



ICAO

Doc 9303

机读旅行证件

第八版, 2021年

第3部分: 所有机读旅行证件的通用规范



经秘书长批准并由其授权出版

国际民用航空组织



| ICAO

Doc 9303

机读旅行证件

第八版, 2021年

第3部分: 所有机读旅行证件的通用规范

经秘书长批准并由其授权出版

国际民用航空组织

国际民用航空组织分别以中文、阿拉伯文、英文、法文、俄文和西班牙文版本出版
999 Robert-Bourassa Boulevard, Montréal, Quebec, Canada H3C 5H7

下载文件和获取额外信息，请登录 www.icao.int/Security/FAL/TRIP。

Doc 9303 号文件 — 《机读旅行证件》
第 3 部分 — 所有机读旅行证件的通用规范
订购编号：9303P3
ISBN 978-92-9265-490-0（印刷版）
ISBN 978-92-9275-347-4（电子版）

© ICAO 2021

保留所有权利。未经国际民用航空组织事先书面许可，不得将本出版物的任何部分复制、存储于检索系统或以任何形式或手段进行发送。

修订

《产品和服务目录》的补篇中公布了各项修订；在国际民航组织网站 www.icao.int 上有本目录及其补篇。以下篇幅供记录修订之用。

修订和更正记录

修订			更正			
编号	日期	换页人	编号	日期	语文	换页人
1	14/11/22	ICAO				
2	20/3/24	ICAO				

本出版物中所用称谓和陈述材料之方式，并不代表国际民航组织对任何国家、领土、城市或地区或其当局的法律地位，或就其边境或疆界的划分，表达了任何意见。

目录

页码

1. 范围	1
2. 机读旅行证件的物理特性	2
3. 视读区 (VIZ)	2
3.1 语文和字符.....	2
3.2 字体和字号.....	3
3.3 说明文字/域.....	4
3.4 持有人姓名的书写惯例.....	4
3.5 签发国或签发机构的表示方式.....	5
3.6 国籍的表示方式.....	5
3.7 出生地的表示方式.....	5
3.8 日期的表示方式.....	6
3.9 显示的持有人识别特征.....	8
4. 机读区 (MRZ)	15
4.1 机读区的用途.....	15
4.2 机读区的特性.....	15
4.3 机读区的限制条件.....	15
4.4 打印规格.....	16
4.5 机读要求和有效阅读区.....	16
4.6 持有人姓名的书写惯例.....	17
4.7 签发国或签发机构和持有人国籍的表示方式.....	19
4.8 日期的表示方式.....	19
4.9 机读区内的校验数位.....	19
4.10 机读区的特征.....	20
4.11 机读区的质量规范.....	20
5. 代表国籍、出生地、签发国/机构所在地和其他情况的代码	21
6. 建议各国使用的转写方法	24
7. 偏差	32
7.1 运作经验.....	32
7.2 偏差列表做法.....	32
7.3 方法.....	33
7.4 公布.....	39

8. 参考文献（规范性）	40
第 3 部分附录 A 校验数位计算方法示例（资料性）	APP A-1
第 3 部分附录 B 机读旅行证件中阿拉伯文字的转写（资料性）	APP B-1
B.1 阿拉伯文字	APP B-1
B.2 机读旅行证件中的阿拉伯文字	APP B-1
B.3 对视读区的建议	APP B-3
B.4 机读区内字符的转写	APP B-5
B.5 关于机读区的建议	APP B-6
B.6 机读区中的逆向转写	APP B-15
B.7 计算机程序	APP B-17
B.8 参考文献（资料性）	APP B-20

1. 范围

第 3 部分明确了通用于 TD1 型、TD2 型和 TD3 型机读旅行证件（MRTDs）的规范，包括为确保通过视读和机读（光学字符识别）手段实现全球互操作性所需遵守的规范。适用于每种证件类型的具体规范，见 Doc 9303 号文件第 4 部分至第 7 部分。

阅读第 3 部分须结合以下部分进行阅读：

- 第 1 部分 — 引言；
- 第 2 部分 — 机读旅行证件的设计、制作和签发的安全性规范；

以及与特定尺寸旅行证件相关的部分：

- 第 4 部分 — 机读护照（MRPs）和其他 TD3 型机读旅行证件规范；
- 第 5 部分 — TD1 型机读官方旅行证件（MROTDs）规范；
- 第 6 部分 — TD2 型机读官方旅行证件（MROTDs）规范；和
- 第 7 部分 — 机读签证。

这些规范同时还适用于包含非接触式集成电路的机读旅行证件，即电子机读旅行证件（eMRTDs）。仅适用于电子机读旅行证件的规范载于 Doc 9303 号文件的以下部分：

- 第 9 部分 — 生物特征识别技术的运用和机读旅行证件的电子数据存储；
- 第 10 部分 — 在非接触式集成电路（IC）中存储生物特征和其他数据的逻辑数据结构（LDS）；
- 第 11 部分 — 机读旅行证件的安全机制；和
- 第 12 部分 — 机读旅行证件的公钥基础设施。

2. 机读旅行证件的物理特性

签发国和签发机构可自由选择用于制作旅行证件的材料。但是，所选择的材料不得对机读旅行证件中的任何其他组成部分产生不利的影响，并且机读旅行证件在其整个有效期的正常使用中应满足下列要求：

- **变形** 机读旅行证件所用材料应具有可弯性（不起皱折），即由于正常使用造成的变形通过读取装置可以使其变平，而不至于影响其使用或阅读器的功能。
- **毒性** 根据[ISO/IEC 7810]的规定，在正常使用过程中，机读旅行证件不得出现毒性危险。
- **抗化学品** 机读旅行证件应能抵抗正常处理和使用过程中出现的化学影响，出于安全原因加入的化学敏感性除外。
- **温度稳定性** 机读旅行证件应在-10°C至+50°C（14°F至 122°F）的操作温度范围内保持机读性。机读旅行证件在暴露于-35°C至+80°C（-31°F至 176°F）的温度下不得失去其功能。
- **湿度** 根据[ISO/IEC 7810]的规定，机读旅行证件在湿球温度最高为 25°C（77°F）、空气相对湿度为 5%至 95%之间，应保持机读性。机读旅行证件储存或暴露在 0%至 100%（不冷凝）的相对湿度下不应失去其可靠性。
- **光** 根据[ISO/IEC 7810]的规定，机读旅行证件在正常使用中暴露在光下不应损坏。

3. 视读区（VIZ）

机读旅行证件视读区由针对视读设计的强制性和选择性数据元素构成。选择性数据元素以及强制性数据元素可以在保持足够的统一性以确保所有机读旅行证件全球互操作性的同时，满足签发国和签发机构的不同需要。

