

Announcement: Data Quality Improvements to the SAM Public Exclusions File

As part of regular data quality improvements to SAM interfaces, we have improved the quality of data that we are providing in the public exclusions file. The new file will be posted within a week of this email.

Attached is the detailed gap analysis report between current file and the new file that we will start deploying on January 13, 2014.

The Gap analysis report is below. The first four columns describe the field name (attribute) and the current data quality issue with examples. The last two columns detail the data quality improvement and the last column provides an example of the improvement.

Summary of Gap of the gaps analysis report:

Public Exclusions File Attribute	Approximate number of records improved	Data Quality issue with current public exclusions file	Example of bad data in current public exclusions file	Data Quality improvement made in the new public exclusions file.	Example of improved data in new public exclusions file
ACTIVE_DATE	43	Reads date as it is	07/27/0012	Fixes bad year like 0999 to 1999	7/27/2012
ADDITIONAL_COMMENTS	52194	1. Keeps End of line character 2. No special character conversion	Abel Martin Carreon, Jr., President, Federal Builders, Incorporated See also Carreon, Abel Inc.	End of line character removed. Special characters converted.	Abel Martin Carreon, Jr., President, Federal Builders, Incorporated See also Carreon, Abel Inc.
ADDRESS_LINE_1	1	1. No special character conversion 2. No case change	Blvd. García Morales No. 257	1. End of line character removed. Special characters converted. 2. Case changed to upper	BLVD. GARCIA MORALES NO. 257
ADDRESS_LINE_2	14	Reads address line 2 as is	Brazilia Building	Concatenates address line 2 and 3	BRAZILIA BUILDING 1ST FLOOR
CITY	112	Read data as it is	Null	Replaces explicit "Null" with null value	
CROSS_REFERENCE	2071	1. Reads data from both older and new version of records in SAM (obsolete records) 2. Mixed case data can cause matching issues.	(also Bernard Grobart, Bernard Grobart, Bernard Grobart)	1. Reads data from active / expired latest version of record in SAM. 2. Case change to upper	(also BERNARD GROBART)
DUNS	376	Published U* duns number (These are not duns numbers)	U00000145	Suppresses U* duns numbers	(Null value)
EXCLUDING_AGENCY	2569	Read data as it is	(Space)	Populated missing excluding agency based on reference table.	DOL
FIRST	416	1. Read data as it 2. No case change	Mary Ann	1. Trims values and remove excess spaces	MARY ANN

				2. Case change to upper	
LAST	119	1. Read data as it 2. No case change	Salinas	1. Trims values and remove excess spaces 2. Case change to upper	SALINAS
MIDDLE	990	1. Read data as it 2. No case change	W III	1. Trims values and remove excess spaces 2. Case change to upper	W III
NAME	69	In certain cases the draft / work in progress SAM record is read.	BP CORPORATION NORTH AMERICA INC.	Reads data from active / expired latest version of record in SAM. Draft records in SAM are ignored.	AMOCO PETROLEUM PRODUCTS
PREFIX	1601	1. Read data as it 2. No case change	(Space)	1. Trims values and removes excess spaces 2. Case change to upper	(Null value)
STATE	4	Read data as it is	"Null" (Explicit Null)	Replaces explicit "Null" with null value	(Null value)
SUFFIX	1696	1. Read data as it 2. No case change	(Space)	1. Trims values and remove excess spaces 2. Case change to upper	(Null value)
ZIP	8940	Read 5 digit US zip codes	91320	Applies SAM Zip code standardization logic. US Zip codes only numeric, either 5 digit or 9 digit, and non-US zip codes are not touched.	913202157