



Informatica® Data Quality
10.1.1

加速器指南

Informatica Data Quality 加速器指南
10.1.1
2017 年 6 月

© 版权所有 Informatica LLC 2009, 2018

本软件和文档仅根据包含使用与披露限制的单独许可协议提供。未事先征得 Informatica LLC 同意，不得以任何形式、通过任何手段（电子、影印、录制或其他手段）复制或传播本文档的任何部分。

美国政府权利 交付给美国政府客户的程序、软件、数据库及相关文档和技术数据是指适用的联邦采购条例和政府机构特定补充条例中定义的“商业计算机软件”或“商业技术数据”。因此，使用、复制、披露、修改和改编应遵循适用的政府合同中规定的限制和许可条款、政府合同条款的适用范围以及 FAR 52.227-19 商用计算机软件许可中规定的额外权利。

Informatica、Informatica 标志和 PowerCenter 是 Informatica LLC 在美国和世界其他许多司法管辖区的商标或注册商标。欲获得 Informatica 商标的最新列表，请访问 <https://www.informatica.com/trademarks.html>。其他公司和产品名称可能是其各自所有者的商业名称或商标。

本软件和/或文档的某些部分受第三方版权制约，包括但不限于：版权所有 DataDirect Technologies。保留所有权利。版权所有 (C) Sun Microsystems。保留所有权利。版权所有 (C) RSA Security Inc. 保留所有权利。版权所有 (C) Ordinal Technology Corp. 保留所有权利。版权所有 (C) Aandacht c.v. 保留所有权利。版权所有 Genivia, Inc. 保留所有权利。版权所有 Isomorphic Software。保留所有权利。版权所有 (C) Meta Integration Technology, Inc. 保留所有权利。版权所有 (C) Intalio。保留所有权利。版权所有 (C) Oracle。保留所有权利。版权所有 (C) Adobe Systems Incorporated。保留所有权利。版权所有 (C) DataArt, Inc. 保留所有权利。版权所有 (C) ComponentSource。保留所有权利。版权所有 (C) Microsoft Corporation。保留所有权利。版权所有 (C) Rogue Wave Software, Inc. 保留所有权利。版权所有 (C) Teradata Corporation。保留所有权利。版权所有 (C) Yahoo! Inc. 保留所有权利。版权所有 (C) Glyph & Cog, LLC。保留所有权利。版权所有 (C) Thinkmap, Inc. 保留所有权利。版权所有 (C) Clearpace Software Limited。保留所有权利。版权所有 (C) Information Builders, Inc. 保留所有权利。版权所有 (C) OSS Nokalva, Inc. 保留所有权利。版权所有 Edifecs, Inc. 保留所有权利。版权所有 Cleo Communications, Inc. 保留所有权利。版权所有 (C) International Organization for Standardization 1986。保留所有权利。版权所有 (C) ej-technologies GmbH。保留所有权利。版权所有 (C) Jaspersoft Corporation。保留所有权利。版权所有 (C) International Business Machines Corporation。保留所有权利。版权所有 (C) yWorks GmbH。保留所有权利。版权所有 (C) Lucent Technologies。保留所有权利。版权所有 (C) University of Toronto。保留所有权利。版权所有 (C) Daniel Veillard。保留所有权利。版权所有 (C) Unicode, Inc. 版权所有 IBM Corp. 保留所有权利。版权所有 (C) MicroQuill Software Publishing, Inc. 保留所有权利。版权所有 (C) PassMark Software Pty Ltd. 保留所有权利。版权所有 (C) LogiXML, Inc. 保留所有权利。版权所有 (C) 2003-2010 Lorenzi Davide。保留所有权利。版权所有 (C) Red Hat, Inc. 保留所有权利。版权所有 (C) The Board of Trustees of the Leland Stanford Junior University。保留所有权利。版权所有 (C) EMC Corporation。保留所有权利。版权所有 (C) Flexera Software。保留所有权利。版权所有 (C) Jinfonet Software。保留所有权利。版权所有 (C) Apple Inc. 保留所有权利。版权所有 (C) Telerik Inc. 保留所有权利。版权所有 (C) BEA Systems。保留所有权利。版权所有 (C) PDFlib GmbH。保留所有权利。版权所有 (C) Orientation in Objects GmbH。保留所有权利。版权所有 (C) Tanuki Software, Ltd. 保留所有权利。版权所有 (C) Ricebridge。保留所有权利。版权所有 (C) Sencha, Inc. 保留所有权利。版权所有 (C) Scalable Systems, Inc. 保留所有权利。版权所有 (C) jQWidgets。保留所有权利。版权所有 (C) Tableau Software, Inc. 保留所有权利。版权所有 (C) MaxMind, Inc. 保留所有权利。版权所有 (C) TMate Software s.r.o. 保留所有权利。版权所有 (C) MapR Technologies Inc. 保留所有权利。版权所有 (C) Amazon Corporate LLC。保留所有权利。版权所有 (C) Highsoft。保留所有权利。版权所有 (C) Python Software Foundation。保留所有权利。版权所有 (C) BeOpen.com。保留所有权利。版权所有 (C) CNRI。保留所有权利。

本产品包括由 Apache Software Foundation (<http://www.apache.org/>) 开发的软件和/或在不同 Apache 许可证版本（以下简称“许可证”）下许可的其他软件。您可从 <http://www.apache.org/licenses/> 获取这些许可证的副本。除非适用法律要求或者有相应书面协议，否则依据这些“许可证”分发的软件以“原样”提供，不附带任何明示或暗示的担保或条件。请参阅“许可证”中规定的具体语言管理权限和限制。

本产品包括由 Mozilla (<http://www.mozilla.org/>) 开发的软件、由 JBoss Group, LLC 开发的软件（版权所有 JBoss Group, LLC 保留所有权利）、由 Bruno Lowagie 和 Paulo Soares 开发的软件（版权所有 (C) 1999-2006 Bruno Lowagie 和 Paulo Soares）以及在 <http://www.gnu.org/licenses/lgpl.html> 网站上的不同版本 GNU Lesser General 公共许可协议下许可的软件。这些材料由 Informatica 按“原样”免费提供，不附带任何明示或暗示的担保，包括但不限于适销性和特定用途适用性的暗示担保。

本产品包括 ACE(TM) 和 TAO(TM) 软件，这些软件版权归 Douglas C. Schmidt 及其在华盛顿大学、加利福尼亚大学欧芬分校以及范德堡大学的研发团队所有（版权所有 (C) 1993-2006，保留所有权利）。

本产品包括由 OpenSSL Project 开发并在 OpenSSL Toolkit（版权所有 OpenSSL Project。保留所有权利）中使用的软件，该软件的再分发受 <http://www.openssl.org> 和 <http://www.openssl.org/source/license.html> 上规定条款之制约。

本产品包括 Curl 软件，版权所有 1996-2013，Daniel Stenberg <daniel@haxx.se>。保留所有权利。有关该软件的权限和限制受 <http://curl.haxx.se/docs/copyright.html> 上规定条款之制约。允许出于任何目的以免费或收费形式使用、复制、修改和分发该软件，但前提是所有副本均应注明上述版权声明以及本许可声明。

本产品包括由 MetaStuff, Ltd. 开发的软件，版权所有 2001-2005 ((C)) MetaStuff, Ltd. 保留所有权利。有关该软件的权限和限制受 <http://www.dom4j.org/license.html> 上规定条款之制约。

本产品包括由 Dojo Foundation 开发的软件，版权所有 (C) 2004-2007, Dojo Foundation。保留所有权利。有关该软件的权限和限制受 <http://dojotoolkit.org/license> 上规定条款之制约。

本产品包括 ICU 软件，版权所有 International Business Machines Corporation 和其他方。保留所有权利。有关该软件的权限和限制受 <http://source.icu-project.org/repos/icu/icu/trunk/license.html> 上规定条款之制约。

本产品包括由 Per Bothner 开发的软件，版权所有 (C) 1996-2006 Per Bothner。保留所有权利。<http://www.gnu.org/software/kawa/Software-License.html> 上的许可证中规定了您使用这些材料的权利。

本产品包括 OSSP UUID 软件，版权所有 (C) 2002 Ralf S. Engelschall，版权所有 (C) 2002 OSSP Project，版权所有 (C) 2002 Cable & Wireless Deutschland。有关该软件的权限和限制受 <http://www.opensource.org/licenses/mit-license.php> 上规定条款之制约。

本产品包括由 Boost (<http://www.boost.org/>) 开发的软件或在 Boost 软件许可证下许可的软件。有关该软件的权限和限制受 http://www.boost.org/LICENSE_1_0.txt 上规定条款之制约。

本产品包括由 University of Cambridge 开发的软件，版权所有 (C) 1997-2007 University of Cambridge。有关该软件的权限和限制受 <http://www.pcre.org/license.txt> 上规定条款之制约。

本产品包括由 The Eclipse Foundation 开发的软件，版权所有 (C) 2007 The Eclipse Foundation。保留所有权利。有关该软件的权限和限制受 <http://www.eclipse.org/org/documents/epl-v10.php> 和 <http://www.eclipse.org/org/documents/edl-v10.php> 上规定条款之制约。

本产品包括在 <http://www.tcl.tk/software/tcltk/license.html>、<http://www.bosrup.com/web/overlib/?License>、<http://www.stlport.org/doc/license.html>、<http://asm.ow2.org/license.html>、<http://www.cryptix.org/LICENSE.TXT>、<http://hsqldb.org/web/hsqldbLicense.html>、<http://httpunit.sourceforge.net/doc/license.html>、<http://jung.sourceforge.net/license.txt>、http://www.gzip.org/zlib/zlib_license.html、<http://www.openldap.org/software/release/license.html>、<http://www.libssh2.org>、<http://slf4j.org/license.html>、<http://www.sente.ch/software/OpenSourceLicense.html>、<http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>、<http://antlr.org/license.html>、<http://aopalliance.sourceforge.net/>、<http://www.bouncycastle.org/licence.html>、<http://www.jgraph.com/jgraphdownload.html>、<http://www.jcraft.com/jsch/LICENSE.txt>、http://jotm.objectweb.org/bsd_license.html、<http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>、<http://www.slf4j.org/license.html>、<http://nanoxml.sourceforge.net/orig/copyright.html>、<http://www.json.org/license.html>、<http://forge.ow2.org/projects/jaservice/>、<http://www.postgresql.org/about/license.html>、<http://www.sqlite.org/copyright.html>、<http://www.tcl.tk/software/tcltk/license.html>、<http://www.jaxen.org/faq.html>、<http://www.jdom.org/docs/faq.html>、<http://www.slf4j.org/license.html>、<http://www.iodbc.org/dataspace/iodbc/wiki/IODBC/License>、<http://www.keplerproject.org/md5/license.html>、<http://www.toedter.com/en/jcalendar/license.html>、<http://www.edankert.com/bounce/index.html>、<http://>

www.net-snmp.org/about/license.html、<http://www.openmdx.org/#FAQ>、http://www.php.net/license/3_01.txt、<http://srp.stanford.edu/license.txt>、<http://www.schneier.com/blowfish.html>、<http://www.jmock.org/license.html>、<http://xsom.java.net>、<http://benalman.com/about/license/>、<https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>、<http://www.h2database.com/html/license.html#summary>、<http://jsoncpp.sourceforge.net/LICENSE>、<http://jdbc.postgresql.org/license.html>、<http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>、<https://github.com/rantav/hector/blob/master/LICENSE>、<http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>、<http://jibx.sourceforge.net/jibx-license.html>、<https://github.com/lyokato/libgeohash/blob/master/LICENSE>、<https://github.com/hjiang/jsonxx/blob/master/LICENSE>、<https://code.google.com/p/lz4/>、<https://github.com/jedisct1/libsodium/blob/master/LICENSE>、<http://one-jar.sourceforge.net/index.php?page=documents&file=license>、<https://github.com/EsotericSoftware/kryo/blob/master/license.txt>、<http://www.scala-lang.org/license.html>、<https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>、<http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>、<https://aws.amazon.com/asl/>、<https://github.com/twbs/bootstrap/blob/master/LICENSE> 和 <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt> 下许可的软件。

本产品包括在 Academic 免费许可证 (<http://www.opensource.org/licenses/afl-3.0.php>)、通用开发和分发许可证 (<http://www.opensource.org/licenses/cddl1.php>)、通用公共许可证 (<http://www.opensource.org/licenses/cpl1.0.php>)、Sun Binary Code 许可协议补充许可条款、BSD 许可证 (<http://www.opensource.org/licenses/bsd-license.php>)、新 BSD 许可证 (<http://opensource.org/licenses/BSD-3-Clause>)、MIT 许可证 (<http://www.opensource.org/licenses/mit-license.php>)、Artistic 许可证 (<http://www.opensource.org/licenses/artistic-license-1.0>) 以及原始开发者公共许可证版本 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>) 下许可的软件。

本产品包括由 Joe Walnes 和 XStream Committers 开发的软件，版权所有 (C) 2003-2006 Joe Walnes, 2006-2007 XStream Committers。保留所有权利。有关该软件的权限和限制受 <http://xstream.codehaus.org/license.html> 上规定条款之制约。本产品包括由 Indiana University Extreme! Lab 开发的软件。有关详细信息，请访问 <http://www.extreme.indiana.edu/>。

本产品包括软件版权所有 (c) 2013 Frank Balluffi 和 Markus Moeller。保留所有权利。有关此软件的权限和限制受 MIT 许可证上规定条款之制约。

请参阅位于以下位置的专利：<https://www.informatica.com/legal/patents.html>。

免责声明：Informatica LLC 以“原样”提供本文档，不附带任何明示或暗示的担保，包括但不限于非侵权、适销性或特定用途适用性的暗示担保。Informatica LLC 不保证本软件和文档中没有错误。本软件或文档中提供的信息可能包括技术上的不准确性或排字错误。本软件和文档中包含的信息随时可能更改，恕不另行通知。

声明

本 Informatica 产品（以下称“软件”）包括由 Progress Software Corporation 的运营公司 DataDirect Technologies（以下称“DataDirect”）提供的某些驱动程序（以下称“DataDirect 驱动程序”），受以下条款和条件制约：

1. DataDirect 驱动程序以“原样”提供，不附带任何明示或暗示的担保，包括但不限于适销性、特定用途适用性以及非侵权的暗示担保。
2. 在任何情况下，DataDirect 或其第三方供应商均不对最终用户客户承担因使用 ODBC 驱动程序而引起的任何直接、间接、偶发、特殊、继发或其他损害赔偿的责任，无论是否已提前告知该种损害的可能性。这些限制适用于所有诉因，包括但不限于违反合同、违反担保、过失、严格责任、虚假陈述以及其他侵权行为。

本文档中的信息如有更改，恕不另行通知。如果您发现本文档中存在任何问题，请以书面形式将问题报告给我们，邮寄地址是 Informatica LLC 2100 Seaport Blvd.Redwood City, CA 94063。

Informatica 产品根据对应协议的条款和条件进行担保。INFORMATICA 按“原样”提供本文档中的信息，无任何明示或暗示的担保，包括但不限于任何适销性和特定用途适用性担保，也没有任何非侵权担保或条件。

发布日期: 2018-06-09

目录

前言	8
Informatica 资源	8
Informatica Network	8
Informatica 知识库	8
Informatica 文档	8
Informatica 产品可用性矩阵	9
Informatica Velocity	9
Informatica Marketplace	9
Informatica 全球客户支持部门	9
 第 1 章： 加速器简介.....	10
加速器概览	10
加速器的结构	10
常规加速器结构	10
数据域加速器结构	11
加速器的安装	11
安装加速器时需要注意的规则和准则	13
导入规则和映射	13
导入数据域和数据域组	14
加速器组件	14
Rules	16
演示映射	16
数据域	17
引用表	17
内容集	17
标记和规则	17
在 PowerCenter 中使用加速器	17
 第 2 章： 核心加速器.....	19
核心加速器概览	19
核心地址数据清理规则	19
核心联系人数据清理规则	21
核心公司数据清理规则	21
核心常规数据清理规则	22
核心匹配和消除重复记录规则	27
核心产品数据清理规则	27
核心演示映射	28
 第 3 章： 数据域加速器.....	29
数据域加速器概览	29

数据域加速器中的数据域.	29
数据域加速器中的列名称规则.	36
数据域加速器中的数据规则.	39
第 4 章： 澳大利亚/新西兰加速器.	45
澳大利亚/新西兰加速器概览.	45
澳大利亚/新西兰地址数据清理规则.	45
澳大利亚/新西兰复合规则.	47
澳大利亚/新西兰联系人数据清理规则.	49
澳大利亚/新西兰公司数据清理规则.	51
澳大利亚/新西兰常规数据清理规则.	52
澳大利亚/新西兰匹配和消除重复记录规则.	53
澳大利亚/新西兰演示映射.	55
第 5 章： 巴西加速器.	56
巴西加速器概览.	56
巴西地址数据清理规则.	56
巴西复合规则.	57
巴西联系人数据清理规则.	58
巴西公司数据清理规则.	59
巴西常规数据清理规则.	60
巴西匹配和消除重复记录规则.	61
巴西演示映射.	62
第 6 章： 金融服务加速器.	63
金融服务加速器概览.	63
金融服务联系人数据清理规则.	63
金融服务财务数据清理规则.	64
金融服务常规数据清理规则.	66
金融服务匹配和消除重复记录规则.	67
第 7 章： 法国加速器.	69
法国加速器概览.	69
法国地址数据清理规则.	69
法国复合规则.	70
法国联系人数据清理规则.	71
法国公司数据清理规则.	73
法国常规数据清理规则.	73
法国匹配和消除重复记录规则.	74
法国演示映射.	75
第 8 章： 德国加速器.	77
德国加速器概览.	77

德国地址数据清理规则.	77
德国复合规则.	78
德国联系人数据清理规则.	79
德国公司数据清理规则.	81
德国常规数据清理规则.	81
德国匹配和消除重复记录规则.	82
德国演示映射.	84
第 9 章：葡萄牙加速器.	85
葡萄牙加速器概览.	85
葡萄牙地址数据清理规则.	85
葡萄牙复合规则.	86
葡萄牙联系人数据清理规则.	87
葡萄牙公司数据清理规则.	89
葡萄牙常规数据清理规则.	89
葡萄牙匹配和消除重复记录规则.	90
葡萄牙演示映射.	91
第 10 章：西班牙加速器.	93
西班牙加速器概览.	93
西班牙地址数据清理规则.	93
西班牙联系人数据清理规则.	94
西班牙公司数据清理规则.	96
西班牙常规数据清理规则.	96
西班牙匹配和消除重复记录规则.	97
西班牙演示映射.	99
第 11 章：英国加速器.	100
英国加速器概览.	100
英国地址数据清理规则.	100
英国复合规则.	102
英国联系人数据清理规则.	103
英国公司数据清理规则.	105
英国金融数据清理规则.	105
英国常规数据清理规则.	106
英国匹配和消除重复记录规则.	107
英国演示映射.	109
第 12 章：美国/加拿大加速器.	110
美国/加拿大加速器概览.	110
美国/加拿大地址数据清理规则.	110
美国/加拿大复合规则.	112
美国/加拿大联系人数据清理规则.	114

美国/加拿大公司数据清理的相关性.	119
美国/加拿大常规数据清理规则.	119
美国/加拿大匹配和消除重复记录规则.	120
美国/加拿大演示映射.	122

前言

《Informatica Data Quality 加速器指南》的目标读者是数据质量开发人员。本指南假设您已了解标准化、解析、添加标签和验证等数据质量概念。

Informatica 资源

Informatica Network

Informatica Network 囊括了 Informatica 全球客户支持部门、Informatica 知识库和其他产品资源。要访问 Informatica Network，请访问 <https://network.informatica.com>。

成员可以执行以下操作：

- 在一个位置访问您的所有 Informatica 资源。
- 在知识库中搜索文档、常见问题和最佳实践等产品资源。
- 查看产品可用性信息。
- 查看支持案例。
- 查找当地的 Informatica 用户组网络并与您的伙伴进行协作。

Informatica 知识库

使用 Informatica 知识库可在 Informatica Network 中搜索文档、入门知识文章、最佳实践和 PAM 等产品资源。

要访问知识库，请访问 <https://kb.informatica.com>。如果您对知识库有任何疑问、意见或建议，请与 Informatica 知识库团队联系，电子邮件地址为 KB_Feedback@informatica.com。

Informatica 文档

要获取有关产品的最新文档，请浏览 Informatica 知识库，网址为 https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx。

如果您对此文档有任何疑问、意见或建议，请与 Informatica 文档团队联系，电子邮件地址为 infa_documentation@informatica.com。

Informatica 产品可用性矩阵

产品可用性矩阵 (PAM) 指明了产品版本支持的操作系统版本、数据库以及其他类型的数据源和目标。如果您是 Informatica Network 成员，您可以访问 PAM，网址为 <https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Informatica Velocity

Informatica Velocity 收集了 Informatica 专业服务开发的一系列提示和最佳实践。Informatica Velocity 基于数以百计的数据管理项目的实际经验而开发，汇集了我们曾在世界各地组织就职的顾问在成功规划、开发、部署和维护数据管理解决方案方面的知识。

如果您是 Informatica Network 成员，您可以访问 Informatica Velocity 资源，网址为 <http://velocity.informatica.com>。

如果您对 Informatica Velocity 有任何疑问、意见或建议，请通过 ips@informatica.com 与 Informatica 专业服务联系。

Informatica Marketplace

Informatica Marketplace 是一个论坛，该论坛中提供的解决方案可补充、扩展或增强您的 Informatica 实现。您可以利用 Informatica 开发人员和合作伙伴提供的数以百计解决方案中的任何方案，提高生产率，加快项目的实现时间。您可以访问 Informatica Marketplace，网址为 <https://marketplace.informatica.com>。

Informatica 全球客户支持部门

您可以通过电话或 Informatica Network 上的联机支持与全球支持中心联系。

要查找您当地的 Informatica 全球客户支持部门电话号码，请访问 Informatica 网站，链接为：
<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>。

如果您是 Informatica Network 成员，您可以使用联机支持，网址为 <http://network.informatica.com>。

第 1 章

加速器简介

本章包括以下主题：

- [加速器概览, 10](#)
- [加速器的结构, 10](#)
- [加速器的安装, 11](#)
- [加速器组件, 14](#)
- [标记和规则, 17](#)
- [在 PowerCenter 中使用加速器, 17](#)

加速器概览

加速器是用于解决某个国家/地区、区域或行业的常见数据质量问题的内容包。加速器可能会包含可用于分析和改善组织数据的 Maplet。加速器还可能包含可用于发现数据所含信息类型的数据域。

请将 Maplet 和数据域添加到模型存储库。Informatica 会配置 Maplet 和数据域，以便与您可能为组织数据定义的业务规则相对应。加速器使用术语 *Maplet* 和 *规则* 来标识 Maplet。在将 Maplet 导入模型存储库时，Developer tool 会在名为 *Rules* 的文件夹中创建 Maplet 对象。

Informatica Data Quality 包含一个核心加速器和一个核心数据域加速器。您可以从 Informatica 购买和下载另外的加速器。

加速器的结构

加速器是一个压缩文件，其目录结构中包含存储库元数据文件和其他文件。目录结构取决于加速器类型。常规的加速器包含规则、引用数据对象、演示映射和演示数据源。数据域加速器包含规则、引用数据对象、数据域和数据域组。

常规加速器结构

常规加速器包括分析和改善组织数据的规则，以及用于演示规则操作的示例映射。常规加速器还包含上述规则和映射使用的引用数据文件和源数据文件。

常规加速器包含以下目录：

- Accelerator_Content

- Accelerator_Sources

Accelerator_Content 目录

Accelerator_Content 目录包含以下组件：

Accelerator XML 文件

包含规则、演示映射、引用表和数据对象的元数据。

引用数据文件

包含规则和映射用于识别不同格式数据值的引用数据。引用数据文件是一个在多个目录下包含字典文件的压缩文件。在导入对应的 XML 文件时，指定该压缩文件。导入过程会将引用数据复制到引用数据数据库中的表内。

注意: 如果您将一个包含规则的映射导出到 PowerCenter(R)，请将字典文件复制到 PowerCenter 集成服务可以读取的目录。

Accelerator_Sources 目录

Accelerator_Sources 目录包含演示数据文件。演示数据文件是一个包含演示映射的源数据的压缩文件。将源数据文件复制到文件系统。

数据域加速器结构

数据域加速器包含决定了数据集内的信息类型的数据域，以及定义了数据域逻辑的规则。加速器还包含供数据域和规则使用的引用数据文件。

数据域加速器包含以下文件：

数据域元数据文件

此文件包含添加到数据域词汇表的数据域和数据域组的元数据。

规则元数据文件

此文件包含定义了数据域逻辑的规则的元数据，以及数据域使用的引用数据对象的元数据。

用于数据域的引用数据文件

此文件包含当您运行含有某个数据域的配置文件时该数据域所使用的引用数据。引用数据文件是一个在多个目录下包含字典文件的压缩文件。在导入对应的 XML 文件时，指定该压缩文件。导入过程会将引用数据复制到引用数据数据库中的表内。

用于数据域规则的引用数据文件

此文件包含当您运行含有某个规则的数据域时该规则所使用的引用数据。引用数据文件是一个在多个目录下包含字典文件的压缩文件。在导入对应的 XML 文件时，指定该压缩文件。导入过程会将引用数据复制到引用数据数据库中的表内。

加速器的安装

要安装加速器，请将存储库对象元数据导入模型存储库项目，并将演示数据文件复制到文件系统。使用 Developer tool 导入存储库对象。

在导入规则和演示映射时，请从对象浏览器中选择存储库项目。在导入数据域时，请从**首选项**对话框选择存储库项目。在上述两种情况下，导入操作都会提示您选择包含 XML 文件所指定的引用数据的压缩文件。