3.1 语文和字符

视读区的数据应该用拉丁字母字符，即 A 到 Z 和 a 到 z，和阿拉伯数字，即 1234567890 表示。可以使用变音符。第 6.A 节 — “多国基于拉丁语的字符转写”中所列出的基于拉丁语的国家字符，如 **Ð** 和 **ß**，也可以用于视读区中，无需转写。当强制性数据元素不使用拉丁字母的语文表示时，还应提供抄写或转写。

3.3 说明文字/域

应使用说明文字标明视读区内强制性数据元素使用的所有域，Doc 9303 号文件第 4 部分至第 7 部分中有关每种证件类型的数据元素目录另有规定的除外。

说明文字可以使用签发国的官方语文或签发机构的工作语文。当这些语文使用拉丁字母时，应采用直体字打印说明文字。

如果签发国的官方语文或签发机构的工作语文不是英文、法文或西班牙文，还应在所印的说明文字之后加上一个斜杠字符 (/) 以及说明文字的英文、法文或西班牙文译文。第二种语文应采用斜体字。

如果签发国的官方语文或签发机构的工作语文是英文、法文或西班牙文，则签发国或签发机构应使用另外两种语文中的一种在斜杠字符 (/) 之后打印出说明文字。第二种语文应采用斜体字。

说明文字应采用清晰的直线型字体打印，字号应在 1.0 毫米至 1.8 毫米 (0.04 英寸至 0.07 英寸) 之间。

当某个域未被使用时，不应在旅行证件上显示说明文字。

3.4 持有人姓名的书写惯例

持有人姓名通常用两部分表示：主要标识符和次要标识符。

签发国或签发机构应确定姓名的哪一部分是主要标识符，它可以是姓、婚前姓或夫姓、主名、别名，或者在某种情况下当持有人姓名无法分为两部分时可以是整个姓名。这一部分应登记到视读区的主要标识符域内。建议使用大写字母，但在有诸如“von、”“Mc”或“de la”等前缀的情况下除外。在这些情况下，混合使用大小写是适宜的。

姓名的余下部分为次要标识符，它可以是名、字、教名、首字母或任何其他辅助名。这些名字应写在视读区次要标识符域内。建议全部使用大写字母。

如果姓名是使用单一一个域，那么次要标识符和主要标识符之间须使用一个逗号 (,) 分开。如果是使用多个域，不需要使用逗号。

前缀和后缀，包括头衔、专业和学术资格、荣誉、奖励和世袭地位不应包括在视读区内。但是，如果签发国或签发机构认为这样的前缀或后缀是姓名法定的一部分，那么前缀或后缀可以出现在视读区。数字字符不应写在视读区的姓名域中，但是如果使用数字字符在签发国是合法的命名惯例，这些字符应该用罗马数字表示。任何前缀、后缀或罗马数字应登记到次要标识符域内。

在视读区可以使用本国字符。如果本国字符不是基于拉丁语，那么应将其抄写或转写成拉丁字符。

3.5 签发国或签发机构的表示方式

当签发国或签发机构和/或签发部门或机关所在地的名称用非拉丁字符的语文表示时，国名或其他地点名称应以本国语文/签发机构的工作语文的形式出现，同时还应：

- 转写成拉丁字符；或
- 翻译成使该名称可能更普遍为国际社会所熟知的一种或多种语文（至少其中一种必须是英文、法文或西班牙文）。

不同语文的名称应该用一个斜杠字符（/）分开，后面至少跟一个空格。

当签发国或签发机构或签发部门或机关所在地的名称用拉丁字母的一种语文表示，但是如果将名称翻译成另外一种语文或多种语文（特别是英文、法文或西班牙文）更为国际社会所熟悉时，应配有一种或多种译文。不同语文的名称应该用一个斜杠字符（/）分开，后面至少跟一个空格。

3.6 国籍的表示方式

对于持有人国籍域为视读区强制性元素的证件，签发国或签发机构应自行决定使用三字代码（见第5节）或全称表示持有人国籍。

当国籍采取全称形式且签发国的本国语文或签发机构的工作语文为非拉丁字符的语文时，国籍应以本国/工作语文的形式出现，同时还应：

- 转写成拉丁字符；或
- 翻译成使该国籍可能更普遍为国际社会所熟知的一种或多种语文（至少其中一种必须是英文、法文或西班牙文）。

不同语文的国籍应该用一个斜杠字符（/）分开，后面至少跟一个空格。

当签发国的本国语文或签发机构的工作语文使用拉丁字母，但是如果将国籍翻译成另外一种语文或多种语文（特别是英文、法文或西班牙文）更为国际社会所熟悉时，用本国/语文语文写入的国籍应配有一种或多种译文。不同语文的国籍应该用一个斜杠字符（/）分开，后面至少跟一个空格。

3.7 出生地的表示方式

可自由选择是否纳入出生地。如果纳入出生地，可用镇、市、城郊和/或国家加以表示。

如果纳入镇、市或城郊的名称且签发国的本国语文或签发机构的工作语文并不使用拉丁字符，则镇、市或城郊的名称应以本国/工作语文的形式出现，同时还应：

- 转写成拉丁字符；或
- 翻译成使该出生地可能更普遍为国际社会所熟知的一种或多种语文（至少其中一种必须是英文、法文或西班牙文）。

用不同语文表示的镇、市或城郊的名称应该用一个斜杠字符 (/) 分开，后面至少跟一个空格。

当签发国的本国语文或签发机构的工作语文使用拉丁字母，但是如果将镇、市或城郊的名称翻译成另外一种语文或多种语文（特别是英文、法文或西班牙文）更为国际社会所熟悉时，用本国/工作语文写入的镇、市或城郊的名称应配有一种或多种译文。用不同语文表示的镇、市或城郊的名称应该用一个斜杠字符 (/) 分开，后面至少跟一个空格。

如果纳入国家名称，其三字代码须采用第 5 节概述的三字代码，出生国没有代码的情况除外，在这种情况下，应该用全称写入出生国的名称，并且还应遵守上文所规定的针对镇、市和城郊的翻译和转写要求。

注：在选择纳入或省略出生地时，旅行证件的签发国或签发机构应该考虑到当前与该 国或该领土相关的任何政治敏感因素，以及其是否是被其他国家的签证发放机关所认可的国家或领土。

