



**DATAFLUX**  
data management

# What's New in DataFlux<sup>®</sup> Quality Knowledge Base for Contact Information 2013A



The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2013. *What's New in DataFlux® Quality Knowledge Base for Contact Information 2013A*. Cary, NC: SAS Institute Inc.

### **What's New in DataFlux® Quality Knowledge Base for Contact Information 2013A**

Copyright © 2013, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

**For a hard-copy book:** No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government Restricted Rights Notice:** Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc, SAS Campus Drive, Cary, North Carolina 27513.

SAS provides a complete selection of books and electronic products to help customers use SAS® software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit [support.sas.com/bookstore](http://support.sas.com/bookstore) or call 1-800-727-3228.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

# What's New in DataFlux Quality Knowledge Base for Contact Information 2013A

This document explains the changes in DataFlux Quality Knowledge Base (QKB) for Contact Information (CI) 2013A.

- [Overview](#)
- [Locale Support](#)
- [Deprecations](#)
- [Enhancements](#)

See the QKB CI online Help for additional information.

## Overview

QKB CI 2013A introduces support for the Hebrew, Israel language and locale.

Support is updated for address related definitions for the English, New Zealand language and locale.

See the [Enhancements](#) section for a detailed list of definitions that have been added, removed, or modified in this release. See the [Deprecations](#) section for a list of definitions that have been deprecated, renamed, or removed in this release.

Note that the file format of libraries in this release differs from the file format of libraries in releases earlier than QKB CI 2010B. Therefore, it is not possible to merge QKB CI 2013A into any QKB that was released prior to QKB CI 2010B. It is possible, however, to merge an earlier QKB into QKB CI 2013A after installing it into a separate location. See the "Merging with an Earlier QKB" section for details.

See the "Configuration Settings" section for recommendations regarding software settings to be used with definitions in this QKB.

## Locale Support

This release includes support for the following locales:

- Afrikaans, South Africa
- Arabic, Egypt (Partial Support; see QKB documentation)
- Chinese, China
- Czech, Czech Republic
- Danish, Denmark

- Dutch, Belgium
- Dutch, Netherlands
- English, Australia
- English, Canada
- English, Hong Kong (Partial Support; see QKB documentation)
- English, India
- English, New Zealand
- English, South Africa
- English, United Kingdom
- English, United States
- Finnish, Finland
- French, Belgium
- French, Canada
- French, France
- German, Germany
- Greek, Greece
- Hebrew, Israel
- Hungarian, Hungary
- Italian, Italy
- Japanese, Japan
- Korean, South Korea
- Malay, Malaysia
- Norwegian, Norway
- Polish, Poland
- Portuguese, Brazil
- Portuguese, Portugal
- Russian, Russia
- Slovak, Slovakia
- Spanish, Mexico
- Spanish, Spain
- Swedish, Sweden

- Thai, Thailand
- Turkish, Turkey

## Deprecations

DataFlux deprecates QKB definitions that are scheduled to be removed or renamed. We also deprecate definitions that use output tokens that are scheduled to be removed or renamed. Deprecated definitions are retained in the QKB for one release and are then removed from the QKB in the following release.

You might notice that some new definition names include a version tag denoting the QKB release in which the definition was introduced. This naming convention indicates that a definition is scheduled to be replaced by a new definition that uses an updated/renamed token list. For example, you might see definitions like the following in this release:

- Phone
- Phone (2013A)

These names indicate that the "Phone" definition will be replaced by the "Phone (2013A)" definition in a subsequent release. In that subsequent release, the "Phone" definition is replaced by the "Phone (2013A)" definition. The name remains simply "Phone", however. The definition with the version tag in its name, "Phone (2013A)", is then deprecated and removed in the following release:

- Phone
- Phone (2013A) → Phone
- Phone (2013A) → Phone (2013A) <deprecated>

The deprecated "Phone (2013A)" definition will then be removed in a future release.

The following is a list of definitions that were affected by this deprecation policy in the 2013A release.

Abbreviation	Description
Add	Definition added
Depr	Definition deprecated in this release
Repl	Definition replaced in this release
Same	Definition not changed in this release
C	Case Definition
G	Gender Analysis Definition
I	Identification Definition
M	Match Definition
P	Parse Definition
A	Pattern Analysis
S	Standardization Definition

*Legend for Type of Deprecated Definitions*

Notes on legend:

- Definitions marked "Add" are new definitions with updated tokens, match code lengths, or definition name changes. We recommend that you replace calls to current definitions with calls to the newly added definitions in all jobs. Note that names of newly added definitions often include a version tag such as "(2013A)".
- Definitions marked "Depr" are deprecated and will be removed or replaced in a future release. Note that names of deprecated definitions often include a version tag such as "(2010B)". We recommend that you remove calls to deprecated definitions and replace them with calls to the replacement definitions.
- Definitions marked "Repl" are definitions that are replaced by a copy of another definition, but have retained their original name. Since definitions that are replaced often have new or different tokens or match code lengths, you might need to update your job to use the enhancements if your job calls a definition that is replaced.
- Definitions marked "Same" are definitions that have not been significantly changed in this release but are related to added or deprecated definitions.

## English, New Zealand

Locale	Action	Type of Definition
Address	Depr	P S M
Address (2013A)	Add	P S M
City - State/Province - Postal Code	Depr	P S M
City - State/Province - Postal Code (2013A)	Add	P S M

*English, New Zealand*

## Enhancements

Note that definitions named with the suffix "(2013A)" are the newest definitions developed. We recommend that you use these new definitions when possible. Be aware, however, that some software version restrictions apply for some definitions. See the sections "Configuration Settings" and "Deprecations" for more details.