常规加速器示例

您可以为核心加速器导入以下元数据文件：

Informatica_Core_Accelerator_1011.xml

在导入元数据文件时，请选择以下引用数据文件：

Informatica_Core_Accelerator_1011.zip

数据域加速器示例

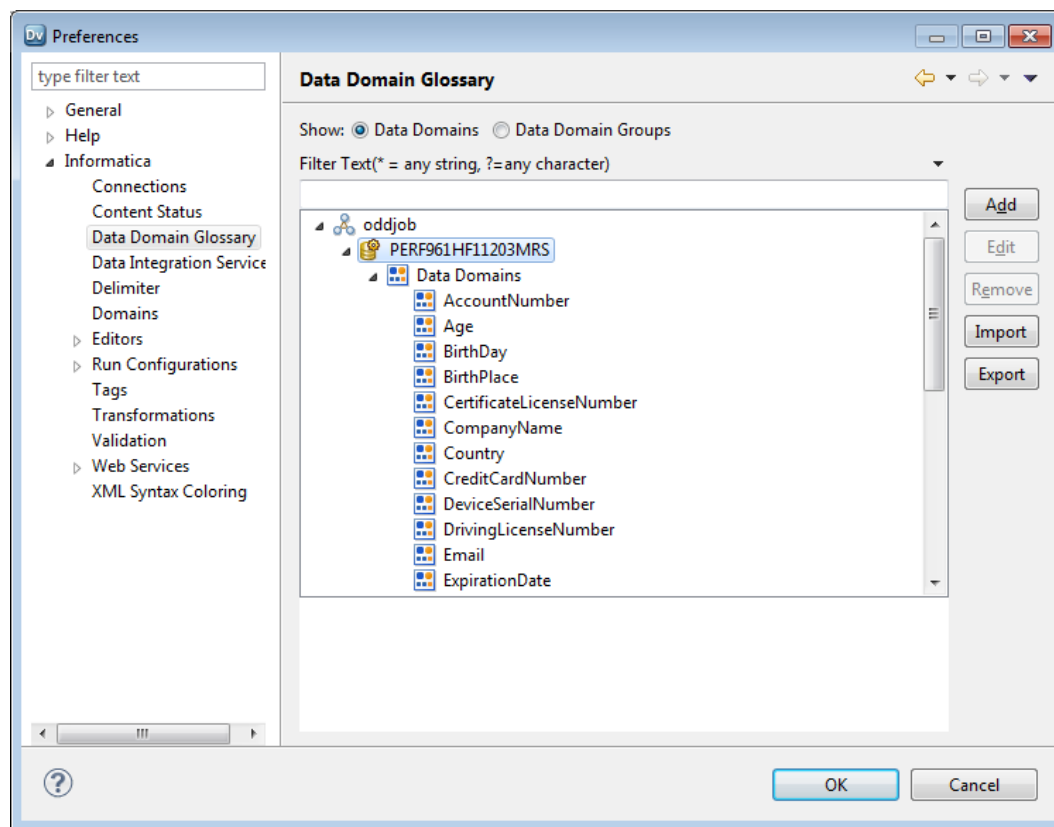
您可以为核心数据域加速器导入以下元数据文件：

Informatica_IDE_DataDomain_1011.xml

在导入元数据文件时，请选择以下引用数据文件：

Informatica_IDE_DataDomain_1011.zip

下图显示了**首选项**对话框中的数据域：



示例映射的源数据

在导入常规加速器时，将演示数据文件复制到数据集成服务计算机上的以下目录：

<Informatica 安装目录>\services\DQContent\INFA_Content\demos\source_data

安装加速器时需要注意的规则和准则

加速器中的存储库对象和数据文件与 Informatica 系统中的其他对象和文件操作方式相同。加速器内容需要遵循某些规则和准则。

安装加速器时，请注意以下规则和准则：

- 在导入或复制文件前，确认您对数据集成服务、内容管理服务和分析服务拥有全部特权。
- 将加速器导入一个模型存储库项目。您需要在导入加速器前创建该项目。
- 先安装核心加速器，然后再安装其他加速器。
- 首先安装核心数据域加速器，然后再安装数据域加速器。
- 如果导入的元数据文件包含一个与之前导入的加速器相同的对象，请在存储库中替换该对象。
- 要使用加速器规则执行地址验证，请下载并安装加速器所指定的国家/地区的地址引用数据文件。要使用加速器规则执行标识匹配分析，请下载并安装加速器所指定的国家/地区的标识填充文件。您从 Informatica 购买地址引用数据文件和标识填充文件。

导入规则和映射

使用 Developer tool 可导入规则、演示映射和映射数据源的元数据。在导入操作期间，选择规则和映射使用的引用数据文件。

1. 在 Developer tool 中，连接到包含元数据目标项目的模型存储库。
2. 在对象浏览器中，选择目标项目。
例如，选择 *Informatica_DQ_Content* 项目。如果需要，在模型存储库中创建一个项目。
3. 选择 **文件 > 导入**。
4. 在 **导入** 对话框中，选择 **Informatica > 导入对象元数据文件(高级)**。
5. 单击 **下一步**。
6. 在加速器目录结构中浏览到 XML 元数据文件，然后选择该文件。
7. 单击 **打开**，然后单击 **下一步**。
8. 在 **源** 窗格中，选择项目节点下显示的项。
9. 在 **目标** 窗格中，选择目标项目。
10. 单击 **添加到目标**。
 - 如果存储库项目中包含某个您要添加的对象，Developer tool 会提示您将该对象与当前对象合并。单击 **是合并对象**。
 - 如果 Developer tool 提示您重命名对象，单击 **否**。
 - 如果有任何对象留在 **源** 窗格中，使用指针将对象移至目标项目。
11. 单击 **下一步**。
12. 在加速器目录结构中浏览到压缩的引用数据文件，然后选择该文件。
13. 单击 **打开**。
14. 验证代码页是否为 UTF-8，然后单击 **下一步**。
15. 在 **目标连接** 字段中，选择引用数据数据库。
16. 单击 **完成**。

导入数据域和数据域组

使用“首选项”对话框导入数据域和数据域组的元数据。在导入操作过程中，请选择数据域使用的引用数据文件。

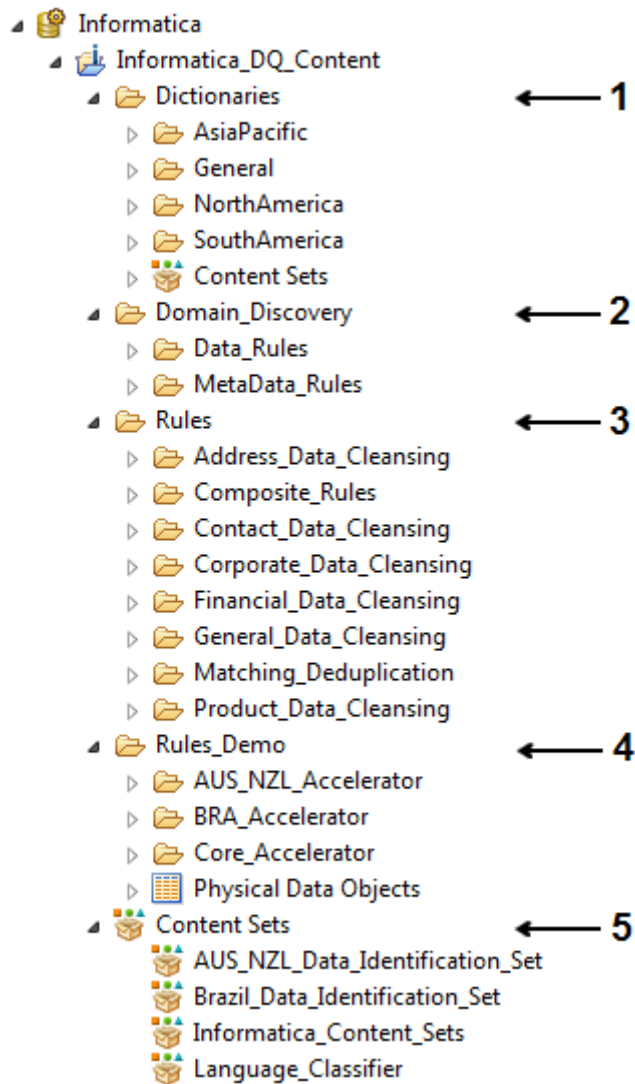
1. 在 Developer tool 中，连接到包含元数据目标项目的模型存储库。
2. 选择窗口 > 首选项。
3. 在首选项对话框中，展开 Informatica 节点并选择**数据域词汇表**。
4. 在存储库窗格中，选择数据域或数据域组的顶级节点。
5. 单击**导入**。
6. 浏览到加速器目录结构中的 XML 元数据文件，并选择该文件。
7. 单击**打开**，然后单击**下一步**。
8. 在**源**窗格中，选择数据域词汇表项目。
9. 在**目标**窗格中，选择目标项目。
10. 选择“解决办法”字段中的以下选项：
目标中的替换选项
11. 单击**向目标添加内容**。
 - 如果 Developer tool 提示您要添加对象，请单击**是**。
 - 如果 Developer tool 提示您要重命名对象，请单击**否**。
12. 单击**下一步**。
13. 如果导入操作识别出相关性，请将相关对象从源项目复制到目标项目。
14. 单击**下一步**。
15. 浏览到加速器目录结构中的压缩引用数据文件，并选择该文件。
16. 单击**打开**。
17. 验证代码页是否为 UTF-8，然后单击**下一步**。
18. 在**目标连接**字段中，选择引用数据数据库。
19. 单击**完成**。

加速器组件

当您导入加速器时，Developer tool 会为该加速器指定的规则、数据域和其他对象创建文件夹。每个文件夹均包含子文件夹，按国家/地区和对对象执行的数据质量操作的类型组织对象。

使用核心加速器在存储库项目中创建文件夹。当您导入其他加速器时，请将对象和文件夹添加到该项目。

下图显示了您向项目导入多个加速器时的 Informatica_DQ_Content 项目文件夹结构：



1. Dictionaries 文件夹
2. Domain_Discovery 文件夹
3. Rules 文件夹
4. Rules_Demo 文件夹
5. Content Sets 文件夹

项目包含以下顶级文件夹：

Dictionaries

Dictionaries 文件夹包含引用表对象。每个对象引用一个位于引用数据数据库中的表。

Domain_Discovery

Domain_Discovery 文件夹包含在您安装的加速器中定义数据域的规则。该文件夹包含 Data_Rules 文件夹和 Metadata_Rules 文件夹。Data_Rules 文件夹中的规则与用来分析列数据值的数据域相对应。Metadata_Rules 文件夹中的规则与用来分析列名称的数据域相对应。

规则

Rules 文件夹包含用于分析和改善数据的规则。

Rules_Demo

Rules_Demo 文件夹包含演示映射和演示数据源。

内容集

Content Sets 文件夹包含不在引用数据数据库中指定数据的引用数据对象。

Rules

加速器规则定义了一系列数据分析和数据转换操作。您可以将单个规则或一系列规则添加到映射。

使用加速器规则可执行以下数据质量任务：

地址验证

验证并改善通信地址记录中的数据。规则要求使用地址引用数据文件。

数据解析

解析记录中的信息。解析规则可以提取多种类型的信息，包括人名、组织名称、电话号码、日期和身份号码。

数据标准化

将数据值的拼写和格式标准化。标准化规则可以识别并更正多种类型的信息，包括人名、组织名称、电话号码、日期和身份号码。

重复分析

查找数据集中的重复记录。重复分析规则会比较数据集中的记录，并生成表示记录之间相似性的数字得分。

重复分析规则可以读取包含常规公司数据的记录和包含标识数据的记录。标识数据规则要求使用标识填充数据文件。

导入操作会将规则添加到以下存储库文件夹：

[Informatica_DQ_Content]\Rules

执行地址验证、数据解析和数据标准化操作的规则位于加速器项目中的 *Data Cleansing* 子文件夹中。执行重复分析的规则位于加速器项目中的 *Matching Deduplication* 子文件夹中。

如果您导入国家或地区的规则，将为复合规则添加一个子文件夹。复合规则以嵌套格式将多个规则组合到一个规则中。

演示映射

演示映射是运行时对象，可将一个或多个规则应用到数据源，并将结果写入另一数据源。您可以使用演示映射作为其他映射的模板。

导入操作会将映射和数据源对象添加到以下存储库文件夹：

[Informatica_DQ_Content]\Rules_Demo

当您导入加速器时，导入操作会将演示映射的数据源添加到 Rules_Demo 文件夹。请将数据源文件从 Accelerator_Sources 目录复制到文件系统。

数据域

数据域描述可代表一列中单一类型业务信息的数据值。使用数据域可以识别列中信息的类型，以及在列中查找指定类型的信息。加速器包括多种信息类型的数据域，其中包括社会保障号、信用卡号、电子邮件地址和职称。

例如，某个数据库表可能在任何用户均可读取的“备注”列中包含社会保障号。您必须识别含有社会保障号的记录，并删除或移动社会保障号。这种情况下，可将 *SSN* 数据域添加到一个配置文件，然后对“备注”列运行该配置文件。

您可以将一个数据域分配给一个或多个数据域组。根据数据域执行的业务分析的类型，使用数据域组来组织数据域。数据域词汇表列出了您向模型存储库添加的数据域和数据域组。请使用 Developer tool 中的**首选项**菜单将数据域添加到数据域词汇表。要更新数据域中的数据定义，请使用数据域加速器中的规则。

注意：您不能在对象浏览器中查看数据域对象。

引用表

引用表包含标准版本和替代版本的一组数据值。规则使用引用表验证数据值是否准确以及是否格式正确。

导入操作将引用表添加到以下存储库文件夹：

```
[Informatica_DQ_Content]\Dictionaries
```

内容集

内容集是不在数据库表中存储数据的引用数据对象。内容集包括字符集、模式集、正则表达式、标志集、概率模型和分类器模型。

导入操作会将规则添加到以下存储库文件夹：

```
[Informatica_DQ_Content]\Content Sets
```

注意：要查看内容集中元素的列表，请在 Developer 工具中打开内容集，然后选择**标记**选项卡。

标记和规则

加速器规则包含标记，这些标记指示了规则可以读取的数据类型和规则可以执行的操作类型。

要查看应用于规则的标记，请在 Developer 工具中打开规则并单击“标记”选项卡。您可以使用 Developer 工具中的“搜索”选项查找包含指定标记的加速器。

在 PowerCenter 中使用加速器

您可以将规则和映射从模型存储库导出到文件系统和 PowerCenter 存储库。导出对象时，请选择引用表、数据对象以及依赖于所导出对象的其他相关项。

导出操作会将引用表数据复制到文件系统。将文件复制到 PowerCenter 集成服务主机。引用数据文件在 PowerCenter 目录结构中的位置必须与引用表在模型存储库文件夹结构中的位置相对应。

以下路径描述了引用数据对象在 PowerCenter 安装中的示例目录结构：

```
<Informatica_installation_directory>\services\<Model_repository_project_name>\<Model_repository_project_folder_name>
```

注意: 如果 PowerCenter 产品版本与 Developer tool 版本不一致, 请验证 PowerCenter 环境是否包含 Data Quality 集成插件。

有关 Data Quality 与 PowerCenter 集成的详细信息, 请参阅《适用于 PowerCenter 的 Informatica Data Quality 集成用户指南》

第 2 章

核心加速器

本章包括以下主题：

- [核心加速器概览, 19](#)
- [核心地址数据清理规则, 19](#)
- [核心联系人数据清理规则, 21](#)
- [核心公司数据清理规则, 21](#)
- [核心常规数据清理规则, 22](#)
- [核心匹配和消除重复记录规则, 27](#)
- [核心产品数据清理规则, 27](#)
- [核心演示映射, 28](#)

核心加速器概览

使用核心加速器中的规则可验证和改善任何国家或地区的业务数据。

核心加速器包括执行以下数据质量过程的规则：

- 地址数据清理
- 联系人数据清理
- 公司数据清理
- 常规数据清理
- 匹配和消除重复记录数据清理
- 产品数据清理

核心加速器包含其他加速器可以重用的 Maplet 和引用数据对象。请先安装核心加速器，然后再安装其他任何加速器。

核心地址数据清理规则

使用地址数据清理规则可解析、标准化和验证地址数据。

您可以在以下存储库位置找到地址数据清理规则：

[Informatica_DQ_Content]\Rules\Address_Data_Cleansing

下表介绍了核心加速器中的地址数据清理规则：

名称	说明
mplt_Global_AddressValidation5_v2_Discrete_Webservice	验证多个国家/地区的通信地址。当您可以将输入地址字段连接到地址验证器转换的离散输入端口时，请使用此 Mapplet。 此 Mapplet 调用地址验证 Web 服务。在设置其他 Web 服务 Mapplet 时，使用该 Mapplet 作为示例。
mplt_Global_AddressValidation5_v2_Hybrid_Webservice	验证多个国家/地区的通信地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用此 Mapplet。 此 Mapplet 调用地址验证 Web 服务。在设置其他 Web 服务 Mapplet 时，使用该 Mapplet 作为示例。
mplt_Global_AddressValidation5_v2_Multiline_Webservice	验证多个国家/地区的通信地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用此 Mapplet。 此 Mapplet 调用地址验证 Web 服务。在设置其他 Web 服务 Mapplet 时，使用该 Mapplet 作为示例。
rule_Calc_Distance_Between_Geocoordinates	计算两组地理坐标之间的距离。
rule_Country_Identification	识别国家/地区。
rule_Country_Name_Standardization	将国家/地区名称标准化。该规则返回国家/地区名称、双字符 ISO 国家/地区代码和三字符 ISO 国家/地区代码。
rule_Geocoordinate_In_Polygon	验证三个或更多地理坐标点所定义的区域内的地理坐标点存在情况。
rule_Global_Address_Parse_Hybrid	将非结构化的地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_Global_Address_Parse_Multiline	将非结构化的地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_Global_Address_Validation_Discrete	验证多个国家/地区的地址记录的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_Global_Address_Validation_Discrete_w_Geocoding	验证多个国家/地区的地址记录的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_Global_Address_Validation_Hybrid	验证多个国家/地区的地址记录的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_Global_Address_Validation_Hybrid_w_Geocoding	验证多个国家/地区的地址记录的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。

名称	说明
rule_Global_Address_Validation_Multiline	验证多个国家/地区的地址记录的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_Global_Address_Validation_Multiline_w_Geocoding	验证多个国家/地区的地址记录的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。

核心联系人数据清理规则

使用联系人数据清理规则解析并验证有关业务联系人和个人的数据。

您可以在以下存储库位置找到联系人地址数据清理规则：

[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

下表介绍了核心加速器中的联系人数据清理规则：

名称	说明
rule_Email_Parse	解析数据字段中的电子邮件地址。
rule_Email_Parse_and_Validate	解析数据字段中的电子邮件地址，并验证每个电子邮件地址的格式。
rule_Email_Parse_Into_Mailbox_Domain	将电子邮件地址解析为邮箱、域和子域端口。例如，此规则按以下方式解析 <i>info@informatica.com</i> ： - 邮箱：info - 子域：informatica - 域：com
rule_Email_Validation	验证电子邮件地址的格式。此规则不验证电子邮件地址是否准确或处于活动状态。此规则将返回 Valid 或 Invalid。
rule_Identify_Suspect_Names	识别可能不是真正人名的姓名。此规则会将输入值与包含可能不真实的姓名的引用表进行比较。例如，引用表包含带有虚构字符的姓名。

核心公司数据清理规则

使用核心加速器中的公司数据清理规则可将公司数据标准化。

您可以在以下存储库位置找到公司数据清理规则：

[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

下表介绍了核心加速器中的公司数据清理规则：

名称	说明
rule_Company_Name_Standardization	使用引用表将公司名称标准化。

核心常规数据清理规则

使用常规数据清理规则可解析、标准化和验证数据。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]\Rules\General_Data_Cleansing

下表介绍了核心加速器中的常规数据清理规则：

名称	说明
mplt_Parse_Tokens_Into_Single_Field	将空格分隔的字符串中的每个单词解析为单独的端口。
rule_Add_Leading_Zero	在字符串开头添加数字“0”。
rule_Add_Parentheses_At_Start_End_of_Line	在字符串开头和结尾添加括号。
rule_Add_Plus_To_Start_of_Line	在字符串开头添加加号。
rule_Add_Space_Around_Ampersand	在字符串中的所有与号前后添加空格。
rule_Add_Space_Around_Hyphen	在字符串中所有短划线和连字符前后添加空格。
rule_Add_Space_Between_Number_Letter	在一个数字字符与一个字母字符组成的字符对之间添加空格。按照从左到右的读取顺序，该 Maplet 向数据中的第一个数字字母字符对添加空格。
rule_Add_Spaces_Around_Period	在字符串中的所有句号前后添加空格。
rule_AllTrim	删除输入数据字段中的所有前导空格和尾随空格。
rule_Assign_DQ_AddressResolutionCode_Description	为地址验证器转换的“地址解析代码”输出分配说明。
rule_Assign_DQ_ElementInputStatus_Description	为地址验证器转换的“元素输入状态”输出分配说明。该说明对应于 Data Quality 9.0 之前版本中的数据质量转换的输出。
rule_Assign_DQ_ElementRelevance_Description	为地址验证器转换的“元素相关性”输出分配说明。该说明对应于 Data Quality 9.0 之前版本中的数据质量转换的输出。
rule_Assign_DQ_ElementResultStatus_Description	为地址验证器转换的“元素结果状态”输出分配说明。该说明对应于 Data Quality 9.0 之前版本中的数据质量转换的输出。
rule_Assign_DQ_ExtendedElementStatus_Description	为地址验证器转换的“扩展元素结果状态”输出分配说明。

名称	说明
rule_Assign_DQ_GeocodingStatus_Description	为地址验证器转换的“地理编码状态”输出分配说明。该说明对应于 Data Quality 9.0 之前版本中的数据质量转换的输出。
rule_Assign_DQ_Mailability_Score_Description	为地址验证器转换的“可邮寄得分”输出分配说明。该说明对应于 Data Quality 9.0 之前版本中的数据质量转换的输出。
rule_Assign_DQ_Match_Code_Description	为地址验证器转换的“匹配代码”输出分配说明。该说明对应于 Data Quality 9.0 之前版本中的数据质量转换的输出。
rule_Classify_Language	<p>将字符串归类为以下一种语言：阿拉伯语、荷兰语、英语、法语、德语、意大利语、葡萄牙语、俄语、西班牙语或土耳其语。该规则使用 Language_Classifier 内容集来识别语言。</p> <p>注意: 对于所分析的每个字符串，该规则将返回一种语言。如果字符串属于该规则无法识别的某种语言，规则将返回与字符串文本最接近的语言。</p>
rule_Compare_Dates	<p>计算两个日期之间的差值。该 Mapplet 使用以下度量单位：</p> <ul style="list-style-type: none"> - 小时 - 日 - 月 - 年 <p>每个输出值均与其他值互斥。不能将输出相加来表示数据值之间的差值。</p>
rule_Completeness	检查单个端口是否包含空值 (NULL)。如果端口包含数据，将返回“完整”。如果端口为空或包含空值 (NULL)，则返回“不完整”。
rule_Completeness_Multi_Port	检查多个端口是否包含空值 (NULL)。如果所有端口都包含数据，将返回“完整”。如果有任何端口为空或包含空值 (NULL)，则返回“不完整”。
rule_Concatenate_Words	连接两个字段。使用字符空格作为分隔符。
rule_Convert_Match_Codes_to_Legacy_Values	将地址验证器转换中“匹配代码”端口的输出转换为 Data Quality 8.6 中与之相当的地址验证匹配代码。
rule_CreditCard_Number_Validation	<p>验证使用 Luhn 算法的信用卡的信用卡号。该验证包括但不限于以下信用卡：</p> <ul style="list-style-type: none"> - American Express - Diners Club Carte Blanche - Diners Club International - Diners Club US & Canada - Discover Card - JCB - Maestro - Master Card - Solo - Switch - Visa - Visa Electron <p>该规则将返回“有效”或“无效”。</p>
rule_Date_Complete	<p>验证输入字符串是否符合规则识别的日期格式。该规则可读取以下引用数据对象：</p> <ul style="list-style-type: none"> - user_defined_dates_infra

名称	说明
rule_Date_of_Birth_Validation	检查出生日期与当前日期之间的年数。如果年数为 120 或更低，除了返回“有效”之外，还将返回“成年人”或“未成年人”。如果年数大于 120，则返回“无效”。
rule_Date_Parse	将字符串中的日期数据解析为此规则指定的端口。该规则可识别以下格式的日期： <ul style="list-style-type: none"> - dd/mm/yyyy - mm/dd/yyyy - yyyy/dd/mm 该规则将返回一个日期，以及一个包含无日期的输入文本的字符串。
rule_Date_Standardization	将日期字符串标准化为您指定的输出格式。要设置输出格式，请打开规则中的 dq_FormatDate 表达式转换，并更新 Output_Date_Format 表达式变量和 Delimiter 表达式变量。如果输入数据未描述有效日期，该规则将为每个输入字符返回数字 0。
rule_Date_Validation	验证数据列中以单一格式出现的日期字符串。要配置该规则执行验证所使用的日期格式，请打开规则中的 dq_ValidateDate 表达式转换并更新 In_Date_Format 表达式变量。默认格式为“MM/DD/YYYY”。该规则将返回“有效”或“无效”。
rule_Date_Validation_Variable_Format	验证数据列中以多种格式出现的日期字符串。当数据源包括以下列时，请使用该规则： <ul style="list-style-type: none"> - 包含多种格式的日期值的列。 - 标识了每行的日期值格式的列。如果列未标识行的日期格式，该规则将为日期值应用“MM/DD/YYYY”格式。 该规则将读取 <i>is_date()</i> 功能识别的所有数据值。该规则将返回“有效”或“无效”。
rule_Days_Between_Dates	计算两个日期之间的天数。
rule_Days_From_Current_Date	计算指定日期与当前日期之间的天数。
rule_EAN13_Algorithm	验证国际商品编码 (International Article Number)。如果编码的校验位正确，该规则将返回“有效”；如果校验位不正确，则返回“无效”。
rule_GTIN_Validation	验证全球贸易项目代码 (Global Trade Item Number, GTIN)。该规则可验证 8 位、12 位、13 位和 14 位代码。如果编码的校验位正确，该规则将返回“有效”；如果校验位不正确，则返回“无效”。
rule_IsNumeric	验证输入数据是否为数值。该规则将返回“True”或“False”。
rule_LowerCase	返回所有小写字母字符。
rule_Luhn_Algorithm	将 Luhn 算法应用于数字字符串。该规则可以验证数字字符串，如信用卡号。
rule_Mask_Profanity	检查输入数据是否包含猥亵语言。在输出数据中将猥亵语言屏蔽为“CENSORED”。
rule_Negative_Number_Validation	验证输入数据是否为负数。

名称	说明
rule_Numeric_Completeness	检查数字输入中是否包含空值 (NULL)。
rule_Parse_Alpha_Chars_from_Non_Alpha_Chars	识别输入字符串中的字母字符和非字母字符，并将每个字符集写入不同输出端口。例如，该规则会解析输入字符串 teststring_123 中的以下值： teststring _123
rule_Parse_First_Word	将输入字符串中的第一个单词解析为此规则指定的端口。
rule_Parse_Number_At_End_Of_Line	将出现在输入字符串结尾的任何数字解析为此规则指定的端口。此规则按照从左到右的顺序读取字符串。
rule_Parse_Number_At_Start_Of_Line	将出现在输入字符串开头的任何数字解析为此规则指定的端口。此规则按照从左到右的顺序读取字符串。
rule_Parse_Profanity	将字符串与包含误用的术语的引用表进行比较，并将任何与引用表值匹配的术语解析为此规则指定的端口。
rule_Parse_Text_Between_Parentheses	将包含于圆括号内的字符串解析为此规则指定的端口。此规则包含一个用于解析的字符串的输出端口，以及一个用于不含解析的字符串的输入文本的输出端口。
rule_Parse_Text_in_Single_Quotes	将包含于引号内的字符串解析为此规则指定的端口。当输入数据包含多个带引号的元素时，该规则将解析最后一个元素。此规则按照从左到右的顺序读取输入字符串。此规则包含一个用于解析的字符串的输出端口，以及一个用于不含解析的字符串的输入文本的输出端口。
rule_Past_Date_Label	判断输入日期是早于系统日期还是晚于系统日期。
rule_Personal_Company_Identification	将人名和公司名称解析为此规则指定的不同端口。该规则具有以下输出： - 人名 - 公司名称 - 数据类别（如人名或公司名称） - 此规则无法解析的数据
rule_Positive_Number_Validation	验证输入数据是否为正数。
rule_Prepending_Zero_to_Single_Digit	在单个数字字符前面添加数字“0”。
rule_Remove_All_Leading_Zeros	删除字符串开头的所有数字字符“0”。
rule_Remove_Apostrophe	删除撇号。该规则将合并撇号两侧的文本字符串。
rule_Remove_Control_Characters	删除文本字符串中的控制字符。该规则将返回一个包含控制字符的字符串，以及一个包含无控制字符的输入文本的字符串。
rule_Remove_Extra_Spaces	将所有连续空格替换为一个空格，并删除前导空格和尾随空格。
rule_Remove_Hyphen	删除连字符。
rule_Remove_Leading_Zero	删除字符串开头的数字字符“0”的单个实例。