3.8 日期的表示方式

机读旅行证件视读区中的日期应根据公历按照下列方式登记：

日

日应该用两位数表示，即 1 到 9 日的前面应加零。可以在该数字后面的月份之前留一个空格，也可以不留空格，直接跟上月份。

月

月份可以用签发国的本国语文或签发机构的工作语文完整打印出来，或使用缩写，最多使用四个字符位置。

如果签发国的本国语文或签发机构的工作语文不是英文、法文或西班牙文，则月份后面应跟一个斜杠字符 (/) 和用下表中所列的三种语文中的一种标明月份或月份的缩写（最多占用四个字符位置）。

如果签发国的本国语文或签发机构的工作语文是英文、法文或西班牙文，签发国或签发机构还可以在斜杠字符 (/) 后面加上其他两种语文中的一种（如表 1 所示）。

根据签发国或签发机构自行作出的选择，月份也可以用数字形式打印出来，特别是如果这样做可能便于那些不使用公历的国家使用机读旅行证件。这时，日期将写成 DDnMMnYY 或者 DDnMMnYYYY，其中 n = 一个空格或一个圆点。

表1 月份的英文、法文和西班牙文缩写

月份	英文	法文	西班牙文
一月	JAN	JAN	ENE
二月	FEB	FÉV	FEB
三月	MAR	MARS	MAR
四月	APR	AVR	ABR
五月	MAY	MAI	MAYO
六月	JUN	JUIN	JUN
七月	JUL	JUIL	JUL
八月	AUG	AOÛT	AGO
九月	SEP	SEPT	SEPT
十月	OCT	OCT	OCT
十一月	NOV	NOV	NOV
十二月	DEC	DÉC	DIC

年

年份用最后两位数或四位数表示，前面留有一个空格，或者也可以直接跟在月份后面，不留空格。两种格式都是可以接受的。

当月份用数字形式表示时，签发国或签发机构可以使用两位或四位数表示年份，并用一个空格或圆点将月份和年份分开。

注：鼓励各国对于所有日期格式使用四位数代表年份。

示例：

用意大利文签发且将月份译成法文的机读旅行证件数据页上的 1942 年 7 月 12 日可表示为：

12nLUGn/JUILn1942

其中，n = 一个空格，即 12 LUG / JUIL 1942

或

12nLUGn/JUILn42

其中，n = 一个空格，即 12 LUG/JUIL 42

或

12 July 1942 或 12 July 42（仅使用英文）

或

12JUIL1942 或 12JUIL42（使用法文缩写）

或

12JUL 1942 或 12JUL 42（使用英文或西班牙文缩写）

或

12 07 42 或 12.07.42（使用数字格式）

或

12 07 1942 或 12.07.1942（使用数字格式，其中年份为四位数）。

未知的出生日期在出生日期一无所知的情况下，该数据元素应该用签发国或签发机构所使用的出生日期格式来表示，用 Xs 表示日期的未知元素（数字和/或字母）。

举例：

XXnXXnXX

XXnXXnXXXX

XXnXXXnXX

其中 n = 一个空格或（如果使用数字格式）一个圆点。

如果出生日期中只有一部分不清楚，则仅该部分（日、月、年）应按照签发国或签发机构所使用的日期格式用 Xs 来表示。

3.9 显示的持有人识别特征

Doc 9303 号文件明确了各类机读旅行证件必须在视读区中显示的强制性的和选择性的持有人识别特征，如人脸图像、签名或常用标记和/或单枚指纹，以及这些识别特征的位置、尺寸和缩放比例。

3.9.1 显示的人脸图像

为确保与脸部识别系统的兼容性，标准照的拍摄必须遵守[ISO/IEC 39794-5]概述的相关规范。

所显示的人脸图像，不论是以纸质或数字格式提供，都必须：

- 以数字形式打印在机读旅行证件上；
- 展现机读旅行证件合法持有人的真实面貌；和
- 不得通过数字手段修改或美化图像，从而使人物面貌发生任何变化。

签发国或签发机构须采取必要措施，确保所显示的标准照是防伪造和防替换的。

3.9.1.1 为提交标准照打印的图像

实际标准照必须产生对人物准确可识别的描绘。原始拍摄的图像质量至少应与纸质照片的最低可接受的质量相当（分辨率相当于每毫米6至8线对）。为了在数字复制中达到这种相当的图像质量，须格外注意图像的拍摄、处理、数字化、压缩和打印技术以及制作标准照所使用的过程。打印过程须保持原始图像的宽高比。

注：许多签发国对证件申请使用打印/重新扫描的程序。这种做法是可以接受的。但是，应按照下面以及[ISO/IEC 39794-5]所载的指南和要求，谨慎确保质量。如果考虑采用新设计的申请过程，则应考虑尽可能将数字提交作为首选技术。

打印分辨率 打印过程应产生能够准确呈现皱纹和痣等面部细节对比的平滑图像。不论人物肤色深浅都应准确打印，并限制明显的热点或阴影消失。应渲染出平滑的面部细节，没有明显的色调反差或轮廓线。

饱和度和颜色 除了小面积镜面（像镜子一样）反射可能造成的眩光或闪光外，所打印的图像只有一小部分应当是白色或黑色饱和。除背景部分外，使用发光度、完全饱和度数值为0的像素须低于0.1%，完全饱和的255像素值的数量须低于0.1%。

不应将背景或主体服装的任何部分打印成全白，并且深色阴影区域的细节应清晰可见。

打印的照片必须是具有平衡色彩灰度的彩色图像。可以假定已经对拍摄装置（数码相机或扫描仪）进行了正确的白平衡。

纸张特性和标准照尺寸 照片须打印在照片专用纸上。以下是此类纸张的示例（也可以接受具有类似特性的其他技术）：

- 即时摄影标准光泽，
- 热升华摄影亚光，
- 卤化银摄影亚光，或

- 干实验室摄影喷墨基础标准光泽。

相纸须具有低粗糙度、非结构化的表面（无颗粒或丝印效果）。所提交标准照的最小宽度应为 35 毫米。眼间距（IED）应至少为 10 毫米。

仍依赖提交打印标准照的新制定的申请过程，应考虑使用较大尺寸的照片，例如 7 厘米乘 10 厘米。加大照片尺寸可降低过程链中出现质量损失的风险。但是，切换到较大尺寸的照片将需要考虑过程的影响。

波纹或可见点图案 所打印照片的数字化，可能会引发诸如波纹之类的人为影像，并且某些打印过程可能会加剧此类人为影像的产生。当使用证件扫描仪以每轴每厘米 120 像素（每英寸 300 像素）的空间采样率扫描其打印件时，所采用的打印过程应能够进行准确的面部识别。