The following is a list of definitions that have been added, removed, or modified in this release. See the QKB CI 2013A online Help for details about these definitions.

Abbreviation	Description
New	Definition added
Mod	Definition modified
Rem	Definition removed
P	Parse Definition
M	Match Definition
S	Standardization Definition
G	Gender Analysis Definition

Abbreviation	Description
I	Identification Definition
C	Case Definition
A	Pattern Analysis Definition
E	Extraction Definition
N	Language Guess Definition
L	Locale Guess Definition

*Legend for Change Descriptions*

### **Afrikaans, South Africa**

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*Afrikaans, South Africa*

### **Arabic, Egypt**

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Arabic, Egypt*

### **Chinese, China**

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*Chinese, China*

### **Czech, Czech Republic**

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Czech, Czech Republic*

### **Danish, Denmark**

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Danish, Denmark*

## Dutch, Belgium

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Dutch, Belgium*

## Dutch, Netherlands

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Dutch, Netherlands*

## English, Australia

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*English, Australia*

## English, Canada

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*English, Canada*

## English, Hong Kong

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*English, Hong Kong*

## English, India

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*English, India*



## English, New Zealand

Locale	Action	Type of Definition
Address	Mod	P M S
Address (2013A)	New	P M S
Address (Extended)	Mod	M S
Address (Full)	New	P
Address (PO Box Only)	New	M
Address (Street Only)	New	M
Character (Script Identification)	Mod	A
City - State/Province - Postal Code	Mod	P M S
City - State/Province - Postal Code (2013A)	New	P M S
City - State/Province - Postal Code (Global)	Mod	P
Field Name	Mod	M I
Phone	Mod	I
Phone (2011A)	Mod	I
Word (Script Identification)	Mod	A

*English, New Zealand*

## English, South Africa

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*English, South Africa*

## English, United Kingdom

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*English, United Kingdom*

## English, United States

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*English, United States*

## Finnish, Finland

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Finnish, Finland*

## French, Belgium

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*French, Belgium*

## French, Canada

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*French, Canada*

## French, France

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*French, France*

## German, Germany

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Field Name	Mod	M I
Word (Script Identification)	Mod	A

*German, Germany*

## Greek, Greece

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Greek, Greece*

## Hebrew, Israel

Locale	Action	Type of Definition
ASCII Non-Printable Character Removal	New	S
Account Number	New	M
Address	New	P M S
Address (Detailed)	New	P
Address (Full)	New	P M
Address (Global)	New	P
Address (PO Box Only)	New	M
Address (Street Only)	New	M
Character	New	A
Character (Script Identification)	New	A
City	New	M S
City - State/Province - Postal Code	New	P M S
City - State/Province - Postal Code (Global)	New	P
City - State/Province - Postal Code (International)	New	S
Date (DMY Validation - Numeric Only)	New	I
Date (MDY Validation - Numeric Only)	New	I
Date (YMD Validation - Numeric Only)	New	I
E-mail	New	P M S
E-mail (Country Identification)	New	I
Field Name	New	M I
Government ID	New	M S
Hyphen/Dash Removal	New	S
Hyphen/Dash Space Replacement	New	S
IBAN	New	P
IBAN (Detailed)	New	P
IBAN (Electronic)	New	S
IBAN (Printed)	New	S
Lower	New	C
Multiple Space Collapse	New	S
Name	New	P M S G
Name (Global)	New	P
Name (Multiple Name)	New	P
Nikud Removal	New	S
Non-Alphanumeric Removal	New	S
Non-Number Removal	New	S
Number Removal	New	S
Organization	New	P M S
Organization (Global)	New	P
Organization (Matching)	New	P

Locale	Action	Type of Definition
Phone	New	P M S
Phone (Electronic)	New	S
Phone (Global)	New	P
Phone (Matching)	New	P
Phone (Multiple Number)	New	P
Phone (Standardization)	New	P
Phone (with Country Code)	New	S
Phone Country Code to Country Name	New	S
Postal Code	New	M S
Postal Code (with Country Code)	New	S
Proper	New	C
Proper (Address Number)	New	C
Proper (Given Name)	New	C
Proper (Name)	New	C
Punctuation Removal	New	S
Punctuation Space Replacement	New	S
Space Removal	New	S
Surrounding Quote Removal	New	S
URL	New	S
Upper	New	C
Website	New	P S
Word	New	A
Word (Script Identification)	New	A

*Hebrew, Israel*

## Hungarian, Hungary

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Hungarian, Hungary*

## Italian, Italy

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Italian, Italy*

## Japanese, Japan

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Japanese, Japan*

## Korean, South Korea

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Korean, South Korea*

## Malay, Malaysia

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Malay, Malaysia*

## Norwegian, Norway

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Norwegian, Norway*

## Polish, Poland

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Polish, Poland*

## Portuguese, Brazil

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Portuguese, Brazil*

## Portuguese, Portugal

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Portuguese, Portugal*

## Russian, Russia

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Russian, Russia*

## Slovak, Slovakia

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Slovak, Slovakia*

## Spanish, Mexico

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Spanish, Mexico*

## Spanish, Spain

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Spanish, Spain*

## Swedish, Sweden

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Swedish, Sweden*

## Thai, Thailand

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Thai, Thailand*

## Turkish, Turkey

Locale	Action	Type of Definition
Character (Script Identification)	Mod	A
Word (Script Identification)	Mod	A

*Turkish, Turkey*