名称	说明
rule_Remove_Limited_Punctuation	删除无关的字符。无关字符包括正斜杠、反斜杠、句号、感叹号、下划线和多个连续空格。
rule_Remove_Non_Numbers	删除所有非数字字符。
rule_Remove_Parentheses	删除左右括号。
rule_Remove_Period	删除句号。
rule_Remove_Period_Parentheses	删除以下字符： - 左右括号 - 句号
rule_Remove_Punctuation	删除标点符号。
rule_Remove_Punctuation_and_Space	删除所有标点和所有空格字符。
rule_Remove_Quotation	删除引号。
rule_Remove_Slashes	删除正斜杠和反斜杠。
rule_Remove_Space	删除所有字符空格。
rule_Replace_Ampersand_With_Space	将与号替换为空格。
rule_Replace_Hyphen_Underscore_with_Space	将连字符和下划线替换为空格。
rule_Replace_Hyphen_with_Space	将连字符替换为空格。
rule_Replace_Limited_Punct_with_Space	将以下标点字符替换为单个空格：短划线、反斜杠、句号、感叹号和下划线。该规则还会将两个、三个或四个连续空格替换为单个空格。
rule_Replace_Non_Alphabetic_with_Space	将数字和标点字符替换为单个空格。
rule_Replace_Period_With_Space	将句号替换为单个空格。
rule_Replace_Punctuation_with_Space	将所有标点替换为空格。
rule_Replace_Slashes_With_Space	将正斜杠和反斜杠替换为空格。
rule_Reverse_String_Input	颠倒输入字符串中的字符顺序。
rule_String_Completeness	检查字符串是否完整。该规则还会在输入字符串中搜索引用表 string_default_values_infa 中的值。该引用表包含 NA、DEFAULT 和 XX 等值。如果某个输入字符串包含该引用表中的值，该规则会将该字符串标识为不完整。
rule_TitleCase	将字符串转换为词首字母大写。在词首字母大写字符串中，每个单词的首字母均为大写。
rule_Translate_Diacritic_Characters	将音调字符替换为与之相当的 ASCII 字符。例如，该规则会将“ä”转换为“a”。

名称	说明
rule_UpperCase	返回所有大写字母字符。
rule_URL_Validation	验证 URL 的格式和结构。
rule_Years_Since_Date_of_Birth	计算自输入日期以来的年数。

核心匹配和消除重复记录规则

使用匹配和消除重复记录规则可识别重复记录。

您可以在以下存储库位置找到匹配和消除重复记录规则：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了核心加速器中的匹配和消除重复记录规则：

名称	说明
mplt_Consolidate_and_Remove_Duplicate_Rows	将重复记录群集整合到单个记录中，然后删除多余的重复记录。

核心产品数据清理规则

使用产品数据清理规则可解析、标准化和验证产品数据。

您可以在以下存储库位置找到产品数据清理规则：

[Informatica_DQ_Content]\Rules\Product_Data_Cleansing

下表介绍了核心加速器中的产品数据清理规则：

名称	说明
rule_Color_Parse	将颜色值解析为此规则指定的端口。
rule_Parse_Quantity_And_UOM	将字符串中某个数量和度量单位的第一个实例解析为此规则指定的端口。此规则按照从左到右的顺序读取字符串，并返回以下数据： <ul style="list-style-type: none"> - 数量。 - 度量单位。 - 无数量和度量单位值的输入字符串。

名称	说明
rule_UOM_Standardization	将度量单位标准化。该规则可返回数量和度量单位的标准化和非标准化值。还会返回一个包含具有标准化度量单位的输入文本的字符串。
rule_UPC_Validation	验证通用产品代码 (Universal Product Code) 并返回一个标准化的通用产品代码。

核心演示映射

核心加速器中的演示映射使用多个规则演示数据质量过程。

您可以在以下存储库位置找到演示映射：

[Informatica_DQ_Content]\Rules_Demo\Core_Accelerator

该加速器包含以下演示映射：

m_customer_data_demo

解析、标准化和验证美国和加拿大数据。

m_product_demo

解析产品说明并验证该说明的质量。

第 3 章

数据域加速器

本章包括以下主题：

- [数据域加速器概览, 29](#)
- [数据域加速器中的数据域, 29](#)
- [数据域加速器中的列名称规则, 36](#)
- [数据域加速器中的数据规则, 39](#)

数据域加速器概览

数据域是一个预定义或用户定义的模型存储库对象，它使用规则来发现列数据或列名称的功能含义。数据域规则定义了与源数据和元数据匹配的数据模式和列名称模式。您可以使用数据域规则来更新数据域逻辑。

使用数据域加速器中的数据域可根据列名称或列数据值发现源数据的功能含义。

数据域加速器包括以下类型的规则：

- 数据规则。查找所含数据与规则所定义的逻辑相匹配的列。
- 列名称规则。查找列名称与规则所定义的列名称逻辑相匹配的列。

数据域规则会返回布尔值，指示列数据或列名称是否符合规则标准。数据域规则使用正则表达式或引用表来查找特定值或模式。例如，可以使用九位数字规则表达式来查找采用社会保障号格式的数据值。

在数据域规则中使用表达式时，某些不相关的数据值也可能符合规则表达式的标准。例如，源数据中的美国邮政编码可能符合社会保障号的格式。要使数据域推理生效，请检查数据域发现结果中是否存在差异。在对数据域发现结果进行检查和验证后，您可以决定将数据域与数据列进行关联。

数据域加速器中的数据域

使用配置文件中的预定义数据域执行数据域发现，并标识企业中的关键数据特征。

注意：在下表中，星号 (*) 为通配符。

下表介绍了数据域加速器中可用的数据域：

名称	说明	相关规则类型	数据域组
Account_Status	发现与引用表中的帐户状态值相匹配的列数据。	数据规则	Account_Bank
AccountNumber	发现含有 “a*c*num” 或 “acc” 字符串的列名称。	列名称规则	Account_Bank、PCI、PHI
Age	发现含有 “age” 字符串或 “dob” 字符串的列名称，并发现值从 1 至 120 的列数据。	列名称规则 数据规则	PII
AlphaNumeric_SpecialCharacter	发现含有未格式化的字母数字数据和特殊字符数据的列数据。	数据规则	常规
Austria_NationalID	发现与奥地利居民身份标识相匹配的列数据。	数据规则	NationalID
BIC_SwiftCode	通过模式识别和国家/地区代码发现与银行识别代码 (Bank Identifier Code, BIC) 或环球银行金融电信协会 (Society for Worldwide Interbank Financial Telecommunication, SWIFT) 代码相匹配的列数据。	数据规则	Account_Bank
BinayValue	发现含有二进制值的列数据。	数据规则	常规
BirthDay	发现含有 “dob” 字符串、“date*of*bir*” 字符串或 “birth*da*” 字符串的列名称，并标识与有效出生日期相匹配的列数据。	列名称规则 数据规则	PII
BirthPlace	发现含有 “birth*place” 字符串或 “location*birth” 字符串的列名称。	列名称规则	PII
Brazil_IDDoc	发现与标题为 Registro Geral 的巴西身份证号码格式相匹配的列数据。	数据规则	NationalID
Brazil_NationalID	发现与巴西居民身份标识相匹配的列数据。	数据规则	NationalID
Bulgaria_NationalID	发现与保加利亚居民身份标识相匹配的列数据。	数据规则	NationalID
Canada_SIN	发现与加拿大社会保险号格式相匹配的列数据。	数据规则	NationalID
CertificateLicense Number	发现含有 “cert*lic*number” 字符串、“cert*lic*no*” 字符串、“lic* nu*” 字符串或 “lic*no*” 字符串的列名称。	列名称规则	PHI

名称	说明	相关规则类型	数据域组
China_NationalId	发现与中国居民身份标识相匹配的列数据。	数据规则	NationalID
CompanyName	发现含有“company”字符串的列名称，并识别与引用表中的组织名称值相匹配的列数据。	列名称规则 数据规则	Contact、PII
ComputerAddress	发现与计算机地址格式相匹配的列数据。	数据规则	常规
国家/地区	发现含有“iso*countr*code”字符串、“iso*country”字符串或“countr*”字符串的列名称，并识别与国家/地区名称相匹配的列数据。	列名称规则 数据规则	地址、PII
CountryCode_Phone	根据国际拨号代码发现电话号码。	数据规则	Contact
CreditCardNumber	发现含有“ccn”字符串、“cr*ca*nu”字符串或“credit*no*”字符串的列名称，并识别与多个信用卡组织的信用卡号格式相匹配的列数据。	列名称规则 数据规则	Account_Bank、PCI、PII
CreditCard_AMEX	发现与 American Express 信用卡号格式相匹配的列数据。	数据规则	Account_Bank
CreditCard_DinersCard	发现与 Diners Club 国际信用卡号格式相匹配的列数据。	数据规则	Account_Bank
CreditCard_DiscoverCard	发现与 Discover 信用卡号格式相匹配的列数据。	数据规则	Account_Bank
CreditCard_JCB	发现与 JCB 国际信用卡号格式相匹配的列数据。	数据规则	Account_Bank
CreditCard_MasterCard	发现与 MasterCard 信用卡号格式相匹配的列数据。	数据规则	Account_Bank
CreditCard_Visa	发现与 Visa 信用卡号格式相匹配的列数据。	数据规则	Account_Bank
Croatia_NationalID	发现与克罗地亚居民身份标识相匹配的列数据。	数据规则	NationalID
Date_AllFormats	发现列数据中的日期值。	数据规则	常规
Date_MM_DD_YYYY	发现源数据中以单一格式显示在日期列中的日期字符串。默认格式为“MM/DD/YYYY”。	数据规则	常规
Denmark_NationalID	发现与丹麦居民身份标识相匹配的列数据。	数据规则	NationalID

名称	说明	相关规则类型	数据域组
DeviceSerialNumber	发现含有“device*number”字符串、“device*no*”字符串、“serial*number”字符串、“serial*no*”字符串或“device*identi*”字符串的列名称。	列名称规则	PHI
DriverLicense_Canada	发现与加拿大（不列颠哥伦比亚、魁北克、马尼托巴和爱德华王子岛等四省除外）的驾驶证编号相匹配的列数据。	数据规则	NationalID
DriverLicense_GB R	发现与英国驾驶证编号相匹配的列数据。	数据规则	NationalID
DriverLicense_USA	发现与美国大多数州的驾驶证编号相匹配的列数据。	数据规则	NationalID
DrivingLicenseNumber	发现含有“license*”字符串或“driver*license”字符串的列名称。根据长度和模式要求，识别与英国、美国和某些加拿大省份的驾驶证编号相匹配的列数据。	列名称规则 数据规则	PII
电子邮件	发现含有“email”字符串的列名称，并识别与预定义的电子邮件 ID 格式相匹配的列数据。	列名称规则 数据规则	Contact、PHI
ExpirationDate	发现含有“exp*da*”字符串或“cr*exp*”字符串的列名称，并识别与信用卡过期日期相匹配的列数据。	列名称规则 数据规则	PCI
Finland_NationalID	发现与芬兰居民身份标识相匹配的列数据。	数据规则	NationalID
FirstName	发现含有“f*nam*”字符串的列名称，并识别与名字引用表中的值相匹配的列数据。	列名称规则 数据规则	Contact、PCI、PII
France_INSEE	发现与法国国家统计与经济研究所 (French Institut National de la Statistique et des Études Économiques, INSEE) 编码格式相匹配的列数据。	数据规则	NationalID
Gender	发现含有“gender”字符串或“female”和“male”等字符串的列名称，并识别与引用表中的性别值相匹配的列数据。	列名称规则 数据规则	Contact、PII
Geocode_Latitude	发现含有“latitude”字符串的列名称，并识别与有效的纬度坐标相匹配的列数据。	列名称规则 数据规则	地址、常规

名称	说明	相关规则类型	数据域组
Geocode_Latitude Longitude	发现含有“latitude”、“longitude”和“geocode”等字符串的列名称，并识别与有效的纬度坐标或经度坐标相匹配的列数据。	列名称规则 数据规则	地址、常规
Geocode_Longitude	发现含有“longitude”字符串的列名称，并识别与有效的经度坐标相匹配的列数据。	列名称规则 数据规则	地址、常规
Grade	发现含有“grade”字符串的列名称。	列名称规则	PII
GreatBritian_NINO	发现与英国国家社会保险号格式相匹配的列数据。	数据规则	NationalID
主机名	发现与有效的主机名相匹配的列数据。	数据规则	常规
IBAN	发现与多个欧洲国家/地区的国际银行账号格式相匹配的列数据。	数据规则	Account_Bank
India_NationalID	发现与印度永久帐号 (Indian Permanent Account Number) 格式相匹配的列数据。	数据规则	NationalID
IPAddress	发现含有“ip”字符串或“inter*port*add”字符串的列名称，并识别与预定义的 IP 地址格式相匹配的列数据。	列名称规则 数据规则	PII
ISBN	发现与国际标准书号格式相匹配的列数据。	数据规则	常规
Italy_FiscalCode	发现与意大利居民身份标识格式相匹配的列数据。	数据规则	NationalID
JobPosition	发现含有“title”字符串、“position”字符串或“designation”字符串的列名称。	列名称规则	PII
Korea_NationalID	发现与韩国居民身份标识相匹配的列数据。	数据规则	NationalID
LastName	发现含有“lname”字符串、“su*name”字符串或“last*name”字符串的列名称，并识别与姓氏引用表中的值相匹配的列数据。	列名称规则 数据规则	Contact、PCI、PII
Norway_NationalID	发现与挪威居民身份标识相匹配的列数据。	数据规则	NationalID

名称	说明	相关规则类型	数据域组
Passport_DEU_MR	发现与计算机可读取的德国护照号格式相匹配的列数据。	数据规则	NationalID
Passport_GBR	发现与英国护照号格式相匹配的列数据。	数据规则	NationalID
Passport_India	发现与印度护照号格式相匹配的列数据。	数据规则	NationalID
Passport_Machine Readable	发现与计算机可读取的所有国家/地区的护照号相匹配的列数据。	数据规则	NationalID
Passport_USA_MR	发现与计算机可读取的美国护照号格式相匹配的列数据。	数据规则	NationalID
PhoneNumber	发现含有“phone”字符串或“fax”字符串的列名称，并识别与美国电话号码格式相匹配的列数据。	列名称规则 数据规则	Contact、PHI
Postcode	发现与多个国家/地区的邮政编码相匹配的列数据。	数据规则	地址
Romania_National ID	发现与罗马尼亚居民身份标识相匹配的列数据。	数据规则	NationalID
Salary	发现含有“compensation”字符串、“salary”字符串或“wages”字符串的列名称。	列名称规则	PII
SouthAfrica_NationalID	发现与南非居民身份标识相匹配的列数据。	数据规则	NationalID
SSN	发现含有“SSN”字符串、“social*sec*no”字符串或“social*sec*num*”字符串的列名称，并识别与社会保障号格式相匹配的列数据。	列名称规则 数据规则	NationalID、PHI
SSN_General	发现与社会保障号格式相匹配的列数据。	数据规则	NationalID、PII
州	发现含有“add*sta”字符串、“state”字符串或“us*sta*”字符串的列名称，并识别与美国州名相匹配的列数据。	列名称规则 数据规则	PII

名称	说明	相关规则类型	数据域组
Street	发现含有以下其中一个字符串的列名称： - street - road - lane - court - avenue - way - blvd - boule*ard	列名称规则	PII
Sweden_NationalID	发现与瑞典居民身份标识相匹配的列数据。	数据规则	NationalID
Taiwan_NationalID	发现与中国台湾居民身份标识相匹配的列数据。	数据规则	NationalID
UniqueIdentifying Number	发现含有“unique*iden*number”字符串或“iden*num”字符串的列名称。	列名称规则	PHI
UPC	发现与通用产品代码相匹配的列数据。	数据规则	常规
URL	发现含有“uni*res*loc”字符串、“URL”字符串或“web”字符串的列名称，并识别与预定义的 URL 格式相匹配的列数据。	列名称规则 数据规则	PHI
USZip_5digit	发现与美国邮政编码相匹配的列数据。	数据规则	地址
VehicleRegPlateNumber	发现含有“registration”字符串、“number*plate”字符串、“license*plate”字符串或“vehicle*registration”字符串的列名称。	列名称规则	PII
ZipCode	发现含有“zip”字符串或“pin”字符串的列名称，并识别与美国邮政编码相匹配的列数据。	列名称规则 数据规则	PII

数据域加速器中的列名称规则

使用数据域列名称规则识别名称与规则所定义的列名称逻辑相匹配的数据列。每个规则都使用一个或多个正则表达式来搜索列名称可能包括的常用字符串。

例如，规则 `dataDomain_MetaDataRule_BIC_SWIFTCode` 包含一个可使用下列正则表达式搜索的标签创建器转换：

```
^*[iI][sS][oO].*[9][3][6][2].*$  
^*[sS][wW][iI][fF][tT]*[bB][iI][cC]$  
^*[bB][iI][cC].*[cC][oO][dD][eE].*$
```

列名称规则分析列名称中的字符。列名称规则不分析列中的数据值。

您可以在以下存储库位置找到列名称规则：

[Informatica_DQ_Content]\Domain_Discovery_MetaData_Rules

下表介绍了数据域加速器中的列名称规则：

名称	说明
dataDomain_MetaDataRule_ABARoutingNumber	搜索可以描述美国银行业协会路由编号列的名称。
dataDomain_MetaDataRule_AccountNumber	搜索可以描述帐号列的名称。
dataDomain_MetaDataRule_AccountStatus	搜索可以描述帐户状态信息列的名称。
dataDomain_MetaDataRule_Address	搜索可以描述地址信息列的名称。
dataDomain_MetaDataRule_AdmissionDate	搜索可以描述入院日期信息列的名称。
dataDomain_MetaDataRule_Age	搜索可以描述年龄或出生日期信息列的名称。
dataDomain_MetaDataRule_BankAccount	搜索可以描述银行帐户信息列的名称。
dataDomain_MetaDataRule_BIC_SwiftCode	搜索可以描述业务标识符代码列的名称。业务标识符代码也称为 SWIFT 代码和 ISO 9362 代码。
dataDomain_MetaDataRule_BirthDay	搜索可以描述出生日期或生日信息列的名称。
dataDomain_MetaDataRule_BirthPlace	搜索可以描述出生地列的名称。
dataDomain_MetaDataRule_CertificateLicenseNumber	搜索可以描述证书许可编号信息列的名称。
dataDomain_MetaDataRule_City	搜索可以描述城市信息列的名称。
dataDomain_MetaDataRule_CompanyName	搜索可以描述公司名称信息列的名称。
dataDomain_MetaDataRule_ComputerAddress	搜索可以描述计算机或 MAC 地址数据列的名称。

名称	说明
dataDomain_MetaDataRule_Country	搜索可以描述国家/地区信息（包括 ISO 国家/地区代码信息）列的名称。
dataDomain_MetaDataRule_CountryCode_Phone	搜索可以描述电话国家/地区代码列的名称。
dataDomain_MetaDataRule_County	搜索可以描述县信息列的名称。
dataDomain_MetaDataRule_CreditCardNumber	搜索可以描述信用卡号列的名称。
dataDomain_MetaDataRule_CreditCardTrack1FormatB	搜索可以描述磁道 1 格式 B 信用卡信息列的名称。
dataDomain_MetaDataRule_Currency	搜索可以描述货币信息列的名称。
dataDomain_MetaDataRule_DeviceSerialNumber	搜索可以描述设备编号或序列号信息列的名称。
dataDomain_MetaDataRule_DischargeDate	搜索可以描述出院日期信息列的名称。
dataDomain_MetaDataRule_DrivingLicenseNumber	搜索可以描述驾驶证信息列的名称。
dataDomain_MetaDataRule_Email	搜索可以描述电子邮件信息列的名称。
dataDomain_MetaDataRule_ExpirationDate	搜索可以描述过期日期信息列的名称，例如信用卡过期日期信息。
dataDomain_MetaDataRule_FirstName	搜索可以描述名字信息列的名称。
dataDomain_MetaDataRule_FullName	搜索可以描述全名信息列的名称。
dataDomain_MetaDataRule_Gender	搜索可以描述性别信息列的名称。
dataDomain_MetaDataRule_Grade	搜索可以描述年级信息列的名称。
dataDomain_MetaDataRule_HealthCareBeneficiaryNumber	搜索可以描述医疗保健受益人编号列的名称。
dataDomain_MetaDataRule_Height	搜索可以描述高度信息列的名称。
dataDomain_MetaDataRule_Hostname	搜索可以描述计算机主机名信息列的名称。
dataDomain_MetaDataRule_IBAN	搜索可以描述国际银行帐号列的名称。
dataDomain_MetaDataRule_ICD10	搜索可以描述来自《国际疾病与相关健康问题统计分类》第十次修订的值列的名称。
dataDomain_MetaDataRule_ICD9	搜索可以描述来自《国际疾病与相关健康问题统计分类》第九次修订的值列的名称。
dataDomain_MetaDataRule_IPAddress	搜索可以描述计算机 IP 地址信息列的名称。

名称	说明
dataDomain_MetaDataRule_ISBN	搜索可以描述国际标准书号列的名称。
dataDomain_MetaDataRule_ITIN_USA	搜索可以描述个体纳税人识别号列的名称。
dataDomain_MetaDataRule_JobPosition	搜索可以描述职称、职位或头衔信息列的名称。
dataDomain_MetaDataRule_LastName	搜索可以描述姓氏信息列的名称。
dataDomain_MetaDataRule_Latitude	搜索可以描述纬度信息列的名称。
dataDomain_MetaDataRule_LatitudeLongitude	搜索可以描述纬度、经度或地理坐标信息列的名称。
dataDomain_MetaDataRule_Longitude	搜索可以描述经度信息列的名称。
dataDomain_MetaDataRule_MaidenName	搜索可以描述婚前姓氏信息列的名称。
dataDomain_MetaDataRule_MiddleName	搜索可以描述中间名信息列的名称。
dataDomain_MetaDataRule_NationalId	搜索可以描述身份证号码列的名称。
dataDomain_MetaDataRule_NDC_USA	搜索可以描述国家药品编码信息列的名称
dataDomain_MetaDataRule_NPI_USA	搜索可以描述国家提供商标识编号列的名称。
dataDomain_MetaDataRule_Passport	搜索可以描述护照信息列的名称。
dataDomain_MetaDataRule_PhoneNumber	搜索可以描述电话号码或传真号码列的名称。
dataDomain_MetaDataRule_Quantity	搜索可以描述数量信息列的名称。
dataDomain_MetaDataRule_Race	搜索可以描述种族或颜色信息列的名称。
dataDomain_MetaDataRule_Religion	搜索可以描述宗教、信仰或信念信息列的名称。
dataDomain_MetaDataRule_Salary	搜索可以描述工资、薪水或报酬信息列的名称。
dataDomain_MetaDataRule_SSN	搜索可以描述社会保障号的名称。
dataDomain_MetaDataRule_State	搜索可以描述美国州信息列的名称。
dataDomain_MetaDataRule_Street	搜索可以描述街道地址信息列的名称。
dataDomain_MetaDataRule_UniqueIdentifyingNumber	搜索可以描述唯一标识号列的名称。
dataDomain_MetaDataRule_UPC_EAN	搜索可以描述通用产品代码或欧洲物品编号列的名称。
dataDomain_MetaDataRule_URL	搜索可以描述统一资源定位符或 Web 地址信息列的名称。

名称	说明
dataDomain_MetaDataRule_VehicleRegPlateNumber	搜索可以描述车辆登记或车牌号码列的名称。
dataDomain_MetaDataRule_Weight	搜索可以描述重量信息列的名称。
dataDomain_MetaDataRule_ZipCode	搜索可以描述邮政编码列的名称。

数据域加速器中的数据规则

使用数据域数据规则识别所含数据与规则标准相匹配的列。

您可以在以下存储库位置找到数据规则：

[Informatica_DQ_Content]\Domain_Discovery\Data_Rules

下表介绍了数据域加速器中的数据规则：

名称	说明
dataDomain_DataRule_ABARoutingNumber	识别与美国银行业协会路由编号的格式相匹配的列数据。路由编号标识金融交易中的金融机构。
dataDomain_DataRule_Account_Status	识别与引用数据中的帐户状态值相匹配的列数据。
dataDomain_DataRule_Address_Data	识别表示地址信息的列数据。该规则识别来自全球多个国家/地区的地址数据。
dataDomain_DataRule_Age	识别值从 1 至 120 的列数据。
dataDomain_DataRule_Alphanumeric_SpecialCharacter	识别含有未格式化的字母数字数据和特殊字符数据的列数据。
dataDomain_DataRule_Amount	识别表示物理数量的列数据。
dataDomain_DataRule_AUT_NATID	识别与奥地利居民身份标识格式相匹配的列数据。
dataDomain_DataRule_BankAccount_USA	识别与美国银行帐号格式相匹配的列数据。
dataDomain_DataRule_BGR_NATID	识别与保加利亚居民身份标识格式相匹配的列数据。
dataDomain_DataRule_BIC_SWIFTCode	通过模式识别和国家/地区代码来识别与银行识别代码 (Bank Identifier Code, BIC) 或环球银行金融电信协会 (Society for Worldwide Interbank Financial Telecommunication, SWIFT) 代码相匹配的列数据。
dataDomain_DataRule_BinaryValues	识别含有二进制值的列数据。
dataDomain_DataRule_BirthDay	识别与有效出生日期相匹配的列数据。此规则验证输入日期与当前日期之间的年数。根据从 1 至 120 的值返回 “Adult”、“Minor” 或 “Valid”。对于所有其他值，返回 “Invalid”。