如果打印的照片是通过定期半色调过程进行的，则对照片的扫描几乎总会出现波纹图案。因此，诸如喷墨打印机和激光打印机等等那些原本采用半色调来模拟连续色调的打印机，应使用非定期（或抖动）的半色调方法。此外，打印过程不应产生肉眼可见的点状图案。

注：向负责照片质量评估的人员提供透明模板通常十分有益。模板将显示头像大小和旋转（摇摆）的限制，并且叠加在照片上时，可以帮助确定打印的照片是否符合要求。此类工具的示例可参见[ISO/IEC 39794-5]。

3.9.1.2 所提交标准照的扫描

提交的标准照须符合 3.9.1.1 节和[ISO/IEC 39794-5]概述的相关规范。

所提交标准照的特性 所提交标准照的尺寸应该为 45.0 毫米×35.0 毫米（1.77 英寸×1.38 英寸）。这一尺寸可提供适当的分辨率，便于将照片缩放至机读旅行证件上使用的所需尺寸，同时具有进行人脸识别的适当分辨率。

申请过程中不得使用多次扫描/打印步骤。如果标准照已经打印提交并随后进行扫描，则所有其余制作步骤均须数字化。

如[ISO/IEC 39794-5]所述，提交的标准照须在申请前六个月内拍摄。不应接受拍摄时间超过三个月的标准照。签发人应考虑使用具有数字图像编码的元数据，以确保照片的最新时效。

如果提交的是打印标准照，则应要求提供拍摄日期的证据。这可以是照片背面打印的制作日期，也可以是摄影室注明日期的发票。如果标准照是相片卡的一部分（例如：一张包含 2x2 张图像的 10x15 打印件），则应提供完整的相片卡。

提交的标准照须整洁、没有弯曲、没有划痕、没有折叠并且没有损坏。打印的标准照上不得有墨迹或皱褶。

如果向签发机关提供数字版的标准照，则必须遵守签发机关规定的要求。

像素计数和调制变换功能（MTF） 最终扫描图像必须具有[ISO/IEC 39794-5]规定的像素数。对于扫描仪，调制变换功能 20 应在扫描仪上出现 4、7 周期每毫米（cy/mm）或更高分辨率。两个轴上的扫描仪调制变换功能应当相同。利用内置硬件或基于软件图像锐化的图像美化，通常不应用于增强调制变换功能。

示例：

在根据[ISO 12233]使用调制变换功能测试方法时，如果数码相机原始图像调制变换功能 20 以大约 80% 或更高的奈奎斯特频率出现，则可以保持图像的光学特性。在脸部照片中应该可以看到 2 至 3 毫米的雀斑/痣。对图像进行测量时，尺子是应手的基准标记。

调制变换功能分析应使用 ISO 12233 中的适当目标进行。信息示例可参见[ISO/IEC 39794-5]。

示例：

所打印的具有 10 mm 眼间距的典型图像应以至少 300 像素的采样率进行扫描。

调制变换功能将受到纸质照片尺寸及其分辨率（细节的精细度）的限制。为了从扫描图像中获得更高的分辨率，签发人应考虑提高对打印标准照的尺寸要求。

在采集过程中须格外小心，以避免在任何方向上进行任何形式的图像尺寸拉伸。

最终图像的宽高比由签发人的申请过程予以界定，典型数值为 7:9。必要的修改须通过裁剪进行，不得拉伸。

颜色、清晰度和饱和度 扫描的标准照必须与提交的标准照具有相同的颜色。在经过颜色校正的显示装置上以及日光条件下观看时，须做到人眼无法检测到标准照与扫描结果之间的差异。标准照须具有自然展现肤色的适当亮度和对比度。

每种颜色量化等级的数量应至少为 256 个等级，每个像素应具有三种颜色。扫描的图像须符合[ISO/IEC 39794-5]概述的颜色要求。

由于红—绿—蓝（RGB）色彩空间及其衍生空间本质上取决于设备装置，因此扫描仪的输出必须转换为[ISO/IEC 39794-5]概述的定义明确、与装置无关的色彩空间之一。

如果采用每种颜色八位量化，则当大量像素数值处于量化极限，即：在 0 或 255 级的水平时便会出现饱和。可接受的面部扫描图像不应在面部区域有大量饱和像素。

扫描标准照须居中、清晰且对焦明确、无阴影。扫描标准照不得有可见压缩的人为影像。

3.9.1.3 制作机读旅行证件采用的图像打印

数据页上打印的标准照须来自与机读旅行证件中以电子形式存储的图像相同的数字图像来源。但是，由于打印技术的影响以及对标准照和数据页适用的多种安保特征，图像可能不完全与之相同。可能的偏差示例包括打印机分辨率、删除打印标准照的背景、图像美化、灰度内容抖动或打印中出现扭索现象。

注：在机读旅行证件上或其中印制标准照，应考虑所使用的不同材料和技术的属性。打印技术本身可能会在打印的标准照中引入具体的特性。

数字复制须产生对主体的准确可识别的体现。为了在证件数据页实现此类图像质量，须特别注意处理、压缩和打印技术以及印制标准照所采用的过程。所打印的标准照具备的具体特性取决于打印技术的类别。

机读旅行证件上的主要的打印图像，可以是灰色调的也可以是彩色的。

所有面部的打印过程均应产生能够准确呈现精细面部细节的平滑图像，例如：在打印图像区域的面部任何位置的直径小至 2 毫米明暗对比的皱纹、明暗对比的痣和明暗对比的疤痕。当用裸眼在 0.3 m 远查看时，这些细节必须能被观察到。

不论是浅肤色还是深肤色主体，所有肤色均应打印准确，并且没有明显的热点或阴影消失。应当呈现平滑的面部细节，没有明显的色调反差或轮廓线。

尺寸 标准照尺寸应符合[ISO/IEC 39794-5]中概述的规格。必要的修改须通过裁剪进行，不得拉伸。如果图像中删除了背景，则可能无法确定打印图像的正确宽度或高度。在这种情况下，如果打印图像的眼间距与口眼距离（EM）之比与标准照的比率相同，则认为高宽比得到了保持。

色调范围 在将打印图像与证件持有人进行比较时，打印图像的色调范围不得干扰对人体识别非常重要的面部细节。

波纹或可见点图案 打印图像中的波纹或光点图案应最小化。将打印的图像与证件持有人进行比较时，所打印图像中的任何此类图案均不得干扰对人体识别非常重要的面部细节。

标准照在机读旅行证件中的位置以及与安保打印的共存 打印的标准照须居中放在 V 区内，头顶（不考虑任何头发）尽量靠近机读旅行证件的上部边缘。人脸图像中的头顶到下颚部分须占 V 区的最长尺寸的 70%到 80%，并保持持有人脸部头顶到下颚以及双耳之间的长宽比。这个 70%到 80%的要求可能意味着需对照片进行裁切，以致看不到全部头发。

