™ software are pipe { | } delimited and include one header record with the source and extract date.
- Files are flat and will have Carriage Return Line Feed (CRLF) present at the end of each line.
- Files are ASCII unless otherwise specified
- Delimited files require field placement
- All dates are in YYYYMMDD format
  - YYYY – four digit year
  - MM – two digit month of the year
  - DD – two digit day of the month
- All times are in HHMMSS format
  - HH – two digit hour in day – military format (00-23)
  - MM – two digit minute in hour (00-59)
  - SS – two digit second in minute (00-59)
- Max Length indicates the maximum allowable length for data in the field;
- If there is a value in a field relating to an attribute – always send a corresponding value in level qualifier

- If there is NOT a value in a field relating to an attribute – do NOT send a corresponding level qualifier
  - For those level qualifiers that relate to multiple fields (i.e. name and address), if all fields are NULL, do NOT send a level qualifier (if any one of the fields in group are populated, the associated level qualifier should contain a value)
  - In all instances, the custom composite view will always render the attribute with the highest level qualifier (irrespective of whether or not the attribute is populated) - A value will always supersede a blank.
  - No double quotes in data
  - The Initiate matching fields are:
    - Name (name as well as alias)
    - Address (mailing, other, and physical)
    - Phone (primary and alternate)
    - Date of Birth
    - Social Security Number
    - Drivers License Number
    - Medicaid ID
    - Gender
    - Email (primary, other, work)
- \*Note – Matching will depend heavily on how well populated the matching fields are.

**Data Characteristics**

Data fields should be:

- Alphanumeric (A) characters, Left-justified
- Number (N) digits, Right-justified

**Sample Data**

The following is a representative sample of the data that you provide to VITA:

\*Note: One delimiter following each field including the last field is required

[SOURCE1| 1001|KIRK|JAMES|T|234545565|10161967|203 West Madison Ave|Suite 123|Chicago|IL|61606|A|M|  
 [SOURCE1| 1021|SPADE|DAVID|T|234545565| 10161967|20 W NORTH Ave| |Chicago|IL|61606|A|M|

**\*\*Preventing Extract Errors\*\***

*We do not process data that fail to meet the described criteria for Position, Length, Format, or Validation. We recognize that any large volume of data is likely to have unforeseen characteristics, so we recommend that you perform thorough quality assurance on your extracts before delivering them to us. Preliminary validation guidelines are included in the checklist below.*

*When we encounter errors we provide you a reject log of the records we could not process. If there are a significant number of errors (greater than 0.05%) you may be required to provide another extract.*

**Reprocessing your data due to excessive extract errors may impact the project schedule and/or price.**

**Data Submission Checklist**

When we receive your data we perform the following series of checks to ensure that the quality, size, and format of the data are what we expected and agreed upon:

Step	Description	Pass
Readable	Are the files readable? If they are zipped, can they be extracted and read in a plain text viewer?	<input type="checkbox"/> Yes <input type="checkbox"/> No
File counts	What is the record count of each file? Does that match our expected record count?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Primary ID	Is the primary identifier unique throughout the source file?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Date Format	Does the date format for all date fields match our expected date format, including only valid dates?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Name Format	Is the name format divided into first, middle, and last?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Non-printable characters	Does the file contain printable characters only? In other words, it does not include any non-printable characters.	<input type="checkbox"/> Yes <input type="checkbox"/> No
End of line terminator	Does the end of line terminator match the CRLF that we are expecting?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Format	Does the file adhere to the agreed upon format? Is the correct delimiter used ( )?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Source	Does the file contain a column that identifies each record as pertaining to a particular source?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Null value	Character(s) are not used to indicate a null value in any field.	<input type="checkbox"/> Yes <input type="checkbox"/> No
Extra characters	Are there any extra characters or fields that we did not agree to in the format?	<input type="checkbox"/> Yes <input type="checkbox"/> No

Any indication of a 'No' in the questions above may require you to provide a new source extract.