名称	说明
dataDomain_DataRule_BRA_IDDoc	识别与巴西 <i>Registro Geral</i> 身份证编号格式相匹配的列数据。
dataDomain_DataRule_BRA_Personal_ID	识别与巴西个人身份标识格式相匹配的列数据。
dataDomain_DataRule_CAN_SIN	识别与加拿大社会保险号码格式相匹配的列数据。
dataDomain_DataRule_CHN_NATID	识别与中国居民身份证格式相匹配的列数据。
dataDomain_DataRule_City	识别包含有效城市名称的列数据。该规则读取包含国际城市名称的引用数据。
dataDomain_DataRule_CompanyName	识别与引用数据中的组织名称值相匹配的列数据。
dataDomain_DataRule_Computer_Address	识别与 IP 地址和 MAC 地址格式相匹配的列数据。
dataDomain_DataRule_Country	识别与 ISO 国家/地区名称匹配的列数据。
dataDomain_DataRule_CountryCode_Phone	识别与基于国际电话区号列表的电话号码相匹配的列数据。
dataDomain_DataRule_County	识别与美国县名称匹配的列数据。
dataDomain_DataRule_CreditCard_AmericanExpress	识别与 American Express 信用卡号格式相匹配的列数据。
dataDomain_DataRule_CreditCard_DinersCard	识别与 Diners Club 国际信用卡号格式相匹配的列数据。
dataDomain_DataRule_CreditCard_DiscoverCard	识别与 Discover 信用卡号格式相匹配的列数据。
dataDomain_DataRule_CreditCard_JCB	识别与 JCB 国际信用卡号格式相匹配的列数据。
dataDomain_DataRule_CreditCard_MasterCard	识别与 MasterCard 信用卡号格式相匹配的列数据。
dataDomain_DataRule_CreditCard_Visa	识别与 Visa 信用卡号格式相匹配的列数据。
dataDomain_DataRule_CreditCardNumber	识别与主要信用卡组织（如 American Express、Diners Club International 和 Maestro）的信用卡号格式相匹配的列数据。
dataDomain_DataRule_CreditCardTrack1FormatB	识别与磁道 1 格式 B 信用卡信息相匹配的列数据。
dataDomain_DataRule_Currency	识别与引用数据中的货币术语相匹配的列数据。
dataDomain_DataRule_Date_Validation	识别源数据中以单一格式显示在日期列中的日期字符串。要配置该规则执行验证所使用的日期格式，请打开规则中的 dq_ValidateDate 表达式转换并更新 In_Date_Format 表达式变量。默认格式为“MM/DD/YYYY”。该规则将返回“有效”或“无效”。

名称	说明
dataDomain_DataRule_Date_Validation_All_Formats	识别列数据中的日期值，并将该列数据标准化为单个日期格式。
dataDomain_DataRule_DEU_Machine_Readable_Passport	识别与计算机可读的德国护照号格式相匹配的列数据。
dataDomain_DataRule_DNK_NATID	识别与丹麦居民身份标识格式相匹配的列数据。
dataDomain_DataRule_DriversLicense	根据数据值的长度和模式，识别与加拿大、英国和美国驾驶证编号相匹配的列数据。
dataDomain_DataRule_DriversLicense_Canada	识别与加拿大（不列颠哥伦比亚、魁北克、马尼托巴和爱德华王子岛等四省除外）的驾驶证编号相匹配的列数据。
dataDomain_DataRule_DriversLicense_Canada_narrow	识别与加拿大（不列颠哥伦比亚、魁北克、马尼托巴和爱德华王子岛等四省除外）的驾驶证编号相匹配的列数据。 该规则类似于 dataDomain_DataRule_DriversLicense_Canada 规则。然而，dataDomain_DataRule_DriversLicense_Canada_narrow 执行更窄的分析，以减少误报的可能性。
dataDomain_DataRule_DriversLicense_GBR	识别与英国驾驶证编号相匹配的列数据。
dataDomain_DataRule_DriversLicense_narrow	识别与英国以及加拿大和美国的许多州和省的驾驶证编号相匹配的列数据。 该规则不验证不列颠哥伦比亚、魁北克、马尼托巴和爱德华王子岛等四省的驾驶证编号。 为减少误报的可能性，该规则不验证包含四到八位数的驾驶证编号。
dataDomain_DataRule_DriversLicense_USA	识别与美国大多数州的驾驶证编号相匹配的列数据。
dataDomain_DataRule_DriversLicense_USA_narrow	识别与美国大多数州的驾驶证编号相匹配的列数据。 为减少误报的可能性，该规则排除了由六到八位数组成的数据值。例如，该规则排除了一个值，如 01012017。
dataDomain_DataRule_Email	识别与预定义的电子邮件 ID 格式相匹配的列数据。
dataDomain_DataRule_ExpirationDate	识别与信用卡过期日期相匹配的列数据。此规则将输入日期与系统日期进行比较验证。
dataDomain_DataRule_FIN_NATID	识别与芬兰居民身份标识格式相匹配的列数据。
dataDomain_DataRule_FirstName	识别与名字的引用数据集中的值相匹配的列数据。
dataDomain_DataRule_FRA_INSEE	识别与法国国家统计与经济研究所 (French Institut National de la Statistique et des Études Économiques, INSEE) 编码格式相匹配的列数据。
dataDomain_DataRule_FullName	识别包含名字、中间名和姓氏的数据列中的字符串。该规则将每个字符串中的单词与引用数据进行比较。
dataDomain_DataRule_GBR_NINO	识别与英国国家社会保险号格式相匹配的列数据。

名称	说明
dataDomain_DataRule_GBR_Passport_Number	识别与英国护照号格式相匹配的列数据。
dataDomain_DataRule_Gender	识别与引用数据中的性别值相匹配的列数据。
dataDomain_DataRule_Height	识别具有值 1 到 8 的列数据，其中 8 表示高度（英尺）。
dataDomain_DataRule_HostName	识别与有效的主机名相匹配的列数据。
dataDomain_DataRule_HRV_NATID	识别与克罗地亚居民身份标识格式相匹配的列数据。
dataDomain_DataRule_IBAN	识别与多个欧洲国家/地区的国际银行帐号格式相匹配的列数据。
dataDomain_DataRule_ICD_10	识别与《国际疾病与相关健康问题统计分类》(International Statistical Classification of Diseases and Related Health Problems, ICD) 第十次修订中的条件名称相匹配的列数据。世界卫生组织 (World Health Organization, WHO) 维护分类。
dataDomain_DataRule_ICD_9	识别与《国际疾病与相关健康问题统计分类》(International Statistical Classification of Diseases and Related Health Problems, ICD) 第九次修订中的条件名称相匹配的列数据。世界卫生组织 (World Health Organization, WHO) 维护分类。
dataDomain_DataRule_IND_NATID	识别与印度永久帐号 (Indian Permanent Account Number) 格式相匹配的列数据。
dataDomain_DataRule_IND_Passport	识别与印度护照号格式相匹配的列数据。
dataDomain_DataRule_IPAddress	识别与预定义的 IP 地址格式相匹配的列数据。
dataDomain_DataRule_ISBN	识别与国际标准书号格式相匹配的列数据。
dataDomain_DataRule_ISIN	识别与国际证券识别码 (international securities identification number, ISIN) 格式相匹配的列数据。ISIN 唯一标识证券，如股票或债券。
dataDomain_DataRule_ItalyFiscalCode	识别与意大利居民身份标识格式相匹配的列数据。
dataDomain_DataRule_ITIN_USA	识别与美国个体纳税人识别号 (Individual Taxpayer Identification Number, ITIN) 的格式相匹配的列数据。美国国家税务局签发识别号。
dataDomain_DataRule_JobPosition	识别与引用数据中的职位名称相匹配的列数据。
dataDomain_DataRule_KOR_NATID	识别与韩国居民身份标识格式相匹配的列数据。
dataDomain_DataRule_LastName	识别姓氏的引用数据集中的值相匹配的列数据。
dataDomain_DataRule_Latitude	识别与有效的纬度坐标相匹配的列数据。
dataDomain_DataRule_LatitudeLongitude	识别与有效的纬度坐标和经度坐标对（每对以分号分隔）相匹配的列数据。
dataDomain_DataRule_Longitude	识别与有效的经度坐标相匹配的列数据。

名称	说明
dataDomain_DataRule_Machine_Readable_Passport	识别与计算机可读取的所有国家/地区的护照号相匹配的列数据。
dataDomain_DataRule_NDC_USA	识别与美国国家药品编码目录中的国家药品编码 (National Drug Code, NDC) 值相匹配的列数据。每个编码都唯一标识制造商开发的人用药品。
dataDomain_DataRule_NOR_NATID	识别与挪威居民身份标识格式相匹配的列数据。
dataDomain_DataRule_NPI_USA	识别与美国国家提供商标识 (National Provider Identifier, NPI) 编号相匹配的列数据。医疗保险和医疗补助服务中心 (Centers for Medicare and Medicaid Services) 向卫生保健服务提供商签发编号。
dataDomain_DataRule_PhoneNumber	识别与美国电话号码格式相匹配的列数据。
dataDomain_DataRule_PostCode	识别与多个国家/地区的邮政编码相匹配的列数据。
dataDomain_DataRule_Quantity	识别描述物理数量并包含度量单位的列数据。
dataDomain_DataRule_Race	识别与引用数据中的人员种族名称相匹配的列数据。
dataDomain_DataRule_Religion	识别与引用数据中的宗教名称相匹配的列数据。
dataDomain_DataRule_ROU_NATID	识别与罗马尼亚居民身份标识格式相匹配的列数据。
dataDomain_DataRule_SouthAfrica_NATID	识别与南非居民身份标识格式相匹配的列数据。
dataDomain_DataRule_Spanish_NIF	识别与西班牙财务标识号 (fiscal identification number, NIF) 格式相匹配的列数据。
dataDomain_DataRule_SSN	识别与美国社会保障号格式相匹配的列数据。
dataDomain_DataRule_State	识别与美国州名相匹配的列数据。
dataDomain_DataRule_Street	识别列数据中描述街道地址信息的字符串，例如，街道、路、大街。该规则使用正则表达式在列数据中查找街道标识符。
dataDomain_DataRule_SWE_NATID	识别与瑞典居民身份标识格式相匹配的列数据。
dataDomain_DataRule_TWN_NATID	识别与中国台湾居民身份标识格式相匹配的列数据。
dataDomain_DataRule_UPC	识别与有效的通用产品代码相匹配的列数据。通用产品代码是一种条形码。
dataDomain_DataRule_UPC_EAN	识别与有效的通用产品代码或欧洲物品编号相匹配的列数据。通用产品代码和欧洲物品编号是条形码类型。
dataDomain_DataRule_URL	识别与预定义的 URL 格式相匹配的列数据。
dataDomain_DataRule_US_Zip5	识别与美国邮政编码相匹配的列数据。
dataDomain_DataRule_USA_Machine_Readable_Passport	识别与计算机可读取的美国护照号格式相匹配的列数据。

名称	说明
dataDomain_DataRule_USA_SSN_post_2011June	识别在长度、数值以及区域、组和序列号各区段的最小值和最大值方面与社会保障号格式相匹配的列数据。根据 2011 年 6 月 25 日生效的 SSN 随机发放方法，该规则不会验证社会保障号的发放情况以及组与区域号的组合。
dataDomain_DataRule_Weight	识别描述重量值的列数据。该规则查找介于 0 和 500 之间的数字。
dataDomain_DataRule_ZipCode	识别与美国邮政编码相匹配的列数据。

第 4 章

澳大利亚/新西兰加速器

本章包括以下主题：

- [澳大利亚/新西兰加速器概览, 45](#)
- [澳大利亚/新西兰地址数据清理规则, 45](#)
- [澳大利亚/新西兰复合规则, 47](#)
- [澳大利亚/新西兰联系人数据清理规则, 49](#)
- [澳大利亚/新西兰公司数据清理规则, 51](#)
- [澳大利亚/新西兰常规数据清理规则, 52](#)
- [澳大利亚/新西兰匹配和消除重复记录规则, 53](#)
- [澳大利亚/新西兰演示映射, 55](#)

澳大利亚/新西兰加速器概览

使用澳大利亚/新西兰加速器中的规则可验证和改善澳大利亚和新西兰的组织的数据。

澳大利亚/新西兰加速器包括执行以下数据质量操作的规则：

- 地址数据清理
- 联系人数据清理
- 公司数据清理
- 常规数据清理
- 匹配和消除重复记录

澳大利亚/新西兰加速器也包括复合规则。复合规则将多个规则合并为单个对象。

该加速器依赖于核心加速器安装的规则。

澳大利亚/新西兰地址数据清理规则

使用地址数据清理规则可解析、标准化和验证地址数据。

您可以在以下存储库位置找到地址数据清理规则：

[Informatica_DQ_Content]\Rules\Address_Data_Cleansing

下表介绍了澳大利亚/新西兰加速器中的地址数据清理规则：

名称	说明
rule_AUS_Address_Parse_Hybrid	将非结构化的澳大利亚地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_AUS_Address_Parse_Multiline	将非结构化的澳大利亚地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_AUS_Address_Validation_Discrete	验证澳大利亚地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_AUS_Address_Validation_Discrete_w_Geocoding	验证澳大利亚地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_AUS_Address_Validation_Hybrid	验证澳大利亚地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_AUS_Address_Validation_Hybrid_w_Geocoding	验证澳大利亚地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_AUS_Address_Validation_Multiline	验证澳大利亚地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_AUS_Address_Validation_Multiline_w_Geocoding	验证澳大利亚地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_NZL_Address_Parse_Hybrid	将非结构化的新西兰地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_NZL_Address_Parse_Multiline	将非结构化的新西兰地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_NZL_Address_Validation_Discrete	验证新西兰地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_NZL_Address_Validation_Discrete_w_Geocoding	验证新西兰地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_NZL_Address_Validation_Hybrid	验证新西兰地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_NZL_Address_Validation_Hybrid_w_Geocoding	验证新西兰地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。

名称	说明
rule_NZL_Address_Validation_Multiline	验证新西兰地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_NZL_Address_Validation_Multiline_w_Geocoding	验证新西兰地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。

澳大利亚/新西兰复合规则

使用澳大利亚/新西兰加速器中的复合规则，将一组规则作为单个对象添加到映射中。复合规则是一种利用其他加速器规则的逻辑的规则。

您可以在以下存储库位置找到复合规则：

[Informatica_DQ_Content]\Rules\Composite_Rules

下表介绍了澳大利亚/新西兰加速器中的复合规则：

名称	说明
rule_AUS_Contact_Data	解析、标准化和验证地址、电话号码和税务档案编号等澳大利亚联系人数据。
rule_NZL_Contact_Data	解析、标准化和验证地址、电话号码和税务局 (Inland Revenue Department, IRD) 编号等新西兰联系人数据。

适用于澳大利亚联系人数据的复合规则

规则 rule_AUS_Contact_Data 从存储库中的多个文件夹读取 Mapplet。该规则还包括一个不可重用的转换。

下表列出了 rule_AUS_Contact_Data 中的规则和转换的名称与存储库位置：

规则	位置
Case_Converter	不可重用的转换
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_AUS_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_AUS_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_AUS_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_AUS_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

规则	位置
rule_AUS_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_AUS_Phone_Number_Validation	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_AUS_Tax_File_Number_Standardization	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_AUS_Tax_File_Number_Validation	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing

适用于新西兰联系人数据的复合规则

规则 rule_NZL_Contact_Data 从存储库中的多个文件夹读取 Mapplet。该规则还包括一个不可重用的转换。

下表列出了 rule_NZL_Contact_Data 中的规则和转换的名称与存储库位置：

规则	位置
Case_Converter	不可重用的转换
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules \General_Data_Cleansing
rule_AUS_Company_Name_Standardization	[Informatica_DQ_Content]\Rules \Corporate_Data_Cleansing
rule_AUS_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_NZL_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules \Address_Data_Cleansing
rule_NZL_Gender_Assignment	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_NZL_IRD_Number_Standardization	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_NZL_IRD_Number_Validate	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing

规则	位置
rule_NZL_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\ Contact_Data_Cleansing
rule_NZL_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\ Contact_Data_Cleansing
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules\ Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules\ Contact_Data_Cleansing

澳大利亚/新西兰联系人数据清理规则

使用联系人数据清理规则可解析、标准化和验证有关业务联系人和个人的数据。

您可以在以下存储库位置找到联系人数据清理规则：

[Informatica_DQ_Content]\Rules\
Contact_Data_Cleansing

下表介绍了澳大利亚/新西兰加速器中的联系人数据清理规则：

名称	说明
rule_AUS_Driver_Licence_Number_Validation	根据长度和模式要求验证澳大利亚驾驶员驾照编号。
rule_AUS_Gender_Assignment	按照名字分配性别。该规则会为男性名字返回“M”，为女性名字返回“F”，为未知性别返回“U”。例如，该规则会为姓名“John Smith”分配代表男性的“M”性别。
rule_AUS_Given_Name_Standard	根据澳大利亚昵称生成名字。
rule_AUS_Multi_Person_Name_Parse	将人名值解析为单独的端口。该规则会为头衔、名字、中间名和姓氏等值创建端口。 规则输出包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。 当姓名数据标识了多位人员时，该规则会为每个全名创建一个输出端口。例如，该规则可读取姓名“John and Jane Smith”，并为“John Smith”和“Jane Smith”创建输出端口。
rule_AUS_Personal_Name_Parsing_FML	将人名中的值解析为单独的端口。 该规则按以下顺序创建端口： - 名字、中间名、姓氏 规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。

名称	说明
rule_AUS_Personal_Name_Parsing_LFM	<p>将人名中的值解析为单独的端口。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 姓氏、名字、中间名 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p>
rule_AUS_Phone_Number_Parse	<p>解析字符串中的澳大利亚电话号码。该规则将解析数据中的第一个电话号码（从右到左读取）。</p> <p>该规则会识别使用前导零、国际拨号代码或以 # 号开头的分机号的电话号码。该规则可处理以下标点符号：加号、括号和 # 号。在运行该规则前，请删除所有其他标点，包括双空格。</p> <p>该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。</p>
rule_AUS_Phone_Number_Standardization	<p>将澳大利亚电话号码标准化为国际和本地拨号格式。该规则会识别使用前导零、国际拨号代码或以 # 号开头的分机号的电话号码。</p>
rule_AUS_Phone_Number_Validation	<p>验证澳大利亚电话号码的区号和长度。该规则将返回电话号码的所属地区，以及表示电话号码的区号和长度是否有效的代码。</p>
rule_AUS_Tax_File_Number_Parse	<p>解析澳大利亚税务档案编号 (Tax File Number, TFN)。</p>
rule_AUS_Tax_File_Number_Standardization	<p>将澳大利亚税务档案编号 (Tax File Number, TFN) 标准化。要配置标准化格式，请编辑 dq_Format_TFN 表达式转换中的 TFN_Format 表达式变量。默认值为 “No_punctuation”。</p>
rule_AUS_Tax_File_Number_Validation	<p>根据每个编号中的校验位验证澳大利亚税务档案编号 (Tax File Number, TFN)。</p>
rule_NZL_Gender_Assignment	<p>按照新西兰名字分配性别。该规则会为男性名字返回 “M”，为女性名字返回 “F”，为未知性别返回 “U”。例如，该规则会为姓名 “John Smith” 分配代表男性的 “M” 性别。</p>
rule_NZL_Given_Name_Standard	<p>根据新西兰昵称生成名字。</p>
rule_NZL_IRD_Number_Parse	<p>将九位数字字符串解析为新西兰税务局 (Inland Revenue Department, IRD) 编号。</p>
rule_NZL_IRD_Number_Standardization	<p>将新西兰税务局 (Inland Revenue Department, IRD) 编号标准化。要配置标准化格式，请编辑 dq_Format_IRD 表达式转换中的 IRD_Format 表达式变量。默认值为 “No_punctuation”。该规则要求输入内容为九位数字字符串。</p>
rule_NZL_IRD_Number_Validate	<p>根据每个编号中的校验位验证新西兰税务局 (Inland Revenue Department, IRD) 编号。</p>

名称	说明
rule_NZL_Phone_Number_Parse	<p>解析字符串中的新西兰电话号码。该规则将解析数据中的第一个电话号码（从右到左读取）。</p> <p>该规则会识别使用前导零、国际拨号代码或以 # 号开头的分机号的电话号码。该规则可处理以下标点符号：加号、括号和 # 号。在运行该规则前，请删除所有其他标点，包括双空格。</p> <p>该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。</p>
rule_NZL_Phone_Number_Standardization	将新西兰电话号码标准化为国际和本地拨号格式。该规则会识别使用前导零、国际拨号代码或以 # 号开头的分机号的电话号码。
rule_NZL_Phone_Number_Validation	验证新西兰电话号码的区号和长度。该规则将返回电话号码的所属地区，以及表示电话号码的区号和长度是否有效的代码。
rule_Prenome_Assignment	根据性别生成敬称。您可以将 female_prenome 表达式变量从 “Ms.” 更改为 “Mrs”。
rule_Salutation_Assignment	根据名字和姓名标志生成正式和非正式问候。例如，当输入数据包含 “Mr. John Smith,” 时，该规则会生成正式问候 “Dear Mr. Smith,” 和非正式问候 “Dear John,”。您可以通过编辑 dq_Generate_Salutation Expression 表达式转换中的变量来更改前缀和标点。

与核心联系人数据清理规则的相关性

澳大利亚/新西兰加速器依赖于核心加速器中的以下联系人数据清理规则：

- rule_Email_Validation

有关这些规则的详细信息，请参阅 [“核心联系人数据清理规则” 页面上 21](#)。

澳大利亚/新西兰公司数据清理规则

使用公司数据清理规则可解析、标准化和验证公司数据。

您可以在以下存储库位置找到公司数据清理规则：

[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

下表介绍了澳大利亚/新西兰加速器中的公司数据清理规则：

名称	说明
rule_AUS_Business_Number_Parse	将 11 位数字字符串解析为澳大利亚商业代码 (Australian Business Number, ABN)。
rule_AUS_Business_Number_Standardize	将澳大利亚商业代码 (Australian Business Number, ABN) 标准化为 “NN NNN NNN NNN” 格式。该规则要求输入内容为 11 位数字字符串。

名称	说明
rule_AUS_Business_Number_Validation	根据每个代码中的校验位验证澳大利亚商业代码 (Australian Business Number, ABN)。
rule_AUS_Company_Name_Standardization	将公司名称标准化为澳大利亚引用表值。

澳大利亚/新西兰常规数据清理规则

使用常规数据清理规则可识别输入字段中所含信息的类型。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]\Rules\General_Data_Cleansing

下表介绍了澳大利亚/新西兰加速器中的常规数据清理规则：

名称	说明
rule_AUS_NZL_NER_Field_Identification	识别输入字段中包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和澳大利亚和新西兰地址数据。该规则会返回一个说明输入数据类型的标签。该规则使用概率匹配技术识别信息类型。

与核心常规数据清理规则的相关性

澳大利亚/新西兰加速器依赖于核心加速器中的以下常规数据清理规则：

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Description
- rule_Remove_Extra_Spaces
- rule_Remove_Hyphen
- rule_Remove_Leading_Zero
- rule_Remove_Period_Parentheses
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Remove_Space
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

有关这些规则的详细信息，请参阅 [“核心常规数据清理规则” 页面上 22](#)。

澳大利亚/新西兰匹配和消除重复记录规则

使用澳大利亚/新西兰加速器中的匹配和消除重复记录规则可度量数据集中记录之间的相似性级别。

您可以在以下存储库位置找到匹配和消除重复记录规则：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了澳大利亚/新西兰加速器中的匹配和消除重复记录规则：

名称	说明
mplt_AUS_Firstname_and_TFN_Match	使用字段匹配策略根据税务档案编号 (Tax File Number, TFN) 和名字识别澳大利亚数据中的重复行。该 Mapplet 会根据 TFN 数据生成组键。
mplt_AUS_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别澳大利亚数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_AUS_IMO_Familyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别澳大利亚数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_AUS_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别澳大利亚数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_AUS_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别澳大利亚数据中的重复行。个人数据列中的字段必须包含单一类型的数据，如电话号码、电子邮件或税务档案编号。该 Mapplet 会根据个人数据生成组键。
mplt_AUS_Individual_Name_and_Address_Match	使用字段匹配策略根据人名和澳大利亚地址数据识别重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_AUS_Individual_Name_and_Date_Match	使用字段匹配策略根据澳大利亚人名和日期识别重复行。该 Mapplet 会生成根据日期数据生成的组键。
mplt_AUS_Individual_Name_and_Email_Match	使用字段匹配策略根据电子邮件地址和澳大利亚人名识别重复行。该 Mapplet 会根据电子邮件地址数据生成组键。
mplt_AUS_Individual_Name_and_Phone_Match	使用字段匹配策略根据澳大利亚人名和电话号码识别重复行。该 Mapplet 会根据电话号码数据生成组键。
mplt_AUS_Individual_Name_and_TFN_Match	使用字段匹配策略根据税务档案编号 (Tax File Number, TFN) 和人名识别澳大利亚数据的重复行。该 Mapplet 会根据 TFN 数据生成组键。
mplt_AUS_Individual_Name_Match	使用字段匹配策略根据澳大利亚人名识别重复行。Mapplet 将根据姓氏值生成 NYSIIS 代码，并使用该 NYSIIS 代码作为组键。
mplt_AUS_NZL_Company_Name_and_Address_Match	使用字段匹配策略根据澳大利亚和新西兰的公司名称和地址数据识别重复行。该 Mapplet 将使用公司名称值和邮政编码值的字符组合生成组键。
mplt_AUS_NZL_Familyname_and_Address_Match	使用字段匹配策略根据澳大利亚和新西兰的姓氏和地址数据识别重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_Company_Name_Match	使用字段匹配策略根据公司名称识别重复行。该 Mapplet 将根据公司名称值生成 Soundex 代码，并使用该 Soundex 代码作为组键。
mplt_NZL_Firstname_and_IRD_Match	使用字段匹配策略根据税务局 (Inland Revenue Department, IRD) 编号和名字识别新西兰数据的重复行。该 Mapplet 会根据 IRD 编号生成组键。