如果出现数字打印的复制件，则必须与 V 区内的背景安保处理共存，即背景安保印刷不得影响所显示标准照的正常观看，反之亦然，但仍然为所显示的标准照提供保护。

与机读旅行证件最后加工处理共存 显示的标准照须与最后加工处理共存，即最后加工处理不得影响所显示的标准照的正常观看，反之亦然。

边界 不得使用边界或边框为数字打印的复制件添加轮廓。

3.9.1.4 遵循国际标准

照片应遵循[ISO/IEC 39794-5]中所列出的相关规定。

3.9.2 显示的签名或常用标记

签名或常用标记显示在 IV 区，其可接受性由签发国或签发机构自行决定。所显示的签名或常用标记应是在机读旅行证件上创建的原件、原件数字打印的复制件，或者在 Doc 9303 号文件第 4 部分至第 7 部分所确定的、与制作不同类型的机读旅行证件具体相关的规范允许的情况下，可将签名或常用标记放在可牢固贴在机读旅行证件上的基材上。签发国或签发机构应采取必要的措施，确保所显示的签名或常用标记是防伪造和防替换的。所显示的签名或常用标记应满足下列要求。

定位 所显示的签名或常用标记在显示时应使 A-尺寸与机读旅行证件的参考（较长）边缘相平行，如图 2 所示。

尺寸 所显示的签名或常用标记的尺寸应使人的眼睛可以辨认（即尺寸缩小比例不超过百分之五十），而且保持原始签名或常用标记的长宽比（A-尺寸比 B-尺寸）。

使用数字打印时复制件的缩放比例 如果所显示的签名或常用标记被放大或缩小，应保持原始签名或常用标记的长宽比（A-尺寸比 B-尺寸）。

使用数字打印时复制件的裁切 签发国或签发机构应采取措施避免裁切或将裁切减少到最低程度。

颜色 所显示的签名或常用标记的颜色应与背景形成明显的对比。

边界 不允许使用边界或边框为所显示的签名或常用标记添加轮廓。

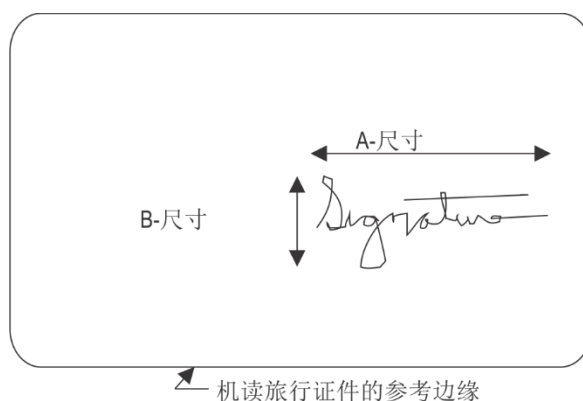


图 2 所显示的签名或常用标记的定位

3.9.3 单枚指纹的显示

如果签发国或签发机构要求显示单枚指纹，它应是持有人在机读旅行证件基材上创建的原件，或者更可能是原件的数字打印复制件。签发国或签发机构应采取必要的措施，确保单枚指纹是防伪造和防替换的。单枚指纹应满足下列要求。

定位 所显示的单枚指纹的 A-尺寸（宽）应与机读旅行证件的参考边缘相平行，如图 3 所示。手指的上部应是单枚指纹距离机读旅行证件参考边缘最远的那个部分。（见 Doc 9303 号文件第 6 部分的图 10 和图 12。）

尺寸 所显示的单枚指纹应是原始指纹的 1 比 1 的复制件（A-尺寸对 B-尺寸）。

使用数字打印时复制件的缩放比例 不允许缩放单枚指纹的尺寸。

使用数字打印时复制件的裁切 签发国或签发机构应采取措施，避免裁切或将裁切减少到最低程度。

颜色 所显示的单枚指纹的颜色应与背景形成明显的对比。

边界 不允许使用边界或边框为所显示的单枚指纹添加轮廓。

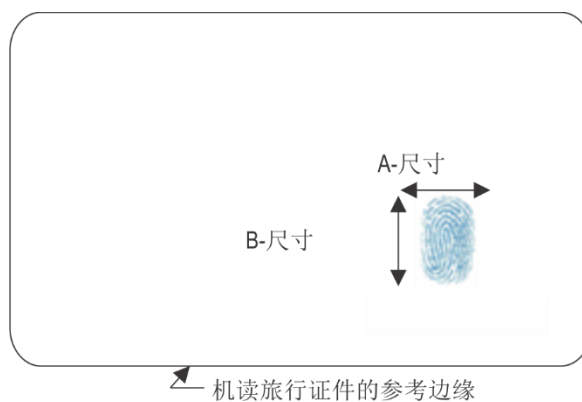


图 3 所显示的单枚指纹的定位

4. 机读区 (MRZ)

4.1 机读区的用途

根据 Doc 9303 号文件制作的机读旅行证件包括一个机读区, 以便于查验旅行证件, 并减少旅行过程中行政程序所占用的时间。另外, 机读区还为视读区的信息提供验证, 并可用来为数据库查询提供搜索字符。它也可用来采集到港和离港的登记数据, 或简单地显示数据库中的现有记录。

机读区以针对各类机读旅行证件的标准格式提供一套基本的数据元素, 所有的接受国, 不管其国家的文字或习惯如何, 都可以使用这一格式。

机读区的数据经格式化后, 全世界具有标准能力的机器都可以阅读。必须强调, 机读区是为依照机读旅行证件的国际标准在国际上使用的数据而预留的。机读区的数据表示法不同于视读区的数据表示法。

4.2 机读区的特性

机读区的数据必须是视读和机读的。数据的表现方式必须符合一个共同的标准, 这样所有按照 Doc 9303 号文件配置的机读设备都能识别每个字符, 并采用标准协议 (例如美国标准信息交换代码) 进行通信。这种协议要与接受国规定的技术基础设施和处理要求相兼容。

为满足这些要求, 规定了 OCR-B 字体作为在机读区存储数据的表现方式。在此界定的机读区被认可为进行全球信息交换必不可少的机读技术, 因此在所有类型的机读旅行证件中都必须采用。

4.3 机读区的限制条件

机读区仅可使用的字符是各国都可以使用的一套公用字符 (图 4)。国家特有字符一般只出现在应用这些字符的国家计算机处理系统中, 在全球范围内并不使用这些字符。因此, 它们不得出现在机读区内。

在机读区内, 不允许使用变音符。尽管变音符对于区分姓名是有用的, 但在机读区内使用会干扰机读设备, 从而降低数据库搜索准确度, 减慢旅客的通关速度。

在机读区内, 数据可用字符位置的数量有限, 并因机读旅行证件的类型而异。插入到机读区的数据元素的长度, 必须符合 Doc 9303 号文件第 4 部分至第 7 部分的数据元素目录中所规定各个域的尺寸。

在某些情况下, 机读区内的名字所采取的表现形式不一定与视读区的相同。在视读区内, 可以使用非拉丁字符和国家特有字符来更准确地代表签发国或签发机构的文字性数据。此类字符不得用在机读区内。

4.4 打印规格

机读数据应使用 OCR-B 字体、尺寸 1 和固定笔画宽度的字符打印，固定的宽度间距为 2.54 毫米（0.1 英寸），即水平打印密度为每 25.4 毫米（1.0 英寸）10 个字符。打印字符应限制在图 4 规定的字符范围内。



图 4 用于机读旅行证件的选自 [ISO 1073-2] 的 OCR-B 字符子集

注：仅作为说明之用 — 所显示字符要比实际尺寸大。

4.5 机读要求和有效阅读区

有效阅读区 按照最大尺寸的机读旅行证件规定了一个固定尺寸的阅读区（大小为 17.0 毫米×118.0 毫米（0.67 英寸×4.65 英寸）的有效阅读区（ERZ）），以便用一个机读者就可以阅读所有尺寸的机读旅行证件。有效阅读区的位置如图 5 所示。提供有效阅读区不是为了增加与制作不同类型机读旅行证件相关的第 4、5、6 和 7 部分中规定的打印位置的公差。有效阅读区是为了兼顾到由于人工放置机读签证（MRV）和打开机读护照内页阅读时出现的册页鼓风效应而造成的不同情况。有效阅读区还兼顾到阅读有两行或三行机读数据的机读官方旅行证件的情况。

为了应对使用诸如复印机等设备对旅行证件安全所带来的威胁，在机读区允许设置安全特征。根据 [ISO 1831] 的规定，这种安全特征不得影响在 B900 范围内准确地阅读 OCR 字符。虽然根据 4.2 中的规定，OCR 字符必须明显可见，以确保所有机读旅行证件，包括那些在机读区带有安全特征的证件，都能被成功地阅读，但在机读区内的 OCR 字符应至少能在光谱的近红外区部分（即 [ISO 1831] 中规定的 B900 谱带）被机器阅读。

注：所示的有效阅读区（ERZ）的尺寸是以可用于所有机读旅行证件的标准化有效阅读区为根据的，以便能够使用单一机器阅读器。

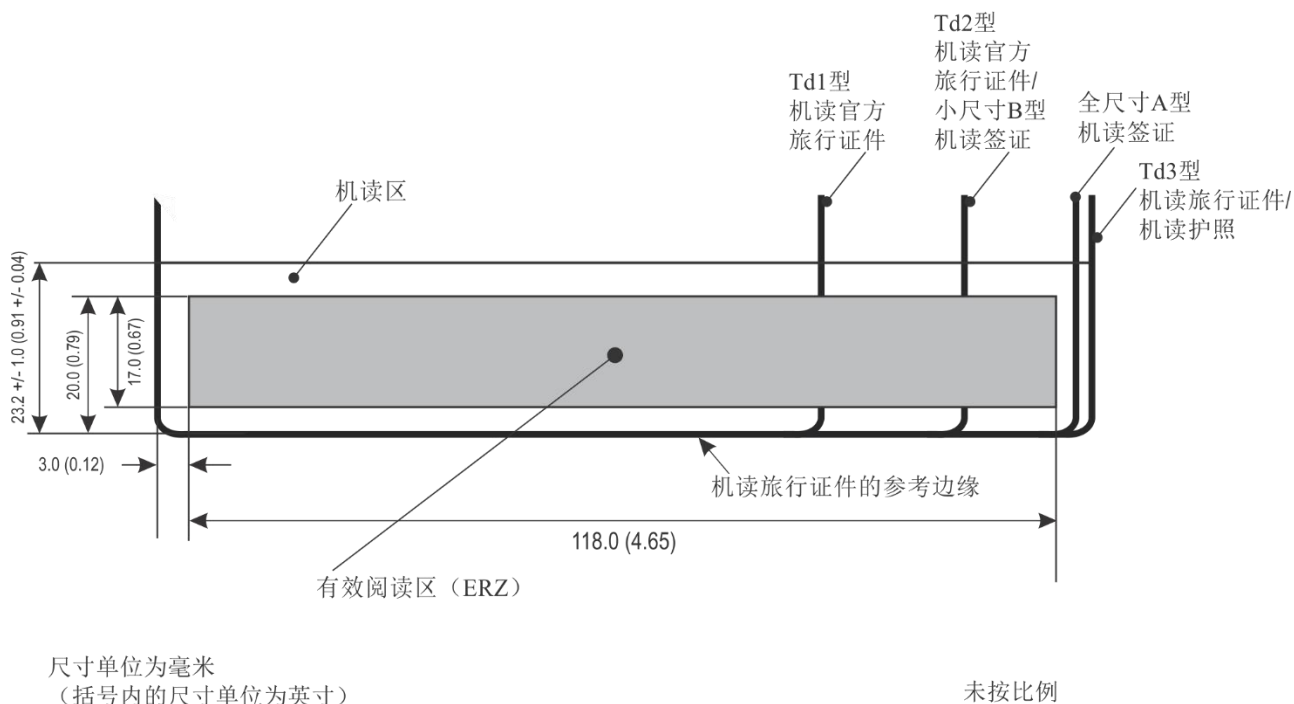


图5 机读旅行证件有效阅读区示意图

4.6 持有人姓名的书写惯例

为实现全球互操作性，机读区内的主要和次要标识符应使用图4所示的大写的OCR-B字符打印，不加变音符，并且应符合可用的字符位置的数量。因此，机读区内姓名的表达方式不同于视读区内姓名的表达方式。签发国或签发机构应仅使用允许的OCR-B字符对本国字符进行转写和/或按照Doc 9303号文件针对特定尺寸旅行证件的第4部分至第7部分中所规定的要求进行截取。第6节载有拉丁语、西里尔语和阿拉伯语系最常用字符的转写表。

使用拉丁字符转写（如果适用的话）的主要标识符应按照Doc 9303号文件针对特定尺寸旅行证件的第4部分至第7部分中所规定的要求写入机读区。主要标识符后面应跟着两个填充符（<<）。使用拉丁字符转写（如果适用的话）的次要标识符应从紧跟着这两个填充符的字符位置开始写入。