名称	说明
mplt_NZL_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别新西兰数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_NZL_IMO_Familyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别新西兰数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_NZL_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别新西兰数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_NZL_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别新西兰数据中的重复行。个人数据列中的字段必须包含单一类型的数据，如电话号码、电子邮件或税务局 (Inland Revenue Department, IRD) 编号。该 Mapplet 会根据个人数据生成组键。
mplt_NZL_Individual_Name_and_Address_Match	使用字段匹配策略根据人名和新西兰地址数据识别重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_NZL_Individual_Name_and_Date_Match	使用字段匹配策略根据新西兰人名和日期识别重复行。该 Mapplet 会生成根据日期数据生成的组键。
mplt_NZL_Individual_Name_and_Email_Match	使用字段匹配策略根据电子邮件地址和新西兰人名识别重复行。该 Mapplet 会根据电子邮件地址数据生成组键。
mplt_NZL_Individual_Name_and_IRD_Match	使用字段匹配策略根据新西兰人名和税务局 (Inland Revenue Department, IRD) 编号识别重复行。该 Mapplet 会根据 IRD 编号生成组键。
mplt_NZL_Individual_Name_and_Phone_Match	使用字段匹配策略根据新西兰人名和电话号码识别重复行。该 Mapplet 会根据电话号码数据生成组键。
mplt_NZL_Individual_Name_Match	使用字段匹配策略根据新西兰人名识别重复行。Mapplet 将根据姓氏值生成 NYSIS 代码，并使用该 NYSIS 代码作为组键。
rule_AUS_NZL_Company_Name_and_Address_MatchScore	根据公司名称和澳大利亚和新西兰地址生成匹配得分。
rule_AUS_NZL_Familyname_and_Address_MatchScore	根据澳大利亚和新西兰的姓氏和地址生成匹配得分。
rule_AUS_NZL_Firstname_and_PID_MatchScore	根据名字和个人身份号码生成匹配得分。
rule_AUS_NZL_Individual_Name_and_Address_MatchScore	根据澳大利亚和新西兰的人名和地址生成匹配得分。
rule_AUS_NZL_Individual_Name_and_PID_MatchScore	根据人名和个人身份号码生成匹配得分。
rule_Company_Name_MatchScore	根据公司名称生成匹配得分。
rule_Individual_Name_and_Date_MatchScore	根据人名和日期生成匹配得分。

名称	说明
rule_Individual_Name_and_Email_MatchScore	根据人名和电子邮件地址生成匹配得分。
rule_Individual_Name_and_Phone_MatchScore	根据人名和电话号码生成匹配得分。
rule_Individual_Name_MatchScore	根据人名生成匹配得分。

澳大利亚/新西兰演示映射

澳大利亚/新西兰加速器中的演示映射使用多个规则演示数据质量过程。

您可以在以下存储库位置找到演示映射：

[Informatica_DQ_Content]\Rules_Demo\AUS_NZL_Accelerator

该加速器包含以下演示映射：

m_AUS_customer_data_demo

解析、标准化和验证澳大利亚和新西兰数据。

m_AUS_customer_matching_demo

解析澳大利亚和新西兰的标识数据并将其标准化，并对该数据执行标识匹配分析。

该映射会分析以下数据组合，并为每个组合生成匹配群集：

- 人名和地址数据
- 人名和电话号码

第 5 章

巴西加速器

本章包括以下主题：

- [巴西加速器概览, 56](#)
- [巴西地址数据清理规则, 56](#)
- [巴西复合规则, 57](#)
- [巴西联系人数据清理规则, 58](#)
- [巴西公司数据清理规则, 59](#)
- [巴西常规数据清理规则, 60](#)
- [巴西匹配和消除重复记录规则, 61](#)
- [巴西演示映射, 62](#)

巴西加速器概览

使用巴西加速器中的规则可验证和改善巴西组织的数据。

巴西加速器包括执行以下数据质量操作的规则：

- 地址数据清理
- 联系人数据清理
- 公司数据清理
- 常规数据清理
- 匹配和消除重复记录

巴西加速器还包括一个复合规则。复合规则将多个规则合并为单个对象。

该加速器依赖于核心加速器安装的规则。

巴西地址数据清理规则

使用地址数据清理规则可解析、标准化和验证地址数据。

您可以在以下存储库位置找到地址数据清理规则：

[Informatica_DQ_Content]\Rules\Address_Data_Cleansing

下表介绍了巴西加速器中的地址数据清理规则：

名称	说明
rule_BRA_Address_Parse_Hybrid	将非结构化的巴西地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_BRA_Address_Parse_Multiline	将非结构化的巴西地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_BRA_Address_Validation_Discrete	验证巴西地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_BRA_Address_Validation_Discrete_w_Geocoding	验证巴西地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_BRA_Address_Validation_Hybrid	验证巴西地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_BRA_Address_Validation_Hybrid_w_Geocoding	验证巴西地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_BRA_Address_Validation_Multiline	验证巴西地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_BRA_Address_Validation_Multiline_w_Geocoding	验证巴西地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。

巴西复合规则

使用巴西加速器中的复合规则，将一组规则作为单个对象添加到映射中。复合规则是一种利用其他加速器规则的逻辑的规则。

您可以在以下存储库位置找到复合规则：

[Informatica_DQ_Content]\Rules\Composite_Rules

下表介绍了巴西加速器中的复合规则：

名称	说明
rule_BRA_Contact_Data	解析、标准化和验证地址、电话号码和个人税号 (Cadastro de Pessoas Físicas, CPF) 等巴西联系人数据。

规则 rule_BRA_Contact_Data 从存储库中的多个文件夹读取 Mapplet。该规则还包括一个不可重用的转换。

下表列出了 rule_BRA_Contact_Data 中的规则和转换的名称与存储库位置：

规则	位置
Case_Converter	不可重用的转换
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_BRA_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_BRA_Company_Suffix_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_BRA_Personal_CPF_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Personal_Name_Parse_Validate	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Phone_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Prenome_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_BRA_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

巴西联系人数据清理规则

使用联系人数据清理规则可解析、标准化和验证有关业务联系人和个人的数据。

您可以在以下存储库位置找到联系人数据清理规则：

[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

下表介绍了巴西加速器中的联系人数据清理规则：

名称	说明
rule_BRA_Gender_Assignmen t	按照名字分配性别。该规则会为男性名字返回“M”，为女性名字返回“F”，为未知性别返回“U”。例如，该规则会为姓名“Joao Coelho”分配代表男性的“M”性别。
rule_BRA_Given_Name_Stan dard	根据巴西昵称生成名字。

名称	说明
rule_BRA_Personal_CPF_Validation	验证个人税号 (Cadastro de Pessoas Físicas, CPF) 的校验位。
rule_BRA_Personal_Name_Parse_Validate	将人名值解析为单独的端口。该规则会为头衔、名字、中间名和姓氏等值创建端口。规则还指示名称是否可能为公司名称，并验证名称拼写。 规则输出包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。
rule_BRA_Personal_PIS_PASEP_Validation	验证巴西社会保险号。
rule_BRA_Personal_Voter_Registration_Validation	验证巴西选民登记号中的校验位。
rule_BRA_Phone_Number_Parse	解析字符串中的巴西电话号码。该规则将解析数据中的第一个电话号码（按照从左到右的顺序读取）。该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。
rule_BRA_Phone_Number_Standardization	将巴西电话号码标准化。该规则可返回以下格式的电话号码： - 标准 - nn nnnn nnnn - 短划线 - nn-nnnn-nnnn - 无空格 - nnnnnnnnnn
rule_BRA_Phone_Validation	验证巴西电话号码的区号和长度。该规则将返回表示电话号码区号和长度是否有效的代码。
rule_BRA_Prenome_Assignment	根据性别生成敬称。您可以将 female_prenome 表达式变量从“Sra”更改为“Sta”。
rule_BRA_Salutation_Assignment	根据名字和姓名标志生成正式和非正式问候。例如，当输入数据包含“Sr. Joao Coelho,”时，该规则会生成正式问候“Prezado Sr. Coelho,”和非正式问候“Prezado Joao,”。您可以通过编辑 dq_Generate_Salutation Expression 表达式转换中的变量来更改前缀和标点。

与核心联系人数据清理规则的相关性

巴西加速器依赖于核心加速器中的以下联系人数据清理规则：

- rule_Email_Parse_Into_Mailbox_Domain
- rule_Email_Validation

有关这些规则的详细信息，请参阅 [“核心联系人数据清理规则” 页面上 21。](#)

巴西公司数据清理规则

使用巴西加速器中的公司数据清理规则可将公司数据标准化并验证这些数据。

您可以在以下存储库位置找到公司数据清理规则：

[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

下表介绍了巴西加速器中的公司数据清理规则：

名称	说明
rule_BRA_Company_CNPJ_Validation	验证国家公司税号 (Cadastro Nacional da Pessoa Jurídica, CNPJ)。CNPJ 号用于标识巴西公司。
rule_BRA_Company_Suffix_Standardization	将巴西公司后缀标准化。

巴西常规数据清理规则

使用常规数据清理规则可识别输入字段中所含信息的类型。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]\Rules\General_Data_Cleansing

下表介绍了巴西加速器中的常规数据清理规则：

名称	说明
rule_BRA_NER_Field_Identification	识别输入字段中包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和巴西地址数据。该规则会返回一个说明输入数据类型的标签。该规则使用引用数据和概率匹配技术识别信息类型。

与核心常规数据清理规则的相关性

巴西加速器依赖于核心加速器中的以下常规数据清理规则：

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Remove_Extra_Spaces
- rule_Remove_Non_Numbers
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Replace_Limited_Punct_with_Space
- rule_TitleCase
- rule_UpperCase

有关这些规则的详细信息，请参阅 [“核心常规数据清理规则” 页面上 22。](#)

巴西匹配和消除重复记录规则

使用匹配和消除重复记录规则可度量数据集中记录之间的相似性级别。

您可以在以下存储库位置找到匹配和消除重复记录规则：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了巴西加速器中的匹配和消除重复记录规则：

名称	说明
mplt_BRA_Company_Name_and_Address_Match	使用字段匹配策略根据公司名称和巴西地址数据识别重复行。该 Mapplet 将使用公司名称值和邮政编码值的字符组合生成组键。
mplt_BRA_Familyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别巴西数据中的重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_BRA_Firstname_and_CPF_Match	使用字段匹配策略根据名字和个人税号 (Cadastro de Pessoas Físicas, CPF) 识别重复行。该 Mapplet 会根据 CPF 号生成组键。
mplt_BRA_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别巴西数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_BRA_IMO_Familyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别巴西数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_BRA_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别巴西数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_BRA_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别巴西数据中的重复行。个人数据列中的字段必须包含单一类型的数据，如电话号码、电子邮件或个人税号。该 Mapplet 会根据个人数据生成组键。
mplt_BRA_Individual_Name_and_Address_Match	使用字段匹配策略根据人名和巴西地址数据识别重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_BRA_Individual_Name_and_CPF_Match	使用字段匹配策略根据巴西人名和个人税号 (Cadastro de Pessoas Físicas, CPF) 识别重复行。该 Mapplet 会根据 CPF 号生成组键。
mplt_BRA_Individual_Name_and_Date_Match	使用字段匹配策略根据巴西人名和日期数据识别重复行。该 Mapplet 会生成根据日期数据生成的组键。
mplt_BRA_Individual_Name_and_Email_Match	使用字段匹配策略根据巴西人名和电子邮件地址识别重复行。该 Mapplet 会根据电子邮件地址数据生成组键。
mplt_BRA_Individual_Name_and_Phone_Match	使用字段匹配策略根据巴西人名和电话号码识别重复行。该 Mapplet 会根据电话号码生成组键。
mplt_Company_Name_Match	使用字段匹配策略根据公司名称识别重复行。该 Mapplet 将根据公司名称值生成 Soundex 代码，并使用该 Soundex 代码作为组键。
rule_BRA_Company_Name_and_Address_MatchScore	根据公司名称和巴西地址数据生成匹配得分。
rule_BRA_Familyname_and_Address_MatchScore	根据姓氏和巴西地址数据生成匹配得分。

名称	说明
rule_BRA_Firstname_and_CPF_MatchScore	根据名字和个人税号 (Cadastro de Pessoas Físicas, CPF) 生成匹配得分。
rule_BRA_Individual_Name_and_Address_MatchScore	根据人名和巴西地址数据生成匹配得分。
rule_BRA_Individual_Name_and_CPF_MatchScore	根据人名和巴西地址数据生成匹配得分。
rule_BRA_Individual_Name_and_Phone_MatchScore	根据人名和电话号码生成匹配得分。
rule_Company_Name_MatchScore	根据公司名称生成匹配得分。
rule_Individual_Name_and_Date_MatchScore	根据人名和日期生成匹配得分。
rule_Individual_Name_and_Email_MatchScore	根据人名和电子邮件地址生成匹配得分。

巴西演示映射

巴西加速器中的演示映射使用多个规则演示数据质量过程。

您可以在以下存储库位置找到演示映射：

[Informatica_DQ_Content]\Rules_Demo\BRA_Accelerator

该加速器包含以下演示映射：

m_BRA_customer_data_demo

解析、标准化和验证巴西数据。

m_BRA_customer_matching_demo

解析巴西的标识数据并将其标准化，并对该数据执行标识匹配分析。

该映射会分析以下数据组合，并为每个组合生成匹配群集：

- 人名和地址数据
- 人名和电话号码

第 6 章

金融服务加速器

本章包括以下主题：

- [金融服务加速器概览, 63](#)
- [金融服务联系人数据清理规则, 63](#)
- [金融服务财务数据清理规则, 64](#)
- [金融服务常规数据清理规则, 66](#)
- [金融服务匹配和消除重复记录规则, 67](#)

金融服务加速器概览

使用金融服务加速器可验证和改善金融服务领域的组织的数据。

金融服务加速器包括执行以下数据质量过程的规则：

- 联系人数据清理
- 金融数据清理
- 常规数据清理
- 匹配和消除重复记录

该加速器依赖于核心加速器安装的数据清理规则。

金融服务联系人数据清理规则

使用联系人数据清理规则可将有关业务联系人和个人的数据标准化。

您可以在以下存储库位置找到联系人数据清理规则：

[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

下表介绍了金融服务加速器中的联系人数据清理规则：

名称	说明
rule_USA_Given_Name_Standard	根据美国昵称生成名字。例如，该规则会将昵称“Bob”标准化为名字“Robert”。

金融服务财务数据清理规则

使用金融数据清理规则可解析、标准化和验证金融数据。

您可以在以下存储库位置找到金融数据清理规则：

[Informatica_DQ_Content]\Rules\Financial_Data_Cleansing

下表介绍了金融服务加速器中的金融数据清理规则：

名称	说明
rule_Account_Status_Validation	验证帐户状态。该规则需要使用帐户状态引用数据。
rule_Accrual_Period_Validation	验证开始日期是否早于结束日期。
rule_Age_For_Account_Validation	验证帐户类型的客户年龄。该规则将使用 age_per_account_infa 引用表。您必须使用自己的数据更新该引用表。
rule_Beta_Coefficient_Validation	验证贝塔系数字符串是否为数字。该规则可指示字符串是正数、负数、零还是非数字。
rule_BIC_SWIFT_Code_Validation	通过模式识别和国家/地区代码验证来验证银行识别代码 (Bank Identifier Code, BIC) 或环球银行金融电信协会 (Society for Worldwide Interbank Financial Telecommunication, SWIFT) 代码。
rule_CAN_Transit_Number_Validation	使用纸质和电子基金交易验证加拿大银行代号 (Transit Number) 的格式。
rule_Credit_Card_Expiry_Check	验证信用卡到期日期。该规则会将信用卡到期日期与系统日期进行比较并识别已过期的日期。该规则接受格式为 MM/YYYY 的七个字符的字符串。
rule_Credit_Card_Security_Code_Validation	验证信用卡的安全码是否为包含三位或四位数字的整数。
rule_Currency_Code_Country_Validation	验证货币代码对于 ISO 三字符国家/地区代码来说是否为有效货币代码。
rule_Currency_Code_Validation	验证货币代码。该规则将返回“有效”或“无效”。
rule_CUSIP_Validation	验证校验位值的格式和长度。该规则会返回一个描述校验位值有效性的状态和一条解释该状态的消息。
rule_Delta_Validation	验证增量值是正值、负值还是零。
rule_Dividend_Yield_Validation	验证股息收益率字符串是否为大于或等于零的数字。该规则会返回该字符串是正数、负数、零还是非数字。
rule_EAD_Drawn_Balance_Validation	验证违约风险暴露 (EAD) 中所列的金额是否不少于已提取的余额。该规则遵循英国金融服务管理局 (Financial Services Authority) 提出的 EAD 计算准则。
rule_EAD_Validation	验证违约风险暴露 (EAD) 字符串是否为数字。该规则会返回该字符串是正数、负数、零还是非数字。

名称	说明
rule_EPS_Validation	验证输入是否为大于或等于零的数字。
rule_Ex_Dividend_Date_Validation	验证除息日期和记录日期是否为有效日期，以及除息日期是否早于记录日期。该规则会将差值大于 15 天的日期标识为无效。该规则将返回记录日期与除息日期之间相差的天数。
rule_Gamma_Validation	验证 Gamma 字符串是否为数字。该规则会返回该字符串是正数、负数、零还是非数字。
rule_GBR_Bank_Account_Parse	将八位数字字符串解析为英国银行帐号。
rule_GBR_Bank_Account_Validation	验证英国银行帐号。该规则将返回表示输入是否为数字以及位数是否正确的代码。
rule_GBR_Bank_Sort_Code_Parse	将六位数字字符串解析为英国银行清算代码 (Sort Code)。该规则可解析以下格式的数字字符串： - 连续数字 (999999) - 使用短划线分隔的数字 (99-99-99)
rule_GBR_Bank_Sort_Code_Standardize	将英国银行清算代码 (Sort Code) 标准化为 “NN-NN-NN” 格式。
rule_GBR_Bank_Sort_Code_Validation	验证已标准化为短划线分隔格式 (99-99-99) 的英国银行清算代码 (Sort Code) 的格式和长度。该规则会返回一个描述该清算代码有效性的状态端口和一个解释该状态的验证说明端口。如果清算代码前缀与某个已知的英国银行代码分配相匹配，验证说明端口将包含该银行名称。
rule_Interest_Rate_Within_Range	验证十进制利率值是否在指定范围内。该范围通过表达式转换中的两个变量端口进行设置。该规则将返回 “True” 或 “False”。
rule_ISIN_Code_Validation	验证输入值是否为国际证券识别码 (International Securities Identification Number, ISIN)。该规则校验值的结构并验证校验数位。
rule_Loan_to_Value_Ratio	计算贷款价值比 (Loan to Value)，即贷款额除以抵押物价值。
rule_Loss_Given_Default_Validation	验证字符串是否为数字以及是正值、负值还是零值。
rule_Market_Cap_Validation	验证输入是否为大于或等于零的数字。
rule_Maturity_Date_Validation	验证到期日期是否大于系统日期。
rule_Positive_Close_Price_Value_Validation	验证输入是否为大于零的数字。
rule_Positive_Coupon_Percent_Validation	验证输入是否为大于零的数字。
rule_Positive_Last_Price_Value_Validation	验证输入是否为大于零的数字。
rule_Positive_Open_Price_Validation	验证输入是否为大于零的数字。

名称	说明
rule_Positive_Volume_Validation	验证输入是否为大于零的数字。
rule_Price_Earnings_Ratio_Validation	验证市盈率是否为 0-100 之间的正数。
rule_Probability_of_Default_Validation	验证默认值的概率是否为数字，并指示其是正值、负值还是零值。如果为正值，该规则将返回以下范围内的值的状态消息： - <= .1 - > .1 并 <= .5 - > .5 并 <= 1 - > 1
rule_Rating_Code_Validation	验证一项评级是属于标准普尔 (Standard & Poor's) 评级表、穆迪 (Moody's) 评级表还是用户定义的列表。
rule_Rating_Date_Validation	验证评级日期是否超过系统日期一年。
rule_Risk_Weighted_Asset_Validation	验证风险加权资产是否为正数。
rule_SEDOL_Validation	通过检查格式和校验位来验证证券交易所每日官定牌价 (Stock Exchange Daily Official List, SEDOL) 代码。
rule_Stock_Exchange_Validation	按名称和符号验证世界上大多数证券交易所。
rule_USA_Routing_Number_Validation	验证采用标准磁性墨水字符识别行 (Magnetic Ink Character Recognition, MICR) 格式的路径号码 (Routing Number)。验证相关联邦储备银行、输入的结构以及校验和计算。
rule_Volatility_Validation	验证波动值是否为大于或等于零的数字。

金融服务常规数据清理规则

使用常规数据清理规则可识别输入字段中所含信息的类型。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]\Rules\General _Data_Cleansing

下表介绍了金融服务加速器中的常规数据清理规则：

名称	说明
rule_Positive_Number_Validation	验证数字是否大于零。

与核心常规数据清理规则的相关性

金融服务加速器依赖于核心加速器中的以下常规数据清理规则：

- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Remove_Space
- rule_UpperCase

有关这些规则的详细信息，请参阅 [“核心常规数据清理规则” 页面上 22。](#)

金融服务匹配和消除重复记录规则

使用匹配和消除重复记录规则可生成匹配得分和识别重复记录。

您可以在以下存储库位置找到匹配和消除重复记录规则：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了金融服务加速器中的匹配和消除重复记录规则：

名称	说明
mplt_Company_Name_and_Address_Match	根据公司名称和美国地址数据识别重复行。该 Mapplet 将使用公司名称值和邮政编码值的字符组合生成组键。
mplt_Company_Name_Match	根据公司名称识别重复行。该 Mapplet 将根据公司名称值生成 Soundex 代码，并使用该 Soundex 代码作为组键。
mplt_Familyname_and_Address_Match	根据姓氏和美国地址数据识别重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_Individual_Name_and_Address_Match	根据人名和美国地址数据识别重复行。Mapplet 将根据姓氏值生成 NYSIIS 代码，并使用该 NYSIIS 代码作为组键。
mplt_Individual_Name_and_Date_Match	根据人名和日期数据识别重复行。该 Mapplet 将生成根据日期数据生成的组键。
mplt_Individual_Name_and_Email_Match	根据人名和电子邮件地址识别重复行。该 Mapplet 匹配将生成根据电子邮件地址数据生成的键。
mplt_Individual_Name_and_Phone_Match	根据人名和电话号码识别重复行。该 Mapplet 会根据电话号码生成组键。
mplt_Individual_Name_Match	根据人名识别重复行。Mapplet 将根据姓氏值生成 NYSIIS 代码，并使用该 NYSIIS 代码作为组键。
rule_Company_Name_and_Address_MatchScore	根据公司名称和美国地址生成匹配得分。
rule_Company_Name_MatchScore	根据公司名称生成匹配得分。

名称	说明
rule_Familyname_and_Address_MatchScore	根据姓氏和美国地址生成匹配得分。
rule_Individual_Name_and_Address_MatchScore	根据人名和美国地址生成匹配得分。
rule_Individual_Name_and_Date_MatchScore	根据人名和日期生成匹配得分。
rule_Individual_Name_and_Email_MatchScore	根据人名和电子邮件地址生成匹配得分。
rule_Individual_Name_and_Phone_MatchScore	根据人名和电话号码生成匹配得分。
rule_Individual_Name_MatchScore	根据人名生成匹配得分。

第 7 章

法国加速器

本章包括以下主题：

- [法国加速器概览, 69](#)
- [法国地址数据清理规则, 69](#)
- [法国复合规则, 70](#)
- [法国联系人数据清理规则, 71](#)
- [法国公司数据清理规则, 73](#)
- [法国常规数据清理规则, 73](#)
- [法国匹配和消除重复记录规则, 74](#)
- [法国演示映射, 75](#)

法国加速器概览

使用法国加速器中的规则可验证和改善法国组织的数据。

法国加速器包括执行以下数据质量过程的规则：

- 地址数据清理
- 联系人数据清理
- 公司数据清理
- 常规数据清理
- 匹配和消除重复记录

该加速器依赖于核心加速器安装的数据清理规则。

法国地址数据清理规则

使用地址数据清理规则可解析、标准化和验证地址数据。

您可以在以下存储库位置找到地址数据清理规则：

[Informatica_DQ_Content]\Rules\Address_Data_Cleansing

下表介绍了法国加速器中的地址数据清理规则：

名称	说明
rule_FRA_Address_Parse_Hybrid	将非结构化的法国地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_FRA_Address_Parse_Multiline	将非结构化的法国地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_FRA_Address_Validation_Discrete_w_Geocoding	验证法国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_FRA_Address_Validation_Discrete	验证法国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_FRA_Address_Validation_Hybrid_w_Geocoding	验证法国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_FRA_Address_Validation_Hybrid	验证法国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_FRA_Address_Validation_Multiline_w_Geocoding	验证法国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_FRA_Address_Validation_Multiline	验证法国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。

法国复合规则

使用法国加速器中的复合规则，将一组规则作为单个对象添加到映射中。复合规则是一种利用其他加速器规则的逻辑的规则。

法国加速器中的复合规则安装在以下存储库位置：

[Informatica_DQ_Content]\Rules\Composite_Rules

适用于法国联系人数据的复合规则

下表介绍了法国加速器中适用于联系人数据的复合规则：

名称	说明
rule_FRA_Contact_Data	解析、标准化和验证地址和电话号码等法国联系人数据。

规则 rule_FRA_Contact_Data 从存储库中的多个文件夹读取 Mapplet。该规则还包括一个不可重用的转换。

下表列出了 rule_FRA_Contact_Data 中的规则和转换的名称与存储库位置：

规则	位置
Case_Converter	不可重用的转换
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_FRA_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_FRA_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Phone_Number_Standardize	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Prenome_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_FRA_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

法国联系人数据清理规则

使用联系人数据清理规则可解析、标准化和验证有关业务联系人和个人的数据。

您可以在以下存储库位置找到联系人数据清理规则：

[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

下表介绍了法国加速器中的联系人数据清理规则：

名称	说明
rule_FRA_Gender_Assignmen t	按照名字分配性别。该规则会为男性名字返回“M”，为女性名字返回“F”，为未知性别返回“U”。例如，该规则会为姓名“Jean Leclerc”分配代表男性的“M”性别。
rule_FRA_Given_Name_Stan dard	根据法国昵称生成名字。

名称	说明
rule_FRA_INSEE_Parse	解析字符串中的法国国家统计与经济研究所 (Institut National de la Statistique et des Études Économiques, INSEE) 编码。
rule_FRA_INSEE_Standardization	将法国 INSEE 编码标准化为一个包含 13 位数字的字符串或 13 位数字后跟一个空格和两位控制码。
rule_FRA_INSEE_Validation	根据性别、日期和 Code Officiel Géographique (COG) 值验证 INSEE 编码。
rule_FRA_Multi_Person_Name_Parse	<p>将人名值解析为单独的端口。该规则会为头衔、名字、中间名和姓氏等值创建端口。</p> <p>规则输出包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p> <p>当姓名数据标识了多位人员时，该规则会为每个全名创建一个输出端口。例如，该规则可读取人名 “Jean et Marianne Leclerc”，并为 “Jean Leclerc” 和 “Marianne Leclerc” 创建输出端口。</p>
rule_FRA_Personal_Name_Parsing_FML	<p>将人名中的值解析为单独的端口。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 名字、中间名、姓氏 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p>
rule_FRA_Personal_Name_Parsing_LFM	<p>将人名中的值解析为单独的端口。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 姓氏、名字、中间名 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p>
rule_FRA_Phone_Number_Parse	<p>解析字符串中的法国电话号码。该规则将解析数据中的第一个电话号码（从右到左读取）。</p> <p>该规则会识别使用前导零、国际拨号代码或以 # 号开头的分机号的电话号码。该规则可处理以下标点符号：加号、括号和 # 号。在运行该规则前，请删除所有其他标点，包括双空格。</p> <p>该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。</p>
rule_FRA_Phone_Number_Standardize	将法国电话号码标准化为国际和本地拨号格式。该规则会识别使用前导零、国际拨号代码或以 # 号开头的分机号的电话号码。
rule_FRA_Phone_Number_Validation	验证法国电话号码的区号和长度。该规则将返回电话号码的所属地区，以及表示电话号码的区号和长度是否有效的代码。
rule_FRA_Prenome_Assignment	根据性别生成敬称。
rule_FRA_Salutation_Assignment	根据名字和姓名标志生成正式和非正式问候。例如，当输入数据包含 “M. Jean Leclerc,” 时，该规则会生成正式问候 “Monsieur Leclerc,” 和非正式问候 “Cher Jean,”。您可以通过编辑 dq_Generate_Salutation Expression 表达式转换中的变量来更改前缀和标点。