如果主要或次要标识符有一个以上的姓名成分，成分之间应该用一个填充符（<）分开。

填充符（<）应紧跟着最后的次要标识符插入（或如果姓名只有主要标识符的话，紧跟着主要标识符插入），一直插到机读行的最后一个字符位置。

姓名域的字符位置数量有限，并因机读旅行证件的类型而异。如果采用上述程序写入相关机读行的主要和次要标识符超过这个可用的字符位置，应按照Doc 9303号文件针对特定尺寸旅行证件的第4部分至第7部分中所规定的程序进行截取。在其他所有情况下，则不得截取姓名。

姓名截取的例子见 Doc 9303 号文件针对特定尺寸旅行证件的第 4 部分至第 7 部分。

前缀和后缀，包括头衔、专业和学术资格、荣誉、奖励和世袭地位（如博士、爵士、小、老、二世和三世），不得包括在机读区内，除非签发国认为它们是姓名的法定的一部分。在这些情况下，前缀或后缀应作为次要标识符的成分出现。

数字字符不得在机读区的姓名域中使用。

在机读区内不允许有标点符号字符。当它们作为姓名的一部分出现时，应对它们做如下处理：

撇号：

应省掉。在机读区，由撇号分开的姓名成分应结合到一起，不得插入填充符替代它的位置。

例子 视读区： D'ARTAGNAN
 机读区： DARTAGNAN

连字符：

如果在两个名字成分之间有连字符，在机读区应该用一个填充符（<）代替它。（即用连字符连接的名字应表示为不同的成分）。

例子 视读区： MARIE-ELISE
 机读区： MARIE<ELISE

逗号：

如果在视读区使用逗号将主要标识符和次要标识符分开，应在机读区省掉逗号。在机读区，主要标识符和次要标识符之间应使用两个填充符（<<）分开。

例子 视读区： ERIKSSON, ANNA MARIA
 机读区： ERIKSSON<<ANNA<MARIA

另外，如果在视读区使用逗号将两个姓名成分分开，该逗号在机读区应显示为一个填充符（<）。

例子 视读区： ANNA, MARIA
 机读区： ANNA<MARIA

其他标点符号字符：

所有其他标点符号字符均应从机读区略去（即在机读区它们的位置不得插入填充符）。

4.7 签发国或签发机构和持有人国籍的表示方式

在机读区，应使用第5节所列的三字代码填写签发国或签发机构和国籍域。

4.8 日期的表示方式

在机读旅行证件机读区内的日期应以六位数表示，其中包括年份的最后两位数（YY），后面紧跟着月份的两位数（MM）和日的两位数（DD）。结构如下：YYMMDD。

根据这个格式，1942年7月12日将被表示为：420712。

如果出生日期的全部或部分不清楚，相关字符位置应该用填充符（<）填充。

4.9 机读区内的校验数位

校验数位由从一系列其他数字计算得出的一个数位组成。机读区内的校验数位是根据机读区内特定的数字数据元素计算而来的。通过校验数位，阅读器可以验证机读区内的数据是否得到正确的解读。

在机读旅行证件中，采用了一种特殊的校验数位计算方法。校验数位应在模数10的基础上进行计算，不断地重复731731……加权，具体方法如下：

步骤1 从左到右，用相应顺序位置上的加权数乘相关数字数据元素的每一位数。

步骤2 将每次乘法运算的乘积相加。

步骤3 将得出的和除以10（模数）。

步骤4 余数即为校验数位。

对于数字没有占满所有可用字符位置的数据元素，应使用符号<填充空白位置，并且应确定它为零值，以便计算校验数位。

在对含有字母字符的数据元素进行校验数位计算时，字符A至Z的赋值应依次为10到35，具体如下：

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35

校验数位计算中所包含的数据元素以及每类证件中的检验数位位置载于Doc 9303号文件针对特定尺寸旅行证件的第4部分至第7部分。校验数位计算的示例见提供了详实信息的本部分附录A。

4.10 机读区的特征

除在此另有说明外，机读旅行证件应在以下事项方面符合[ISO 1831]：

- 所采用的基材的光学特性；
- 构成 OCR 字符的图像图案的光学和尺寸特性；和
- 与 OCR 字符在基材上的位置相关的基本要求。

机读数据应在固定长度域中从左到右分两行排列（上和下），但 TD1 型旅行证件除外，该类证件分三行（上、中、下）排列。应按照 Doc 9303 号文件针对特定尺寸旅行证件的第 4 部分至第 7 部分中的数据结构表所规定的顺序排列数据，并且数据在证件上的位置应遵循这些部分所规定的要求。在每个域中登记数据时应从左边字符位置开始。

如果登记的数据没有占满相关域的所有字符位置，应使用符号<填充未占位置。

4.11 机读区的质量规范

除在此另有规定外，打印质量一般应符合[ISO 1831] Range X。除另有说明外，下文规定的全部质量规范都应符合本部分第 2 节的要求，并且应适用于后期加工后的机读旅行证件以及已经放进护照或其他旅行证件的签证。

基材质量 [ISO 1831]，4.3 一直到 4.3.2，应只用作参考。

基材的不透明性 在后期加工前后使用和测定的基材（对于签证，应为在将其放置进护照或其他旅行证件之前）应至少在[ISO 1831]的 4.4.1 和 4.4.3 所规定的中等不透明性的定义范围内。

基材光泽 光泽度没有规定。

荧光性 在紫外线光的照射下，基材在可见光谱中的反射系数不得显示出眼睛看得到的荧光，但那种出于安全原因而设定的可预测的荧光除外。

替代基材 不管是哪种基材，都应遵循前文所述的质量规范。

光谱带 OCR 打印应清晰可见，而且应使用黑色（[ISO 1831]中规定的 B425 到 B680）。OCR 打印还应在[ISO 1831]规定的 B900 谱带下吸光（即近红外）。如果需透过可能加在证件表面的保护性材料对字符进行机读，这一特征必须顺利通过测试。

打印对比度（PCS） 后期加工后，根据[ISO 1831]规定测定的最小打印对比度（PCS/min）在 B900 光谱带下应为 $PCS/min \geq 0.6$ 。

字符笔画宽度 后期加工后的笔画宽度应符合[ISO 1831]的 5.3.1 中关于 Range X 的规定。

对比度变化率（CVR） 后期加工后，对比度变化率应如[ISO 1831]中所示的 Range X 数值，即 $CVR < 1.50$ 。

污点和外部标记 [ISO 1831]的 5.4.4.6 和 5.4.5.12 应适用于阅读表面（另见[ISO 1831]附录 B 的 B.6 和附录 C 的 C.5.10）。