与核心联系人数据清理规则的相关性

法国加速器依赖于核心加速器中的以下联系人数据清理规则：

- rule_Email_Validation

有关这些规则的详细信息，请参阅[“核心联系人数据清理规则”](#)页面上 21。

法国公司数据清理规则

使用公司数据清理规则可将公司数据标准化。

您可以在以下存储库位置找到公司数据清理规则：

[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

下表介绍了法国加速器中的公司数据清理规则：

名称	说明
rule_FRA_Company_Name_Standardization	将公司名称标准化为引用表值。
rule_FRA_SIRET_Number_Parse	解析字符串中的法国国家识别企业索引系统 (système d'identification du répertoire des établissements, SIRET) 编码。
rule_FRA_SIRET_Number_Standardize	将 14 位数字的编码标准化为 NNN NNN NNN NNNNN 格式，无论字符串中的空格或标点如何。 少于 14 位数字的字符串不进行标准化。
rule_FRA_SIRET_Number_Validation	验证法国国家识别企业索引系统 (système d'identification du répertoire des établissements, SIRET) 编码。该规则假设应用 rule_FRA_SIRET_Number_Standardization 规则后，编码采用的是标准格式。

法国常规数据清理规则

使用常规数据清理规则可识别输入字段中所含信息的类型。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]\Rules\General_Data_Cleansing

下表介绍了法国加速器中的常规数据清理规则：

名称	说明
rule_FRA_NER_Field_Identification	识别输入字段中包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和法国地址数据。该规则会返回一个说明输入数据类型的标签。它使用引用数据识别信息类型。该规则使用概率匹配技术识别信息类型。

法国加速器依赖于核心加速器中的以下常规数据清理规则：

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description

- rule_Assign_DQ_Match_Code_Description
- rule_Luhn_Algorithm
- rule_Remove_Extra_Spaces
- rule_Remove_Parentheses
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

有关这些规则的详细信息，请参阅 [“核心常规数据清理规则” 页面上 22。](#)

法国匹配和消除重复记录规则

使用匹配和消除重复记录功能生成匹配得分并识别重复记录。

法国加速器中的匹配和消除重复记录规则安装在以下存储库位置：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了法国加速器中的匹配和消除重复记录规则：

名称	说明
mplt_Company_Name_Match	使用字段匹配策略根据公司名称识别重复行。该 Mapplet 将根据公司名称值生成 Soundex 代码，并使用该 Soundex 代码作为组键。
mplt_FRA_Company_Name_and_Address_Match	使用字段匹配策略根据公司名称和地址识别重复行。该 Mapplet 将使用公司名称值和邮政编码值的字符组合生成组键。
mplt_FRA_Familyname_and_Address_Match	使用字段匹配策略根据姓氏和地址识别重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_FRA_Firstname_and_INSEE_Match	使用字段匹配策略根据法国国家统计与经济研究所 (Institut National de la Statistique et des Études Économiques, INSEE) 编码识别重复行。该 Mapplet 会根据 INSEE 编码数据生成组键。
mplt_FRA_Firstname_Surname_DOB_and_Postcode_Match	使用字段匹配策略识别人名、出生日期和邮政编码的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_FRA_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别法国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_FRA_IMO_Familyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别法国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_FRA_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别法国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。

名称	说明
mplt_FRA_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别法国数据中的重复行。个人数据列中的字段必须包含单一类型的数据，如电话号码、电子邮件或法国国家统计局与经济研究所 (Institut National de la Statistique et des Études Économiques, INSEE) 编码。该 Mapplet 会根据个人数据生成组键。
mplt_FRA_Individual_Name_and_Date_Match	使用字段匹配策略根据法国人名和日期数据识别重复行。该 Mapplet 会根据日期生成组键。
mplt_FRA_Individual_Name_and_Email_Match	使用字段匹配策略根据法国人名和电子邮件地址识别重复行。该 Mapplet 会根据电子邮件地址数据生成组键。
mplt_FRA_Individual_Name_and_INSEE_Match	使用字段匹配策略根据法国人名和 INSEE 编码识别重复行。该 Mapplet 会根据 INSEE 编码数据生成组键。
mplt_FRA_Individual_Name_Match	使用字段匹配策略根据法国人名标识重复行。Mapplet 将根据姓氏值生成 NYSIIS 代码，并使用该 NYSIIS 代码作为组键。
rule_Company_Name_Match Score	根据公司名称生成匹配得分。
rule_FRA_Company_Name_and_Address_MatchScore	根据公司名称和法国地址生成匹配得分。
rule_FRA_Familyname_and_Address_MatchScore	根据姓氏和法国地址生成匹配得分。
rule_FRA_Firstname_and_INSEE_MatchScore	根据名字和个人数据列中的任意数据（如电话号码、电子邮件或 INSEE 编码等）生成匹配得分。
rule_FRA_Firstname_Surname_DOB_and_Postcode_MatchScore	根据姓氏、出生日期和邮政编码生成匹配得分。
rule_FRA_Individual_Name_and_INSEE_MatchScore	根据人名和 INSEE 编码生成匹配得分。
rule_Individual_Name_and_Date_MatchScore	根据人名和日期生成匹配得分。
rule_Individual_Name_and_Email_MatchScore	根据人名和电子邮件地址生成匹配得分。
rule_Individual_Name_Match Score	根据人名生成匹配得分。

法国演示映射

法国加速器中的演示映射使用多个规则演示数据质量过程。

您可以在以下存储库位置找到演示映射：

[Informatica_DQ_Content]\Rules_Demo\FRA_Accelerator

该加速器包含以下演示映射：

m_FRA_customer_data_demo

解析、标准化和验证法国数据。

m_FRA_customer_matching_demo

解析澳大利亚和新西兰的标识数据并将其标准化，并对该数据执行标识匹配分析。

该映射会分析以下数据组合，并为每个组合生成匹配群集：

- 人名和地址数据
- 人名和电话号码

第 8 章

德国加速器

本章包括以下主题：

- [德国加速器概览, 77](#)
- [德国地址数据清理规则, 77](#)
- [德国复合规则, 78](#)
- [德国联系人数据清理规则, 79](#)
- [德国公司数据清理规则, 81](#)
- [德国常规数据清理规则, 81](#)
- [德国匹配和消除重复记录规则, 82](#)
- [德国演示映射, 84](#)

德国加速器概览

使用德国加速器中的规则可验证和改善德国组织的数据。

德国加速器包括执行以下数据质量过程的规则：

- 地址数据清理
- 联系人数据清理
- 公司数据清理
- 常规数据清理
- 匹配和消除重复记录

该加速器依赖于核心加速器安装的数据清理规则。

德国地址数据清理规则

使用地址数据清理规则可解析、标准化和验证地址数据。

您可以在以下存储库位置找到地址数据清理规则：

[Informatica_DQ_Content]\Rules\Address_Data_Cleansing

下表介绍了德国加速器中的地址数据清理规则：

名称	说明
rule_DEU_Address_Parse_Hybrid	将非结构化的德国地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_DEU_Address_Parse_Multiline	将非结构化的德国地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_DEU_Address_Validation_Discrete_w_Geocoding	验证德国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_DEU_Address_Validation_Discrete	验证德国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_DEU_Address_Validation_Hybrid_w_Geocoding	验证德国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_DEU_Address_Validation_Hybrid	验证德国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_DEU_Address_Validation_Multiline_w_Geocoding	验证德国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_DEU_Address_Validation_Multiline	验证德国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。

德国复合规则

使用德国加速器中的复合规则，将一组规则作为单个对象添加到映射中。复合规则是一种利用其他加速器规则的逻辑的规则。

您可以在以下存储库位置找到复合规则：

[Informatica_DQ_Content]\Rules\Composite_Rules

适用于德国联系人数据的复合规则

下表介绍了德国加速器中适用于联系人数据的复合规则：

名称	说明
rule_DEU_Contact_Data	解析、标准化和验证地址和电话号码等德国联系人数据。

规则 rule_DEU_Contact_Data 从存储库中的多个文件夹读取 Mapplet。该规则还包括一个不可重用的转换。

下表列出了 rule_DEU_Contact_Data 中的规则和转换的名称与存储库位置：

规则	位置
Case_Converter	不可重用的转换
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_DEU_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_DEU_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_DEU_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_DEU_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

德国联系人数据清理规则

使用联系人数据清理规则可解析、标准化和验证有关业务联系人和个人的数据。

您可以在以下存储库位置找到联系人数据清理规则：

[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

下表介绍了德国加速器中的联系人数据清理规则：

名称	说明
rule_DEU_Gender_Assignme nt	按照名字分配性别。该规则会为男性名字返回“M”，为女性名字返回“F”，为未知性别返回“U”。例如，该规则会为姓名“Hans Adler”分配代表男性的“M”性别。
rule_DEU_Given_Name_Stan dard	根据德国昵称生成名字。

名称	说明
rule_DEU_Multi_Person_Name_Parse	<p>将人名值解析为单独的端口。该规则会为头衔、名字、中间名和姓氏等值创建端口。</p> <p>规则输出包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p> <p>当姓名数据标识了多位人员时，该规则会为每个全名创建一个输出端口。例如，该规则可读取人名“Hans und Maria Adler”，并为“Hans Adler”和“Maria Adler”创建输出端口。</p>
rule_DEU_Personal_Name_Parsing_FML	<p>将人名中的值解析为单独的端口。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 名字、中间名、姓氏 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p>
rule_DEU_Personal_Name_Parsing_LFM	<p>将人名中的值解析为单独的端口。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 姓氏、名字、中间名 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p>
rule_DEU_Phone_Number_Parse	<p>解析字符串中的德国电话号码。该规则将解析数据中的第一个电话号码（从右到左读取）。</p> <p>该规则会识别使用前导零、国际拨号代码或以 # 号开头的分机号的电话号码。该规则可处理以下标点符号：加号、括号和 # 号。在运行该规则前，请删除所有其他标点，包括双空格。</p> <p>该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。</p>
rule_DEU_Phone_Number_Standardization	<p>将德国电话号码标准化为国际和本地拨号格式。该规则会识别使用前导零、国际拨号代码或以 # 号开头的分机号的电话号码。</p>
rule_DEU_Phone_Number_Validation	<p>验证德国电话号码的区号和长度。该规则将返回电话号码的所属地区，以及表示电话号码的区号和长度是否有效的代码。</p>
rule_DEU_Prenam_Assignment	<p>根据性别生成敬称。</p>
rule_DEU_Salutation_Assignment	<p>根据名字和姓名标志生成正式和非正式问候。例如，当输入数据包含“Herr Hans Adler,”时，该规则会生成正式问候“Sehr geehrter Herr Adler,”和非正式问候“Lieber Hans,”。您可以通过编辑 dq_Generate_Salutation Expression 表达式转换中的变量来更改前缀和标点。</p>

与核心联系人数据清理规则的相关性

德国加速器依赖于核心加速器中的以下联系人数据清理规则：

- rule_Email_Validation

有关这些规则的详细信息，请参阅 [“核心联系人数据清理规则” 页面上 21。](#)

德国公司数据清理规则

使用公司数据清理规则可将公司数据标准化。

您可以在以下存储库位置找到公司数据清理规则：

[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

下表介绍了德国加速器中的公司数据清理规则：

名称	说明
rule_DEU_Company_Name_Standardization	使用引用表将公司名称标准化。

德国常规数据清理规则

使用常规数据清理规则可识别输入字段中所含信息的类型。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]\Rules\General_Data_Cleansing

下表介绍了德国加速器中的常规数据清理规则：

名称	说明
rule_DEU_NER_Field_Identification	识别输入字段中包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和德国地址数据。该规则会返回一个说明输入数据类型的标签。它使用引用数据识别信息类型。该规则使用概率匹配技术识别信息类型。

与核心常规数据清理规则的相关性

德国加速器依赖于核心加速器中的以下常规数据清理规则：

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Remove_Extra_Spaces
- rule_Remove_Hyphen
- rule_Remove_Leading_Zero
- rule_Remove_Parentheses
- rule_Remove_Period_Parentheses
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Remove_Space
- rule_Replace_Limited_Punct_with_Space

- rule_UpperCase

有关这些规则的详细信息，请参阅 [“核心常规数据清理规则” 页面上 22。](#)

德国匹配和消除重复记录规则

使用匹配和消除重复记录规则可生成匹配得分和识别重复记录。

您可以在以下存储库位置找到匹配和消除重复记录规则：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了德国加速器中的匹配和消除重复记录规则：

名称	说明
mplt_Company_Name_Match	使用字段匹配策略根据公司名称识别重复行。该 Mapplet 将根据公司名称值生成 Soundex 代码，并使用该 Soundex 代码作为组键。
mplt_DEU_Company_Name_and_Address_Match	使用字段匹配策略根据公司名称和地址数据识别德国数据中的重复行。该 Mapplet 将使用公司名称值和邮政编码值的字符组合生成组键。
mplt_DEU_Familyname_and_Address_Match	使用字段匹配策略根据姓氏和地址数据识别德国数据中的重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_DEU_Firstname_3Chars_Surname_DOB_and_Postcode_Match	使用字段匹配策略根据名字、姓氏的前三个字符、出生日期和邮政编码识别德国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_DEU_Firstname_and_PID_Match	使用字段匹配策略根据分组的名字和个人 ID 识别德国数据中的重复行。该 Mapplet 会根据个人 ID 数据生成组键。
mplt_DEU_Firstname_Surname_2ElementsDOB_and_Postcode_Match	使用字段匹配策略根据人名、两个出生日期元素和邮政编码识别德国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_DEU_Firstname_Surname_DOB_and_Postcode_Match	使用字段匹配策略根据人名、出生日期和邮政编码识别德国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_DEU_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别德国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_DEU_IMO_Familyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别德国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_DEU_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别德国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_DEU_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别德国数据中的重复行。个人数据列中的字段必须包含单一类型的数据，如电话号码、电子邮件或个人 ID。该 Mapplet 会根据个人数据生成组键。
mplt_DEU_Individual_Name_and_Date_Match	使用字段匹配策略根据人名和按日期分组的日期数据识别重复行。该 Mapplet 会生成根据日期数据生成的组键。

名称	说明
mplt_DEU_Individual_Name_and_Email_Match	使用字段匹配策略根据人名和电子邮件地址识别德国数据中的重复行。该 Mapplet 会根据电子邮件地址数据生成组键。
mplt_DEU_Individual_Name_and_Phone_Match	使用字段匹配策略根据人名和电话号码识别德国数据中的重复行。该 Mapplet 会根据电话号码数据生成组键。
mplt_DEU_Individual_Name_and_PID_Match	使用字段匹配策略根据人名和个人 ID 识别德国数据中的重复行。该 Mapplet 会根据个人 ID 数据生成组键。
mplt_DEU_Individual_Name_Match	使用字段匹配策略根据人名识别德国数据中的重复行。该 Mapplet 将根据姓氏值生成 NYSIIS 代码，并使用该 NYSIIS 代码作为组键。
rule_Company_Name_MatchScore	根据公司名称生成匹配得分。
rule_DEU_Company_Name_and_Address_MatchScore	根据公司名称和地址生成匹配得分。
rule_DEU_Familyname_and_Address_MatchScore	根据姓氏和地址生成匹配得分。
rule_DEU_Firstname_3Chars_Surname_DOB_and_Postcode_MatchScore	根据名字、姓氏的前三个字符、出生日期和邮政编码生成匹配得分。
rule_DEU_Firstname_and_PID_MatchScore	根据名字和个人数据列中的任意数据（如电话号码、电子邮件或个人 ID 等）生成匹配得分。
rule_DEU_Firstname_Surname_2ElementsDOB_and_Postcode_MatchScore	根据名字、出生日期和邮政编码生成匹配得分。 注意: 出生日期的输入格式假定为 DD/MM/YYYY。
rule_DEU_Firstname_Surname_DOB_and_Postcode_MatchScore	根据姓氏、出生日期和邮政编码生成匹配得分。
rule_DEU_Individual_Name_and_Phone_MatchScore	根据人名和电话号码生成匹配得分。
rule_Familyname_and_Address_MatchScore	根据姓氏和地址生成匹配得分。
rule_Individual_Name_and_Date_MatchScore	根据人名和日期生成匹配得分。
rule_Individual_Name_and_Email_MatchScore	根据人名和电子邮件地址生成匹配得分。
rule_Individual_Name_and_SSN_MatchScore	根据名字和个人数据列中的任意数据（如电话号码、电子邮件或 SSN 编号等）生成匹配得分。
rule_Individual_Name_MatchScore	根据人名生成匹配得分。

德国演示映射

德国加速器中的演示映射使用多个规则演示质量过程。

您可以在以下存储库位置找到演示映射：

[Informatica_DQ_Content]\Rules_Demo\DEU_Accelerator

该加速器包含以下演示映射：

m_DEU_customer_data_demo

解析、标准化和验证德国数据。

m_DEU_customer_matching_demo

解析德国的标识数据并将其标准化，并对该数据执行标识匹配分析。

该映射会分析以下数据组合，并为每个组合生成匹配群集：

- 人名和地址数据
- 人名和电话号码

第 9 章

葡萄牙加速器

本章包括以下主题：

- [葡萄牙加速器概览, 85](#)
- [葡萄牙地址数据清理规则, 85](#)
- [葡萄牙复合规则, 86](#)
- [葡萄牙联系人数据清理规则, 87](#)
- [葡萄牙公司数据清理规则, 89](#)
- [葡萄牙常规数据清理规则, 89](#)
- [葡萄牙匹配和消除重复记录规则, 90](#)
- [葡萄牙演示映射, 91](#)

葡萄牙加速器概览

使用葡萄牙加速器中的规则可验证和改善葡萄牙组织的数据。

葡萄牙加速器包括执行以下数据质量过程的规则：

- 地址数据清理
- 联系人数据清理
- 公司数据清理
- 常规数据清理
- 匹配和消除重复记录

该加速器依赖于核心加速器安装的数据清理规则。

葡萄牙地址数据清理规则

使用地址数据清理规则可解析和验证地址数据。

您可以在以下存储库位置找到地址数据清理规则：

[Informatica_DQ_Content]\Rules\Address_Data_Cleansing

下表介绍了葡萄牙加速器中的地址数据清理规则：

名称	说明
rule_PRT_Address_Parse_Hybrid	将非结构化的葡萄牙地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_PRT_Address_Parse_Multiline	将非结构化的葡萄牙地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_PRT_Address_Validation_Discrete_w_Geocoding	验证葡萄牙地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_PRT_Address_Validation_Discrete	验证葡萄牙地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_PRT_Address_Validation_Hybrid_w_Geocoding	验证葡萄牙地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_PRT_Address_Validation_Hybrid	验证葡萄牙地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_PRT_Address_Validation_Multiline_w_Geocoding	验证葡萄牙地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_PRT_Address_Validation_Multiline	验证葡萄牙地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。

葡萄牙复合规则

使用葡萄牙加速器中的复合规则，将一组规则作为单个对象添加到映射中。复合规则是一种利用其他加速器规则的逻辑的规则。

您可以在以下存储库位置找到复合规则：

[Informatica_DQ_Content]\Rules\Composite_Rules

适用于葡萄牙联系人数据的复合规则

下表介绍了葡萄牙加速器中适用于葡萄牙联系人数据的复合规则：

名称	说明
rule_PRT_Contact_Data	解析、标准化和验证地址、电话号码和纳税人税号 (Número de Identificação Fiscal, NIF) 等葡萄牙联系人数据。

规则 rule_PRT_Contact_Data 从存储库中的多个文件夹读取 Mapplet。该规则还包括一个不可重用的转换。

下表列出了 rule_PRT_Contact_Data 中的规则和转换的名称与存储库位置：

规则	位置
Case_Converter	不可重用的转换
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\ \General_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\ \Contact_Data_Cleansing
rule_PRT_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\ \Address_Data_Cleansing
rule_PRT_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\ \Corporate_Data_Cleansing
rule_PRT_NIF_Standardization	[Informatica_DQ_Content]\Rules\ \Contact_Data_Cleansing
rule_PRT_NIF_Validate	[Informatica_DQ_Content]\Rules\ \Contact_Data_Cleansing
rule_PRT_Personal_Name_Parse_Validate	[Informatica_DQ_Content]\Rules\ \Contact_Data_Cleansing
rule_PRT_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\ \Contact_Data_Cleansing
rule_PRT_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\ \Contact_Data_Cleansing
rule_PRT_Prename_Assignment	[Informatica_DQ_Content]\Rules\ \Contact_Data_Cleansing
rule_PRT_Salutation_Assignment	[Informatica_DQ_Content]\Rules\ \Contact_Data_Cleansing

葡萄牙联系人数据清理规则

使用联系人数据清理规则可解析、标准化和验证有关业务联系人和个人的数据。

您可以在以下存储库位置找到联系人数据清理规则：

[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

下表介绍了葡萄牙加速器中的联系人数据清理规则：

名称	说明
rule_PRT_Gender_Assignment	按照名字分配性别。该规则会为男性名字返回“M”，为女性名字返回“F”，为未知性别返回“U”。例如，该规则会为姓名“Artur Cruz”分配代表男性的“M”性别。
rule_PRT_Given_Name_Standard	根据葡萄牙昵称生成名字。
rule_PRT_NIF_Parse	解析字符串中的纳税人税号 (Número de Identificação Fiscal, NIF)。该规则将返回一个 ID 号，以及一个包含已删除该 ID 号的输入文本的字符串。
rule_PRT_NIF_Standardization	将纳税人税号 (Número de Identificação Fiscal, NIF) 标准化为一个九位数字字符串。该规则将删除字母字符、符号和空格。
rule_PRT_NIF_Validate	根据每个税号中的校验位验证纳税人税号 (Número de Identificação Fiscal, NIF)。该规则要求输入内容为一个无空格的九位数字字符串。
rule_PRT_Personal_Name_Parse_Validate	将人名值解析为单独的端口。该规则会为头衔、名字、中间名和姓氏等值创建端口。规则还指示名称是否可能为公司名称，并验证名称拼写。 规则输出包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。
rule_PRT_Phone_Number_Parse	解析字符串中的葡萄牙电话号码。该规则将解析数据中的第一个电话号码（从右到左读取）。该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。
rule_PRT_Phone_Number_Standardization	将葡萄牙电话号码标准化为国际和本地拨号格式。
rule_PRT_Phone_Number_Validation	验证葡萄牙电话号码的区号和长度。该规则将返回电话号码的所属地区，以及表示电话号码的区号和长度是否有效的代码。
rule_PRT_Prenome_Assignment	根据性别生成敬称。您可以将 female_prenome 表达式变量从“Sra”更改为“Sta”。
rule_PRT_Salutation_Assignment	根据名字和姓名标志生成正式和非正式问候。例如，当输入数据包含“Sr. Artur Cruz,”时，该规则会生成正式问候“Prezado Sr. Cruz,”和非正式问候“Prezado Artur,”。您可以通过编辑 dq_Generate_Salutation Expression 表达式转换中的变量来更改前缀和标点。

与核心联系人数据清理规则的相关性

葡萄牙加速器依赖于核心加速器中的以下联系人数据清理规则：

- rule_Email_Validation

有关这些规则的详细信息，请参阅[“核心联系人数据清理规则”页面上 21](#)。

葡萄牙公司数据清理规则

使用公司数据清理规则可解析、标准化和验证公司数据。

您可以在以下存储库位置找到公司数据清理规则：

[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

下表介绍了葡萄牙加速器中的公司数据清理规则：

名称	说明
rule_PRT_Company_Name_Standardization	将葡萄牙公司名称标准化为引用表值。
rule_PRT_NIPC_Parse	解析葡萄牙企业税号 (Número de Identificação Pessoa Colectiva, NIPC)。该规则将返回 NIPC，以及一个包含已删除 NIPC 的输入文本的字符串。
rule_PRT_NIPC_Standardize	将企业税号 (Número de Identificação Pessoa Colectiva, NIPC) 标准化为一个九位数字字符串。该规则将删除字母字符、符号和空格。
rule_PRT_NIPC_Validate	根据每个税号中的校验位验证企业税号 (Número de Identificação Pessoa Colectiva, NIPC)。该规则要求输入内容为九位数字字符串。

葡萄牙常规数据清理规则

使用常规数据清理规则可识别输入字段中所含信息的类型。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]\Rules\General_Data_Cleansing

下表介绍了葡萄牙加速器中的常规数据清理规则：

名称	说明
rule_PRT_NER_Field_Identification	识别输入字段中包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和葡萄牙地址数据。该规则会返回一个说明输入数据类型的标签。它使用引用数据识别信息类型。该规则使用概率匹配技术识别信息类型。

与核心常规数据清理规则的相关性

葡萄牙加速器依赖于核心加速器中的以下常规数据清理规则：

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_ElementResultStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Parse_First_Word
- rule_Remove_Extra_Spaces
- rule_Remove_Non_Numbers

- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

有关这些规则的详细信息，请参阅[“核心常规数据清理规则” 页面上 22。](#)

葡萄牙匹配和消除重复记录规则

使用匹配和消除重复记录规则可生成匹配得分和识别重复记录。

您可以在以下存储库位置找到匹配和消除重复记录规则：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了葡萄牙加速器中的匹配和消除重复记录规则：

名称	说明
mplt_Company_Name_Match	使用字段匹配策略根据公司名称识别重复行。该 Mapplet 将根据公司名称值生成 Soundex 代码，并使用该 Soundex 代码作为组键。
mplt_PRT_Company_Name_and_Address_Match	使用字段匹配策略根据公司名称和地址数据识别葡萄牙数据中的重复行。该 Mapplet 将使用公司名称值和邮政编码值的字符组合生成组键。
mplt_PRT_Famillyname_and_Address_Match	使用字段匹配策略根据姓氏和地址数据识别葡萄牙数据中的重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_PRT_Firstname_and_NIF_BI_Match	使用字段匹配策略根据名字以及纳税人税号 (Número de Indentificação Fiscal, NIF) 和身份证 (Bilhete de Identidade, BI) 等个人身份号码识别葡萄牙数据中的重复行。该 Mapplet 会根据个人身份号码数据生成组键。
mplt_PRT_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别葡萄牙数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_PRT_IMO_Famillyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别葡萄牙数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_PRT_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别葡萄牙数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_PRT_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别葡萄牙数据中的重复行。个人数据列中的字段必须包含单一类型的数据，如电话号码、电子邮件或纳税人税号 (Número de Indentificação Fiscal, NIF)。该 Mapplet 会根据个人数据生成组键。
mplt_PRT_Individual_Name_and_Address_Match	使用字段匹配策略根据人名和地址数据识别葡萄牙数据中的重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_PRT_Individual_Name_and_Date_Match	使用字段匹配策略根据人名和日期数据识别葡萄牙数据中的重复行。该 Mapplet 会生成根据日期数据生成的组键。
mplt_PRT_Individual_Name_and_Email_Match	使用字段匹配策略根据人名和电子邮件地址识别葡萄牙数据中的重复行。该 Mapplet 会根据电子邮件地址数据生成组键。

名称	说明
mplt_PRT_Individual_Name_and_Phone_Match	使用字段匹配策略根据人名和电话号码识别葡萄牙数据中的重复行。该 Mapplet 会根据电话号码数据生成组键。
mplt_PRT_Individual_Name_Match	使用字段匹配策略根据人名识别葡萄牙数据中的重复行。Mapplet 将根据姓氏值生成 NYSIIS 代码，并使用该 NYSIIS 代码作为组键。
rule_Company_Name_MatchScore	根据公司名称生成匹配得分。
rule_PRT_Company_Name_and_Address_MatchScore	根据公司名称和葡萄牙地址数据生成匹配得分。
rule_PRT_Familyname_and_Address_MatchScore	根据姓氏和葡萄牙地址数据生成匹配得分。
rule_PRT_Firstname_and_NIF_BI_MatchScore	根据名字数据、纳税人税号 (Número de Identificação Fiscal, NIF) 和身份证 (Bilhete de Identidade, BI) 号码生成匹配得分。
rule_PRT_Individual_Name_and_Address_MatchScore	根据人名和葡萄牙地址数据生成匹配得分。
rule_PRT_Individual_Name_and_Date_MatchScore	根据人名和日期生成匹配得分。
rule_PRT_Individual_Name_and_Email_MatchScore	根据人名和电子邮件地址生成匹配得分。
rule_PRT_Individual_Name_and_Phone_MatchScore	根据人名和电话号码生成匹配得分。
rule_PRT_Individual_Name_MatchScore	根据人名生成匹配得分。

葡萄牙演示映射

葡萄牙加速器中的演示映射使用多个规则演示数据质量过程。

您可以在以下存储库位置找到演示映射：

[Informatica_DQ_Content]\Rules_Demo\PRT_Accelerator

该加速器包含以下演示映射：

m_PRT_customer_data_demo

解析、标准化和验证葡萄牙数据。

m_PRT_customer_matching_demo

解析葡萄牙的标识数据并将其标准化，并对该数据执行标识匹配分析。

该映射会分析以下数据组合，并为每个组合生成匹配群集：

- 人名和地址数据

- 人名和电话号码

第 10 章

西班牙加速器

本章包括以下主题：

- [西班牙加速器概览, 93](#)
- [西班牙地址数据清理规则, 93](#)
- [西班牙联系人数据清理规则, 94](#)
- [西班牙公司数据清理规则, 96](#)
- [西班牙常规数据清理规则, 96](#)
- [西班牙匹配和消除重复记录规则, 97](#)
- [西班牙演示映射, 99](#)

西班牙加速器概览

使用西班牙加速器中的规则可验证和改善西班牙组织的数据。

西班牙加速器包括执行以下数据质量过程的规则：

- 地址数据清理
- 联系人数据清理
- 公司数据清理
- 常规数据清理
- 匹配和消除重复记录

该加速器依赖于核心加速器安装的数据清理规则。

西班牙地址数据清理规则

使用地址数据清理规则可解析、标准化和验证地址数据。

您可以在以下存储库位置找到地址数据清理规则：

[Informatica_DQ_Content]\Rules\Address_Data_Cleansing

下表介绍了西班牙加速器中的地址数据清理规则：

名称	说明
rule_ESP_Address_Parse_Hybrid	将非结构化的西班牙地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_ESP_Address_Parse_Multiline	将非结构化的西班牙地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_ESP_Address_Validation_Discrete_w_Geocoding	验证西班牙地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_ESP_Address_Validation_Discrete	验证西班牙地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_ESP_Address_Validation_Hybrid_w_Geocoding	验证西班牙地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_ESP_Address_Validation_Hybrid	验证西班牙地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_ESP_Address_Validation_Multiline_w_Geocoding	验证西班牙地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_ESP_Address_Validation_Multiline	验证西班牙地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。

西班牙联系人数据清理规则

使用联系人数据清理规则可解析、标准化和验证有关业务联系人和个人的数据。

您可以在以下存储库位置找到联系人数据清理规则：

[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

下表介绍了西班牙加速器中的联系人数据清理规则：

名称	说明
rule_ESP_CIF_Parse	解析西班牙企业税号 (Certificado de Identificación Fiscal, CIF)。
rule_ESP_CIF_Standardization	将西班牙企业税号 (Certificado de Identificación Fiscal, CIF) 标准化。
rule_ESP_CIF_Validation	验证西班牙企业税号 (Certificado de Identificación Fiscal, CIF)。
rule_ESP_DNI_Parse	解析西班牙身份证 (Documento Nacional de Identidad, DNI)。
rule_ESP_DNI_Standardization	将西班牙身份证 (Documento Nacional de Identidad, DNI) 标准化。
rule_ESP_DNI_Validation	验证西班牙身份证 (Documento Nacional de Identidad, DNI)。
rule_ESP_Gender_Assignment	按照名字分配性别。该规则会为男性名字返回“M”，为女性名字返回“F”，为未知性别返回“U”。例如，该规则会为姓名“Juan Garcia”分配代表男性的“M”性别。
rule_ESP_Given_Name_Standard	根据西班牙昵称生成名字。
rule_ESP_NIE_Parse	解析西班牙的外国人身份号码 (Número de Identidad de Extranjero, NIE)。
rule_ESP_NIE_Standardization	将西班牙的外国人身份号码 (Número de Identidad de Extranjero, NIE) 标准化。
rule_ESP_NIE_Validation	验证西班牙的外国人身份号码 (Número de Identidad de Extranjero, NIE)。
rule_ESP_NIF_Parse	解析字符串中的西班牙纳税人税号 (Número de Identificación Fiscal, NIF)。
rule_ESP_NIF_Standardization	将西班牙纳税人税号 (Número de Identificación Fiscal, NIF) 标准化。
rule_ESP_NIF_Validation	验证西班牙纳税人税号 (Número de Identificación Fiscal, NIF)。
rule_ESP_Personal_Name_Parse	解析西班牙人名。
rule_ESP_Phone_Number_Parse	解析字符串中的西班牙电话号码。该规则将解析数据中的第一个电话号码（从右到左读取）。该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。
rule_ESP_Phone_Number_Standardization	将西班牙电话号码标准化为国际和本地拨号格式。该规则会识别使用前导零、国际拨号代码或以#号开头的分机号的电话号码。

名称	说明
rule_ESP_Phone_Number_Validation	验证西班牙电话号码的区号和长度。该规则将返回电话号码的所属地区，以及表示电话号码的区号和长度是否有效的代码。
rule_ESP_Phone_Parse_Standardize_Validate	解析西班牙电话号码并将格式标准化。验证西班牙电话号码的区号和长度。
rule_ESP_Prename_Assignment	根据性别生成敬称。
rule_ESP_Salutation_Assignment	根据名字和姓名标志生成正式和非正式问候。例如，当输入数据包含“Sr. Juan Garcia,”时，该规则会生成正式问候“Estimado Sr. Garcia,”和非正式问候“Querido Juan,”。您可以通过编辑 dq_Generate_Salutation Expression 表达式转换中的变量来更改前缀和标点。

与核心联系人数据清理规则的相关性

西班牙加速器依赖于核心加速器中的以下联系人数据清理规则：

- rule_Email_Validation

有关这些规则的详细信息，请参阅[“核心联系人数据清理规则”页面上 21](#)。

西班牙公司数据清理规则

使用公司数据清理规则可将公司数据标准化。

西班牙加速器依赖于核心加速器中的以下公司数据清理规则：

- rule_Company_Name_Standardization

西班牙常规数据清理规则

使用常规数据清理规则可识别输入字段中所含信息的类型。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]Rules\General_Data_Cleansing

下表介绍了西班牙加速器中的常规数据清理规则：

名称	说明
rule_ESP_NER_Field_Identification	识别输入字段中包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和西班牙地址数据。该规则会返回一个说明输入数据类型的标签。该规则使用概率匹配技术识别信息类型。

西班牙加速器依赖于核心加速器中的以下常规数据清理规则：

- rule_Assign_DQ_ElementResultStatus_Description
- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Remove_Extra_Spaces
- rule_Remove_Leading_Zero
- rule_Remove_Limited_Punctuation
- rule_Remove_Non_Numbers
- rule_Remove_Punctuation_and_Space
- rule_Remove_Punctuation
- rule_Replace_limited_Punct_with_Space
- rule_Translate_Diacritic_Characters
- rule_UpperCase

西班牙匹配和消除重复记录规则

使用匹配和消除重复记录规则可生成匹配得分和识别重复记录。

您可以在以下存储库位置找到匹配和消除重复记录规则：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了西班牙加速器中的匹配和消除重复记录规则：

名称	说明
mplt_Company_Name_Match	使用字段匹配策略根据公司名称识别重复行。
mplt_ESP_Company_Name_and_Address_Match	使用字段匹配策略根据公司名称和地址数据识别西班牙数据中的重复行。该 Mapplet 将使用公司名称值和邮政编码值的字符组合生成组键。
mplt_ESP_Familyname_and_Address_Match	使用字段匹配策略根据姓氏和地址数据识别西班牙数据中的重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_ESP_Firstname_and_NIF_BI_Match	使用字段匹配策略根据名字和纳税人识别号 (Número de Identificación Fiscal, NIF) 等个人身份号码识别西班牙数据中的重复行。该 Mapplet 会根据个人身份号码数据生成组键。
mplt_ESP_IMO_Company_Name_Match	使用标识匹配策略根据公司名称识别西班牙数据中的重复行。该 Mapplet 将根据公司名称值生成 Soundex 代码，并使用该 Soundex 代码作为组键。
mplt_ESP_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别西班牙数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。

名称	说明
mplt_ESP_IMO_Familyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别西班牙数据中的重复行。该 Mapplet 会生成根据邮政编码数据生成的组键。
mplt_ESP_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别西班牙数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_ESP_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别西班牙数据中的重复行。个人数据列中的字段必须包含单一类型的数据，如电话号码或电子邮件。该 Mapplet 会根据个人数据生成组键。
mplt_ESP_Individual_Name_Match	使用字段匹配策略根据人名识别西班牙数据中的重复行。Mapplet 将根据姓氏值生成 NYSIIS 代码，并使用该 NYSIIS 代码作为组键。
mplt_ESP_Individual_Name_and_Address_Match	使用字段匹配策略根据人名和地址数据识别西班牙数据中的重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_ESP_Individual_Name_and_Date_Match	使用字段匹配策略根据人名和日期识别西班牙数据中的重复行。该 Mapplet 会生成根据日期数据生成的组键。
mplt_ESP_Individual_Name_and_Email_Match	使用字段匹配策略根据电子邮件地址和人名识别西班牙数据中的重复行。该 Mapplet 会根据电子邮件地址数据生成组键。
mplt_ESP_Individual_Name_and_Phone_Match	使用字段匹配策略根据人名和电话号码识别西班牙数据中的重复行。该 Mapplet 会根据电话号码数据生成组键。
rule_Company_Name_MatchScore	根据公司名称生成匹配得分。
rule_ESP_Company_Name_and_Address_MatchScore	根据公司名称和西班牙地址数据生成匹配得分。
rule_ESP_Familyname_and_Address_MatchScore	根据姓氏和西班牙地址数据生成匹配得分。
rule_ESP_Firstname_and_NIF_BI_Matchscore	根据名字和纳税人税号 (Número de Identificación Fiscal, NIF) 生成匹配得分。
rule_ESP_Individual_Name_MatchScore	根据人名生成匹配得分。
rule_ESP_Individual_Name_and_Address_MatchScore	根据人名和西班牙地址数据生成匹配得分。
rule_ESP_Individual_Name_and_Date_MatchScore	根据人名和日期生成匹配得分。
rule_ESP_Individual_Name_and_Email_MatchScore	根据人名和电子邮件地址生成匹配得分。
rule_ESP_Individual_Name_and_Phone_MatchScore	根据人名和电话号码生成匹配得分。

西班牙演示映射

西班牙加速器中的演示映射使用多个规则演示数据质量过程。

您可以在以下存储库位置找到演示映射：

[Informatica_DQ_Content]\Rules_Demo\ESP_Accelerator

该加速器包含以下演示映射：

m_ESP_customer_data_demo

解析、标准化和验证西班牙数据。

m_ESP_customer_matching_demo

解析西班牙的标识数据并将其标准化，并对该数据执行标识匹配分析。

该映射会分析以下数据组合，并为每个组合生成匹配群集：

- 人名和地址数据
- 人名和电话号码

第 11 章

英国加速器

本章包括以下主题：

- [英国加速器概览, 100](#)
- [英国地址数据清理规则, 100](#)
- [英国复合规则, 102](#)
- [英国联系人数据清理规则, 103](#)
- [英国公司数据清理规则, 105](#)
- [英国金融数据清理规则, 105](#)
- [英国常规数据清理规则, 106](#)
- [英国匹配和消除重复记录规则, 107](#)
- [英国演示映射, 109](#)

英国加速器概览

使用英国加速器中的规则可验证和改善英国组织的数据。

英国加速器包括执行以下数据质量过程的规则：

- 地址数据清理
- 联系人数据清理
- 金融数据清理
- 匹配和消除重复记录

该加速器依赖于核心加速器安装的数据清理规则。

英国地址数据清理规则

使用地址数据清理规则可解析、标准化和验证地址数据。

您可以在以下存储库位置找到地址数据清理规则：

[Informatica_DQ_Content]\Rules\Address_Data_Cleansing

下表介绍了英国加速器中的地址数据清理规则：

名称	说明
rule_GBR_Address_Parse_Hybrid	将非结构化的英国地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_GBR_Address_Parse_Multiline	将非结构化的英国地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_GBR_Address_Validation_Discrete_w_Geocoding	验证英国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_GBR_Address_Validation_Discrete	验证英国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_GBR_Address_Validation_Hybrid_w_Geocoding	验证英国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_GBR_Address_Validation_Hybrid	验证英国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_GBR_Address_Validation_Multiline_w_Geocoding	验证英国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_GBR_Address_Validation_Multiline	验证英国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_GBR_Postcode_Parse	解析英国邮政编码。
rule_GBR_Postcode_Standardize	<p>将英国邮政编码标准化。该规则要求输入内容遵循预定义的格式。该规则会将符合以下模式的输入内容标准化：</p> <ul style="list-style-type: none"> - A9 9AA - A99 9AA - AA9 9AA - AA99 9AA - A9A 9AA - AA9A 9AA - GIR 0AA <p>字母 A 代表字母字符，数字 9 代表数字。</p>
rule_GBR_Postcode_Validate	验证英国 邮政编码。该规则会使用有效的英国邮政编码将邮政编码标准化。如果该规则未发现匹配的邮政编码，会验证邮政编码是否符合标准的英国 邮政编码模式。

英国复合规则

使用英国加速器中的复合规则，将一组规则作为单个对象添加到映射中。复合规则是一种利用其他加速器规则的逻辑的规则。

您可以在以下存储库位置找到复合规则：

[Informatica_DQ_Content]\Rules\Composite_Rules

适用于联系人数据的英国复合规则

下表介绍了英国加速器中适用于联系人数据的复合规则：

名称	说明
rule_GBR_Contact_Data	解析、标准化和验证地址、电话号码和国家社会保险号 (National Insurance Number, NINO) 等英国联系人数据。

规则 rule_GBR_Contact_Data 从存储库中的多个文件夹读取 Mapplet。该规则还包括一个不可重用的转换。

下表列出了 rule_GBR_Contact_Data 中的规则和转换的名称与存储库位置：

名称	位置
Case_Converter	不可重用的转换
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_GBR_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_GBR_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_NINO_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_NINO_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_GBR_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

名称	位置
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules\ Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules\ Contact_Data_Cleansing

英国联系人数据清理规则

使用联系人数据清理规则可解析、标准化和验证有关业务联系人和个人的数据。

您可以在以下存储库位置找到联系人数据清理规则：

[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

下表介绍了英国加速器中的联系人数据清理规则：

名称	说明
rule_GBR_Driver_Number_Parse	解析与英国驾驶员驾照编号格式匹配的字符串。
rule_GBR_Driver_Number_Validation	根据英国政府数据标准目录 (United Kingdom Government Data Standards Catalog) 的要求验证英国驾驶员驾照编号。
rule_GBR_Gender_Assignment	按照名字分配性别。该规则会为男性名字返回“M”，为女性名字返回“F”，为未知性别返回“U”。例如，该规则会为姓名“John Smith”分配代表男性的“M”性别。
rule_GBR_Given_Name_Standard	根据英国昵称生成名字。
rule_GBR_Multi_Person_Name_Parse	将人名值解析为单独的端口。该规则会为头衔、名字、中间名和姓氏等值创建端口。 规则输出包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。 当姓名数据标识了多位人员时，该规则会为每个全名创建一个输出端口。例如，该规则可读取姓名“John and Jane Smith”，并为“John Smith”和“Jane Smith”创建输出端口。
rule_GBR_NHS_Number_Parse	解析字符串中的英国国民医疗保健服务 (National Health Service, NHS) 编号。该规则将返回一个 NHS 编号，以及一个包含已删除该 NHS 编号的输入文本的字符串。
rule_GBR_NHS_Number_Standardize	将英国国民医疗保健服务 (National Health Service, NHS) 编号标准化为标准格式 (999 999 9999)。该规则要求输入内容为 10 位数字字符串。
rule_GBR_NHS_Number_Validate	根据每个编号中的校验位验证国民医疗保健服务 (National Health Service, NHS) 编号。该规则要求输入内容为 10 位数字字符串。
rule_GBR_NINO_Conformity_Check	验证英国国家社会保险号 (United Kingdom National Insurance Number, NINO) 的标准模式。该规则不验证 NINO 是否准确或处于活动状态。

名称	说明
rule_GBR_NINO_Parse	解析字符串中的英国国家社会保险号 (United Kingdom National Insurance Number, NINO)。该规则将返回 NINO，以及一个包含已删除该 NINO 的输入文本的字符串。
rule_GBR_NINO_Standardization	<p>将英国国家社会保险号 (United Kingdom National Insurance Number, NINO) 标准化为两种最常见的格式。该规则将返回以下格式，其中 C 代表字母字符，N 代表数字：</p> <ul style="list-style-type: none"> - CC NN NN NN C - CCNNNNNNC <p>该规则会将所有字母字符格式化为大写字母。该规则要求输入内容符合 NINO 的模式。</p>
rule_GBR_NINO_Validation	验证英国国家社会保险号 (United Kingdom National Insurance Number, NINO)。该规则不验证 NINO 是否处于活动状态。
rule_GBR_Passport_Number_MR_Parse	解析采用扩展格式的英国护照号。扩展格式是一种计算机可读取的护照号格式。
rule_GBR_Passport_Number_Parse	解析采用政府数据标准目录指定格式的英国护照号。该规则将解析所有九位数字字符串。
rule_GBR_Passport_Number_Validation	验证采用政府数据标准目录指定格式的英国护照号。
rule_GBR_Personal_Name_Parsing_FML	<p>将人名中的值解析为单独的端口。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 名字、中间名、姓氏 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p>
rule_GBR_Personal_Name_Parsing_LFM	<p>将人名中的值解析为单独的端口。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 姓氏、名字、中间名 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p>
rule_GBR_Phone_Number_Parse	<p>解析字符串中的英国电话号码。该规则将解析数据中的第一个电话号码（从右到左读取）。</p> <p>该规则会识别使用前导零、“+44”国际拨号代码和以 # 号开头的分机号的电话号码。该规则可处理以下标点符号：加号、括号和 # 号。在运行该规则前，请删除所有其他标点，包括双空格。</p> <p>该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。</p>
rule_GBR_Phone_Number_Standardization	将英国电话号码标准化为国际和本地拨号格式。该规则会识别使用前导零、“+44”国际拨号代码和以 # 号开头的分机号的电话号码。
rule_GBR_Phone_Number_Validation	验证英国电话号码的区号和长度。该规则将返回电话号码的所属地区，以及表示电话号码的区号和长度是否有效的代码。

名称	说明
rule_Prename_Assignment	根据性别生成敬称。您可以将 female_prename 表达式变量从 “Ms.” 更改为 “Mrs”。
rule_Salutation_Assignment	根据名字和姓名标志生成正式和非正式问候。例如，当输入数据包含 “Mr. John Smith,” 时，该规则会生成正式问候 “Dear Mr. Smith,” 和非正式问候 “Dear John,”。您可以通过编辑 dq_Generate_Salutation Expression 表达式转换中的变量来更改前缀和标点。

英国公司数据清理规则

使用公司数据清理规则可将公司数据标准化。

您可以在以下存储库位置找到公司数据清理规则：

[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing

下表介绍了英国加速器中的公司数据清理规则：

名称	说明
rule_GBR_Company_Name_Standardization	对公司名称进行标准化，并在可能的情况下提供名称的首字母缩略词。

英国金融数据清理规则

使用金融数据清理规则可解析、标准化和验证金融数据。

您可以在以下存储库位置找到金融数据清理规则：

[Informatica_DQ_Content]\Rules\Financial_Data_Cleansing

下表介绍了英国加速器中的金融数据清理规则：

名称	说明
rule_GBR_Bank_Account_Parse	将八位数字字符串解析为英国银行帐号。
rule_GBR_Bank_Account_Validation	验证英国银行帐号。该规则将返回表示输入是否为数字以及位数是否正确的代码。
rule_GBR_Bank_Sort_Code_Parse	将六位数字字符串解析为英国银行清算代码 (Sort Code)。该规则可解析以下格式的数字字符串： <ul style="list-style-type: none"> - 连续数字 (999999) - 使用短划线分隔的数字 (99-99-99)

名称	说明
rule_GBR_Bank_Sort_Code_Validation	验证已标准化为短划线分隔格式 (99-99-99) 的英国银行清算代码 (Sort Code) 的格式和长度。该规则会返回一个描述该清算代码有效性的状态端口和一个解释该状态的验证说明端口。如果清算代码前缀与某个已知的英国银行代码分配相匹配，验证说明端口将包含该银行名称。
rule_GBR_Bank_Sort_Code_Standardize	将英国银行清算代码 (Sort Code) 标准化为 “NN-NN-NN” 格式。

英国常规数据清理规则

使用常规数据清理规则可识别输入字段中所含信息的类型。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]\Rules\General_Data_Cleansing

下表介绍了英国加速器中的常规数据清理规则：

名称	说明
rule_GBR_NER_Field_Identification	识别输入字段中包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和英国地址数据。该规则会返回一个说明输入数据类型的标签。它使用引用数据识别信息类型。该规则使用概率匹配技术识别信息类型。

英国加速器依赖于核心加速器中的以下常规数据清理规则：

- rule_Assign_DQ_GeocodingStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Remove_Extra_Spaces
- rule_Remove_Leading_Zero
- rule_Remove_Period_Parentheses
- rule_Remove_Punctuation
- rule_Remove_Punctuation_and_Space
- rule_Remove_Space
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

有关这些规则的详细信息，请参阅 [“核心常规数据清理规则”](#) 页面上 22。

英国匹配和消除重复记录规则

使用匹配和消除重复记录规则可度量数据集中记录之间的相似性级别。

您可以在以下存储库位置找到匹配和消除重复记录规则：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了英国加速器中的匹配和消除重复记录规则：

名称	说明
mplt_GBR_Company_Name_Postcode_Match	使用字段匹配策略根据公司名称和邮政编码识别英国数据中的重复行。该 Mapplet 会根据邮政编码生成组键。
mplt_GBR_Familyname_and_NINO_Match	使用字段匹配策略根据姓氏和国家社会保险号 (National Insurance Number, NINO) 识别英国数据中的重复行。该 Mapplet 会根据 NINO 数据生成组键。
mplt_GBR_Familyname_and_Postcode_Match	使用字段匹配策略根据姓氏和英国邮政编码识别英国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_GBR_Firstname_3Chars Surname_DOB_and_Postcode_Match	使用字段匹配策略根据以下数据识别英国数据中的重复行： <ul style="list-style-type: none">- 名字- 姓氏中的前三个字符- 出生日期- 邮政编码 该 Mapplet 会根据邮政编码数据生成组键。
mplt_GBR_Firstname_Surname_2ElementsDOB_and_Postcode_Match	使用字段匹配策略根据以下数据识别英国数据中的重复行： <ul style="list-style-type: none">- 人名- 任意两个出生日期元素，例如月份和年份- 英国邮政编码 该 Mapplet 会根据邮政编码数据生成组键。
mplt_GBR_Firstname_Surname_DOB_and_Postcode_Match	使用字段匹配策略根据以下数据识别行： <ul style="list-style-type: none">- 人名- 出生日期- 邮政编码 该 Mapplet 会根据邮政编码数据生成组键。
mplt_GBR_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别英国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_GBR_IMO_Familyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别英国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_GBR_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别英国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_GBR_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别英国数据中的重复行。个人数据列中的字段必须包含单一类型的数据，如电话号码、电子邮件或国家社会保险号。该 Mapplet 会根据个人数据生成组键。
mplt_GBR_Individual_Name_and_Date_Match	使用字段匹配策略根据人名和日期数据识别英国数据中的重复行。该 Mapplet 会生成根据日期数据生成的组键。
mplt_GBR_Individual_Name_and_Email_Match	使用字段匹配策略根据人名和电子邮件地址数据识别英国数据中的重复行。该 Mapplet 会根据电子邮件地址数据生成组键。

名称	说明
mplt_GBR_Individual_Name_and_NINO_Match	使用字段匹配策略根据人名和国家社会保险号 (National Insurance Number, NINO) 识别英国数据中的重复行。该 Mapplet 会根据 NINO 数据生成组键。
mplt_GBR_Individual_Name_and_Phone_Match	使用字段匹配策略根据人名和电话号码识别英国数据中的重复行。该 Mapplet 会根据电话号码数据生成组键。
mplt_GBR_Individual_Name_and_Postcode_Match	使用字段匹配策略根据人名和邮政编码数据识别英国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_GBR_Individual_Name_Match	使用字段匹配策略根据人名识别英国数据中的重复行。Mapplet 将根据姓氏值生成 NYSIIS 代码，并使用该 NYSIIS 代码作为组键。
rule_GBR_Familyname_and_NINO_MatchScore	根据姓氏和英国国家社会保险号 (United Kingdom National Identification Numbers, NINO) 生成匹配得分。
rule_GBR_Familyname_and_Postcode_MatchScore	根据姓氏和英国邮政编码生成匹配得分。
rule_GBR_Firstname_3CharsSurname_DOB_and_Postcode_MatchScore	根据以下信息生成匹配得分： <ul style="list-style-type: none"> - 名字 - 姓氏中的前三个字符 - 出生日期 - 邮政编码
rule_GBR_Firstname_Surname_2ElementsDOB_and_Postcode_MatchScore	根据以下信息生成匹配得分： <ul style="list-style-type: none"> - 人名 - 任意两个出生日期元素，例如月份和年份 - 英国邮政编码
rule_GBR_Firstname_Surname_DOB_and_Postcode_MatchScore	根据人名、出生日期和邮政编码生成匹配得分。
rule_GBR_Individual_Name_and_NINO_MatchScore	根据人名和英国国家社会保险号 (United Kingdom National Insurance Number, NINO) 生成匹配得分。
rule_GBR_Individual_Name_and_Phone_MatchScore	根据人名和电话号码生成匹配得分。
rule_GBR_Individual_Name_and_Postcode_MatchScore	根据人名和英国邮政编码生成匹配得分。
rule_Individual_Name_and_Date_MatchScore	根据人名和日期生成匹配得分。
rule_Individual_Name_MatchScore	根据人名生成匹配得分。
rule_GBR_Company_Name_Postcode_MatchScore	根据公司名称和英国邮政编码生成匹配得分。
rule_Individual_Name_and_Email_MatchScore	根据人名和电子邮件地址生成匹配得分。

英国演示映射

英国加速器中的演示映射使用多个规则演示数据质量过程。

您可以在以下存储库位置找到演示映射：

[Informatica_DQ_Content]\Rules_Demo\GBR_Accelerator

英国加速器包含以下演示映射：

m_GBR_customer_data_demo

解析、标准化和验证英国客户数据。

m_GBR_customer_matching_demo

解析英国的标识数据并将其标准化，并对该数据执行标识匹配分析。

该映射会分析以下数据组合，并为每个组合生成匹配群集：

- 人名和地址数据
- 人名和电话号码

第 12 章

美国/加拿大加速器

本章包括以下主题：

- [美国/加拿大加速器概览, 110](#)
- [美国/加拿大地址数据清理规则, 110](#)
- [美国/加拿大复合规则, 112](#)
- [美国/加拿大联系人数据清理规则, 114](#)
- [美国/加拿大公司数据清理规则, 119](#)
- [美国/加拿大常规数据清理规则, 119](#)
- [美国/加拿大匹配和消除重复记录规则, 120](#)
- [美国/加拿大演示映射, 122](#)

美国/加拿大加速器概览

使用美国/加拿大加速器中的规则可验证和改善美国和加拿大组织的数据。

美国/加拿大加速器包括执行以下数据质量过程的规则：

- 地址数据清理
- 联系人数据清理
- 常规数据清理
- 匹配和消除重复记录

该加速器依赖于核心加速器安装的数据清理规则。

美国/加拿大地址数据清理规则

使用地址数据清理规则可解析、标准化和验证地址数据。

地址数据清理规则位于以下存储库位置：

[Informatica_DQ_Content]\Rules\Address_Data_Cleansing

下表介绍了美国/加拿大加速器中的地址数据清理规则：

名称	说明
rule_CAN_Address_Certification_Hybrid	<p>对照加拿大邮政为加拿大地址维护的软件评估和认证计划 (Software Evaluation and Recognition Program, SERP) 标准，验证加拿大地址的可交付性。该规则在可能的情况下更正输入地址中的错误，并且包括一个标识符合 SERP 标准的地址的状态端口。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。</p> <p>注意: 当您对照 SERP 标准认证一组地址记录时，您必须向加拿大邮政提交一份认证报告。地址验证器转换包括可以使用报告信息填充的属性字段。保存或打印该报告，并在报告中包含向加拿大邮政提交的地址详情。</p>
rule_CAN_Address_Parse_Hybrid	<p>将非结构化的加拿大地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。</p>
rule_CAN_Address_Parse_Multiline	<p>将非结构化的加拿大地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。</p>
rule_CAN_Address_Validation_Discrete_w_Geocoding	<p>验证加拿大地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。</p>
rule_CAN_Address_Validation_Discrete	<p>验证加拿大地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。</p>
rule_CAN_Address_Validation_Hybrid_w_Geocoding	<p>验证加拿大地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。</p>
rule_CAN_Address_Validation_Hybrid	<p>验证加拿大地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。</p>
rule_CAN_Address_Validation_Multiline_w_Geocoding	<p>验证加拿大地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。</p>
rule_CAN_Address_Validation_Multiline	<p>验证加拿大地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。</p>
rule_CAN_Postcode_Validation	<p>验证加拿大邮政编码。该规则将返回“有效”或“无效”。</p>
rule_CAN_Province_Validation	<p>验证加拿大省份名称。该规则将返回“有效”或“无效”。</p>
rule_USA_Address_Certification_Hybrid	<p>对照美国邮政总局 (United States Postal Service) 为美国地址维护的编码准确性支持系统 (Coding Accuracy Support System, CASS) 标准，验证美国地址的可交付性。该规则在可能的情况下更正输入地址中的错误，并且包括一个标识符合 CASS 标准的地址的状态端口。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。</p> <p>注意: 当您对照 CASS 标准认证一组地址记录时，您必须向 USPS 提交一份认证报告。地址验证器转换包括可以使用报告信息填充的属性字段。保存或打印该报告，并在报告中包含向 USPS 提交的地址详情。</p>

名称	说明
rule_USA_Address_Parse_Hybrid	将非结构化的美国地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_USA_Address_Parse_Multiline	将非结构化的美国地址解析为地址元素。该规则并不验证地址。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_USA_Address_Validation_Discrete_w_Geocoding	验证美国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_USA_Address_Validation_Discrete	验证美国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的离散值输入端口时，请使用该规则。
rule_USA_Address_Validation_Hybrid_w_Geocoding	验证美国地址中地址记录的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_USA_Address_Validation_Hybrid	验证美国地址中地址记录的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的混合输入端口时，请使用该规则。
rule_USA_Address_Validation_Multiline_w_Geocoding	验证美国地址的可送达性，并将纬度和经度坐标添加到每个输出地址。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_USA_Address_Validation_Multiline	验证美国地址的可送达性。如果可能，该规则会更正输入地址中的错误。当您可以将输入地址字段连接到地址验证器转换的多行输入端口时，请使用该规则。
rule_USA_County_Validation	验证美国县名。该规则会将输入数据与所有州的县名进行比较。该规则将返回“有效”或“无效”。
rule_USA_State_Validation	验证美国州名。该规则将返回“有效”或“无效”。
rule_USA_ZIPCode_Validation	验证五位数字的美国地区改进计划 (Zone Improvement Plan, ZIP) 编码。该规则将返回“有效”或“无效”。

美国/加拿大复合规则

使用美国/加拿大加速器中的复合规则，将一组规则作为单个对象添加到映射中。复合规则是一种利用其他加速器规则的逻辑的规则。

您可以在以下存储库位置找到复合规则：

[Informatica_DQ_Content]\Rules\Composite_Rules

下表介绍了美国/加拿大加速器中的复合规则：

名称	说明
rule_CAN_Contact_Data	解析、标准化和验证地址、电话号码和社会保险号 (Social Insurance Numbers, SIN) 等加拿大联系人数据。
rule_USA_Contact_Data	解析、标准化和验证地址、电话号码和社会保障号 (Social Security Numbers, SSN) 等美国联系人数据。

适用于加拿大联系人数据的复合规则

规则 rule_CAN_Contact_Data 从存储库中的多个文件夹读取 Mapplet。该规则还包括一个不可重用的转换。

下表列出了 rule_CAN_Contact_Data 中的规则和转换的名称与存储库位置：

规则	位置
Case_Converter	不可重用的转换
rule_Assign_DQ_Mailability_Score_Description	Informatica_DQ_Content]\Rules \General_Data_Cleansing
rule_CAN_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules \Address_Data_Cleansing
rule_CAN_Gender_Assignment	Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_CAN_Multi_Person_Name_Parse	Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_CAN_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_CAN_Phone_Number_Validation	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_CAN_SIN_Standardization	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_CAN_SIN_Validation	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_Company_Name_Standardization	[Informatica_DQ_Content]\Rules \Corporate_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules \Contact_Data_Cleansing

适用于美国联系人数据的复合规则

规则 rule_USA_Contact_Data 从存储库中的多个文件夹读取 Mapplet。该规则还包括一个不可重用的转换。

下表列出了 rule_USA_Contact_Data 中的规则和转换的名称与存储库位置：

规则	位置
Case_Converter	不可重用的转换
rule_Assign_DQ_Mailability_Score_Description	[Informatica_DQ_Content]\Rules\General_Data_Cleansing
rule_Company_Name_Standardization	[Informatica_DQ_Content]\Rules\Corporate_Data_Cleansing
rule_Email_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Prename_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_Salutation_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_Address_Validation_Hybrid	[Informatica_DQ_Content]\Rules\Address_Data_Cleansing
rule_USA_Gender_Assignment	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_Multi_Person_Name_Parse	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_Phone_Number_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_Phone_Number_Validation	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_SSN_Standardization	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing
rule_USA_SSN_Validation_post_June2011	[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

美国/加拿大联系人数据清理规则

使用联系人数据清理规则可解析、标准化和验证有关业务联系人和个人的数据。

您可以在以下存储库位置找到联系人数据清理规则：

[Informatica_DQ_Content]\Rules\Contact_Data_Cleansing

下表介绍了美国/加拿大加速器中的联系人数据清理规则：

名称	说明
rule_CAN_Gender_Assignment	按照名字分配性别。该规则会为男性名字返回“M”，为女性名字返回“F”，为未知性别返回“U”。例如，该规则会为姓名“John Smith”分配代表男性的“M”性别。
rule_CAN_Given_Name_Standard	根据加拿大昵称生成名字。例如，该规则会将昵称“Bob”标准化为名字“Robert”。
rule_CAN_Multi_Person_Name_Parse	将人名值解析为单独的端口。该规则会为头衔、名字、中间名和姓氏等值创建端口。 规则输出包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。 当姓名数据标识了多位人员时，该规则会为每个全名创建一个输出端口。例如，该规则可读取姓名“John and Jane Smith”，并为“John Smith”和“Jane Smith”创建输出端口。
rule_CAN_Personal_Name_Parse_and_Standardize_FML	将人名中的值解析为单独的端口。该规则还会将人名值标准化。 该规则按以下顺序创建端口： - 名字、中间名、姓氏 规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。
rule_CAN_Personal_Name_Parse_and_Standardize_LFM	将人名中的值解析为单独的端口。该规则还会将人名值标准化。 该规则按以下顺序创建端口： - 姓氏、名字、中间名 规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。
rule_CAN_Personal_Name_Parsing_FML	将人名中的值解析为单独的端口。 该规则按以下顺序创建端口： - 姓氏、名字、中间名 规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。 注意: 该规则不将人名值标准化。要按照规则定义的顺序对加拿大人名值进行标准化和解析，请选择 rule_CAN_Personal_Name_Parse_and_Standardize_FML。
rule_CAN_Personal_Name_Parsing_LFM	将人名中的值解析为单独的端口。 该规则按以下顺序创建端口： - 姓氏、名字、中间名 规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。 注意: 该规则不将人名值标准化。要按照规则定义的顺序对加拿大人名值进行标准化和解析，请选择 rule_CAN_Personal_Name_Parse_and_Standardize_LFM。
rule_CAN_Phone_Number_Parse	解析字符串中的加拿大电话号码。该规则将解析数据中的第一个电话号码（从右到左读取）。该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。

名称	说明
rule_CAN_Phone_Number_Standardization	将加拿大电话号码标准化。该规则可返回以下格式的电话号码： <ul style="list-style-type: none"> - 标准 - (nnn) nnn-nnnn - 短划线 - nnn-nnn-nnnn - 无空格 - nnnnnnnnn
rule_CAN_Phone_Number_Validation	验证加拿大电话号码的区号和长度。该规则将返回表示电话号码类型和有效性的代码。这些类型将说明各种类别，如“免费号码”。
rule_CAN_Phone_Parse_Standardize_Validate	从文本字符串解析电话号码，并验证区号是否对加拿大有效。 如果区号有效，规则将以三种标准格式返回电话号码。此外，该规则还会返回一个状态值，指示数据是否符合加拿大电话号码的标准格式。
rule_CAN_Phone_w_Extension_Parse	如果电话号码符合加拿大电话号码的标准格式，则从文本字符串解析电话号码。解析电话号码时，该规则包括任何电话分机数据。
rule_CAN_SIN_Parse	解析字符串中的加拿大社会保险号 (Social Insurance Number, SIN)。该规则将返回 SIN，以及一个包含已删除 SIN 的输入文本的字符串。
rule_CAN_SIN_Standardization	将加拿大社会保险号 (Social Insurance Number, SIN) 标准化。该规则可以输出以下格式： <ul style="list-style-type: none"> - 无标点 - nnnnnnnnn - 空格 - nnn nnn nnn - 短划线 - nnn-nnn-nnn 要更改格式，请编辑 dq_Format_SIN 表达式转换中的 SIN_Format 表达式变量。默认值为“No_Punctuation”。
rule_CAN_SIN_Validation	验证加拿大社会保险号 (Social Insurance Number, SIN)。该规则将使用 Luhn 算法验证 SIN 是否有效。该规则将返回“有效”或“无效”。
rule_Prename_Assignment	根据性别生成敬称。您可以将 female_prename 表达式变量从“Ms.”更改为“Mrs”。
rule_Salutation_Assignment	根据名字和姓名标志生成正式和非正式问候。例如，当输入数据包含“Mr. John Smith,”时，该规则会生成正式问候“Dear Mr. Smith,”和非正式问候“Dear John,”。您可以通过编辑 dq_Generate_Salutation Expression 表达式转换中的变量来更改前缀和标点。
rule_USA_Gender_Assignment	按照名字分配性别。该规则会为男性名字返回“M”，为女性名字返回“F”，为未知性别返回“U”。例如，该规则会为姓名“John Smith”分配代表男性的“M”性别。
rule_USA_Given_Name_Standard	根据美国昵称生成名字。例如，该规则会将昵称“Bob”标准化为名字“Robert”。
rule_USA_Multi_Person_Name_Parse	将人名值解析为单独的端口。该规则会为头衔、名字、中间名和姓氏等值创建端口。 规则输出包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。 当姓名数据标识了多位人员时，该规则会为每个全名创建一个输出端口。例如，该规则可读取姓名“John and Jane Smith”，并为“John Smith”和“Jane Smith”创建输出端口。

名称	说明
rule_USA_Personal_Name_Parse_and_Standardize_FML	<p>将人名中的值解析为单独的端口。该规则还会将人名值标准化。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 名字、中间名、姓氏 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p>
rule_USA_Personal_Name_Parse_and_Standardize_LFM	<p>将人名中的值解析为单独的端口。该规则还会将人名值标准化。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 姓氏、名字、中间名 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p>
rule_USA_Personal_Name_Parse_Validation	<p>验证姓名的性别分配。该规则计算数据值为男性名字或女性名字的概率。如果性别未知，该规则会使用概率计算为姓名分配性别。</p>
rule_USA_Personal_Name_Parsing_FML	<p>将人名中的值解析为单独的端口。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 名字、中间名、姓氏 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p> <p>注意: 该规则不将人名值标准化。要按照规则定义的顺序对美国人名值进行标准化和解析，请选择 rule_USA_Personal_Name_Parse_and_Standardize_FML。</p>
rule_USA_Personal_Name_Parsing_LFM	<p>将人名中的值解析为单独的端口。</p> <p>该规则按以下顺序创建端口：</p> <ul style="list-style-type: none"> - 姓氏、名字、中间名 <p>规则输出还包括一个包含记录中人员全名的端口。您可以在标识匹配分析映射中使用全名端口作为匹配转换的输入。</p> <p>注意: 该规则不将人名值标准化。要按照规则定义的顺序对美国人名值进行标准化和解析，请选择 rule_USA_Personal_Name_Parse_and_Standardize_LFM。</p>
rule_USA_Phone_Number_Parse	<p>解析字符串中的美国电话号码。该规则将解析数据中的第一个电话号码（从右到左读取）。该规则将返回一个电话号码，以及一个包含输入文本（但已删除该电话号码）的字符串。</p>
rule_USA_Phone_Number_Standardization	<p>将美国电话号码标准化。该规则可返回以下格式的电话号码：</p> <ul style="list-style-type: none"> - 标准 - (nnn) nnn-nnnn - 短划线 - nnn-nnn-nnnn - 无空格 - nnnnnnnnnn
rule_USA_Phone_Number_Validation	<p>验证美国电话号码的区号和长度。该规则将返回表示电话号码区号和长度是否有效的代码。</p>
rule_USA_Phone_Parse_Standardize_Validate	<p>从文本字符串解析电话号码，并验证区号是否对美国有效。</p> <p>如果区号有效，规则将以三种标准格式返回电话号码。此外，该规则还返回一个状态值，指示数据是否符合美国电话号码的标准格式。</p>
rule_USA_Phone_w_Extension_Parse	<p>如果电话号码符合美国电话号码的标准格式，则从文本字符串解析电话号码。解析电话号码时，该规则包括任何电话机数据。</p>
rule_USA_SSN_Parse	<p>解析美国社会保障号 (Social Security Numbers, SSN)。</p>

名称	说明
rule_USA_SSN_Parse_Standardize_and_Validate	<p>从一个较大的文本字符串解析、标准化和验证美国社会保障号。该规则可以解析包括或忽略破折号的编号。</p> <p>默认情况下，该规则写入不带任何标点和社会保障号。要更改标准化格式，请打开规则中的 dq_SSN_Format 转换，更新 SSN_Format 端口上的表达式。</p>
rule_USA_SSN_Standardization	<p>将美国社会保障号 (Social Security Numbers, SSN) 标准化。该规则可以输出以下格式：</p> <ul style="list-style-type: none"> - 无标点 - nnnnnnnnnn - 空格 - nnn nnn nnn - 短划线 - nnn-nnn-nnn <p>要更改格式，请编辑 dq_SSN_Format 表达式转换中的 SSN_format 表达式变量。默认值为 “No_Punctuation”。</p>
rule_USA_SSN_Validation	<p>验证美国社会保障号 (Social Security Numbers, SSN)。该规则将验证每个 SSN 的长度、数值以及 “区域”、“组” 和 “序列号” 各区段的已知最小值和最大值。</p> <p>“区域” 区段由 SSN 的前三位构成，“组” 区段由第四位和第五位构成。</p> <p>“序列号” 区段由最后四位构成。</p> <p>如果 SSN 是在 2011 年 6 月之前发放的，该规则还将验证 “区域” 值和 “组” 值是否为有效组合。该规则不验证 SSN 是否为已发放的号码。该规则将返回 “有效” 或 “无效”。</p>
rule_USA_SSN_Validation_post_June2011	<p>验证美国社会保障号 (Social Security Numbers, SSN)。该规则将验证每个 SSN 的长度、数值以及 “区域”、“组” 和 “序列号” 各区段的已知最小值和最大值。</p> <p>“区域” 区段由 SSN 的前三位构成，“组” 区段由第四位和第五位构成。</p> <p>“序列号” 区段由最后四位构成。</p> <p>该规则不验证 “区域” 值和 “组” 值是否为有效组合。该规则不验证 SSN 是否为已发放的号码。该规则将返回 “有效” 或 “无效”。</p>

与核心联系人数据清理规则的相关性

美国/加拿大加速器依赖于核心加速器中的以下联系人数据清理规则：

- rule_Email_Validation

有关这些规则的详细信息，请参阅 [“核心联系人数据清理规则” 页面上 21](#)。

美国/加拿大公司数据清理规则

使用美国/加拿大加速器中的公司数据清理规则可解析、标准化和验证公司数据。

下表介绍了美国/加拿大加速器中的公司数据清理规则：

名称	说明
rule_NAICS_Code_Validation	验证北美工业分类系统 (North American Industry Classification System, NAICS) 代码。
rule_USA_SIC_Code_Validation	验证标准工业分类 (Standard Industrial Classification, SIC) 代码。

美国/加拿大加速器依赖于核心加速器中的以下公司数据清理规则：

- rule_Company_Name_Standardization

美国/加拿大常规数据清理规则

使用常规数据清理规则可识别输入字段中所含信息的类型。

您可以在以下存储库位置找到常规数据清理规则：

[Informatica_DQ_Content]\Rules\General_Data_Cleansing

下表介绍了美国/加拿大加速器中的常规数据清理规则：

名称	说明
rule_CAN_Field_Identification	识别输入字段包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和加拿大地址数据。该规则会返回一个说明输入数据类型的标签。它使用引用数据识别信息类型。
rule_CAN_NER_Field_Identification	识别输入字段包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和加拿大地址数据。该规则会返回一个说明输入数据类型的标签。它使用引用数据识别信息类型。该规则使用概率匹配技术识别信息类型。
rule_USA_Field_Identification	识别输入字段包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和美国地址数据。该规则会返回一个说明输入数据类型的标签。它使用引用数据识别信息类型。
rule_Field_North_American_Data	识别以下类型的字段：名称、头衔、公司、地址、城市、州或省、邮政编码、国家/地区、个人 ID、电子邮件、电话、信用卡和日期。 该规则生成一个分数，指示字段标识的置信度。较高的分数表示较高的置信度。 如果规则无法分配字段类型，该规则将在 Out_Undetermined 端口上写入数据。
rule_USA_NER_Field_Identification	识别输入字段包含的信息类型。该规则可以识别姓名、个人 ID、公司名称、日期和美国地址数据。该规则会返回一个说明输入数据类型的标签。它使用引用数据识别信息类型。该规则使用概率匹配技术识别信息类型。

与核心常规数据清理规则的相关性

美国/加拿大加速器依赖于核心加速器中的以下常规数据清理规则：

- rule_Assign_DQ_GeocodinStatus_Description
- rule_Assign_DQ_Mailability_Score_Description
- rule_Assign_DQ_Match_Code_Descriptions
- rule_Date_Validation
- rule_Remove_Extra_Spaces
- rule_Remove_Punctuation
- rule_Replace_Limited_Punct_with_Space
- rule_UpperCase

有关这些规则的详细信息，请参阅 [“核心常规数据清理规则” 页面上 22。](#)

美国/加拿大匹配和消除重复记录规则

使用匹配和消除重复记录规则可度量数据集中记录之间的相似性级别。

您可以在以下存储库位置找到匹配和消除重复记录规则：

[Informatica_DQ_Content]\Rules\Matching_Deduplication

下表介绍了美国/加拿大加速器中的匹配和消除重复记录规则：

名称	说明
mplt_CAN_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别加拿大数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_CAN_IMO_Famillyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别加拿大数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_CAN_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别加拿大数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_CAN_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别加拿大数据中的重复行。个人数据列中的字段应包含单一类型的数据，如电话号码、电子邮件或社会保险号。该 Mapplet 会根据个人数据生成组键。
mplt_Company_Name_and_Address_Match	使用字段匹配策略根据公司名称和地址数据识别重复行。该 Mapplet 将使用公司名称值和邮政编码值的字符组合生成组键。
mplt_Company_Name_Match	使用字段匹配策略根据公司名称识别重复行。该 Mapplet 将根据公司名称值生成 Soundex 代码，并使用该 Soundex 代码作为组键。
mplt_Famillyname_and_Address_Match	使用字段匹配策略根据姓氏和地址数据识别重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_Firstname_and_SSN_Match	使用字段匹配策略根据名字和美国社会保障号识别重复行。该 Mapplet 会根据社会保障号数据生成组键。

名称	说明
mplt_Individual_Name_and_Address_Match	使用字段匹配策略根据人名和美国地址数据识别重复行。该 Mapplet 将使用姓氏值和邮政编码值的字符组合生成组键。
mplt_Individual_Name_and_Date_Match	使用字段匹配策略根据人名和日期数据识别重复行。该 Mapplet 会生成根据日期数据生成的组键。
mplt_Individual_Name_and_Email_Match	使用字段匹配策略根据人名和电子邮件地址识别重复行。该 Mapplet 会根据电子邮件地址数据生成组键。
mplt_Individual_Name_and_Phone_Match	使用字段匹配策略根据人名和电话号码识别重复行。该 Mapplet 会根据电话号码数据生成组键。
mplt_Individual_Name_and_SSN_Match	使用字段匹配策略根据人名和美国社会保障号识别重复行。该 Mapplet 会根据社会保障号数据生成组键。
mplt_Individual_Name_Match	使用字段匹配策略根据人名识别重复行。Mapplet 将根据姓氏值生成 NYSIIS 代码，并使用该 NYSIIS 代码作为组键。
mplt_USA_Address_Match	使用字段匹配策略根据美国地址数据识别美国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_USA_IMO_Company_Name_and_Address_Match	使用标识匹配策略根据公司名称和地址识别美国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_USA_IMO_Familyname_and_Address_Match	使用标识匹配策略根据姓氏和地址识别美国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_USA_IMO_Individual_Name_and_Address_Match	使用标识匹配策略根据人名和地址识别美国数据中的重复行。该 Mapplet 会根据邮政编码数据生成组键。
mplt_USA_IMO_Personal_Name_and_Data_Match	使用标识匹配策略根据人名和个人数据识别美国数据中的重复行。个人数据列中的字段必须包含单一类型的数据，如电话号码、电子邮件或社会保障号。该 Mapplet 会根据个人数据生成组键。
rule_Company_Name_and_Address_MatchScore	根据公司名称和美国地址数据生成匹配得分。
rule_Company_Name_MatchScore	根据公司名称生成匹配得分。
rule_Familyname_and_Address_MatchScore	根据姓氏和美国地址数据生成匹配得分。
rule_Firstname_and_SSN_MatchScore	根据名字和美国地址数据生成匹配得分。
rule_Individual_Name_and_Address_MatchScore	根据人名和美国地址数据生成匹配得分。
rule_Individual_Name_and_Date_MatchScore	根据人名和日期生成匹配得分。
rule_Individual_Name_and_Email_MatchScore	根据人名和电子邮件地址生成匹配得分。

名称	说明
rule_Individual_Name_and_Phone_MatchScore	根据人名和电话号码生成匹配得分。
rule_Individual_Name_and_SN_MatchScore	根据人名、社会保障号和身份数据生成匹配得分。
rule_Individual_Name_MatchScore	根据人名生成匹配得分。
rule_USA_Address_MatchScore	根据美国地址数据生成匹配得分。

美国/加拿大演示映射

美国/加拿大加速器中的演示映射使用多个规则演示数据质量过程。

您可以在以下存储库位置找到演示映射：

[Informatica_DQ_Content]\Rules_Demo\US_Canada_Accelerator

美国/加拿大加速器包括以下演示映射：

m_customer_data_US_demo

解析、标准化和验证美国和加拿大数据。

m_customer_matching_US_demo

解析美国的标识数据并将其标准化，并对该数据执行标识匹配分析。

该映射会分析以下数据组合，并为每个组合生成匹配群集：

- 人名和地址数据
- 人名和电话号码