失效 在[ISO 1831]的 5.4.5.9 中定义的“d”值在阅读表面上应等于 0.4。

行分离 见 Doc 9303 号文件针对特定尺寸旅行证件的第 4 部分至第 7 部分。

行间隔 见 Doc 9303 号文件针对特定尺寸旅行证件的第 4 部分至第 7 部分。

机读区行的歪斜 机读区行的实际歪斜率和机读区字符的歪斜率从参考边缘测量不得超过 3 度，同时机读区的歪斜或字符偏差不得导致机读区的行或其中任何部分超出 Doc 9303 号文件针对特定尺寸旅行证件的第 4 部分至第 7 部分所规定的打印区。

5. 代表国籍、出生地、签发国/机构所在地和其他情况的代码

A 部分 — 字母代码

二字母和三字母代码，须从[ISO 3166-1]维护机构 — [ISO 3166/MA]获得，该机构是国际标准化组织在国家代码方面的联系单位。这些代码在[ISO 3166-1]中定期更新，并且载于（<https://www.iso.org/iso-3166-country-codes.html>）。

下表概述了[ISO 3166-1]中未包含的代码，例如其他国家和机构的扩展名或其他例外情况：

实体（简称）	二字代码	三字代码	实体（简称）	二字代码	三字代码
英国海外领土公民		GBD	受英国保护人士		GBP
英国国民（海外）		GBN	德国	DE	D
英国海外公民		GBO	科索沃 ¹	KS	RKS
英籍人士		GBS			

B 部分 — ISO 3166/MA 预留的其他代码

欧盟（EU）	EU	EUE
--------	----	-----

1. 虽然[ISO 3166-1]未注明，但代码 KS 和 RKS 均在使用。

C 部分 — 联合国旅行证件上使用的代码

联合国组织或该组织的一名官员	UN	UNO
联合国专门机构或该机构的一名官员	UN	UNA
持有联合国驻科索沃临时管理特派团（UNMIK）签发的旅行证件的科索沃居民		UNK

D 部分 — 其他签发机关的代码

非洲开发银行（ADB）		XBA
非洲进出口银行（AFREXIM）		XIM
加勒比共同体或其一名使者（CARICOM）		XCC
欧洲理事会		XCE
东部和南部非洲共同市场（COMESA）		XCO
西非国家经济共同体（ECOWAS）		XEC
国际刑事警察组织（INTERPOL）		XPO
东加勒比国家组织（OECS）		XES
地中海议会大会（PAM）		XMP
马耳他最高军教团或其一名使者		XOM
南部非洲发展共同体		XDC

E 部分 — 没有确定国籍的人的代码

1954 年《无国籍人地位公约》第 1 条定义的无国籍人		XXA
经 1967 年议定书修订的 1951 年《难民地位公约》第 1 条定义的难民		XXB
不在上述 XXB 代码定义之内的难民		XXC
未确定国籍的人，签发国认为不管该人的地位是什么，没有必要为其确定上述任何 XXA、XXB 或 XXC 代码。该类可以包括既不是无国籍		XXX

人，也不是难民，而是在签发国合法居住的不明国籍的人。		
----------------------------	--	--

F 部分 — [ISO 3166]中的弃用代码（为向后兼容而引用）

荷属安的列斯群岛	AN	ANT
中立区	NT	NTZ

G 部分 — 示例证件中使用的代码

为了制定识别示例证件的标准化方法，建议将示例证件的证件持有人国籍设置为“乌托邦”。

乌托邦	UT	UTO
-----	----	-----

H 部分 — 国际民航组织使用的代码

以下代码在[ISO 3166]中未注明，仅由国际民航组织在对总列表进行数字签名时使用。

国际民用航空组织（ICAO）	IA	IAO
----------------	----	-----

6. 建议各国使用的转写方法

以下表格载有拉丁语、西里尔语和阿拉伯语系最常用的国家字符。

A. 基于拉丁语的多国字符的转写

统一码	国家字符	说明	建议的转写
00C0	À	A 加沉音符号	A
00C1	Á	A 加尖重音符号	A
00C2	Â	A 加抑扬音调符号	A
00C3	Ã	A 加鼻音化符号	A
00C4	Ä	A 加分音符号	AE 或 A
00C5	Å	A 上加小圆圈	AA 或 A
00C6	Æ	AE 连字	AE
00C7	Ç	C 加变音符号	C
00C8	È	E 加沉音符号	E
00C9	É	E 加尖重音符号	E
00CA	Ê	E 加抑扬音调符号	E
00CB	Ë	E 加分音符号	E
00CC	Ì	I 加沉音符号	I
00CD	Í	I 加尖重音符号	I
00CE	Î	I 加抑扬音调符号	I
00CF	Ï	I 加分音符号	I
00D0	Ð	Eth 齿间擦音	D
00D1	Ñ	N 加鼻音化符号	N 或 NXX
00D2	Ò	O 加顿音符号	O
00D3	Ó	O 加尖重音符号	O
00D4	Ô	O 加抑扬音调符号	O
00D5	Õ	O 加鼻音化符号	O
00D6	Ö	O 加分音符号	OE 或 O
00D8	Ø	O 加斜线符号	OE
00D9	Ù	U 加沉音符号	U
00DA	Ú	U 加尖重音符号	U
00DB	Û	U 加抑扬音调符号	U
00DC	Ü	U 加分音符号	UE 或 UXX 或 U
00DD	Ý	Y 加尖重音符号	Y
00DE	Þ	北欧字母 (冰岛)	TH

统一码	国家字符	说明	建议的转写
0100	Ā	A 加长音符号	A
0102	Ǻ	A 加短音符号	A
0104	Ą	A 加元音鼻音化符号	A
0106	Ć	C 加尖重音符号	C
0108	Ĉ	C 加抑扬音调符号	C
010A	Ċ	C 加顿音符号	C
010C	Č	C 加倒折符号	C
010E	Ď	D 加倒折符号	D
0110	Đ	D 加横线符号	D
0112	Ē	E 加长音符号	E
0114	Ĕ	E 加短音符号	E
0116	Ė	E 加顿音符号	E
0118	Ę	E 加元音鼻音化符号	E
011A	Ě	E 加倒折符号	E
011C	Ĝ	G 加抑扬音调符号	G
011E	Ǻ	G 加短音符号	G
0120	Ġ	G 加顿音符号	G
0122	Ģ	G 加变音符号	G
0124	Ĥ	H 加抑扬音调符号	H
0126	Ħ	H 加横线符号	H
0128	İ	I 加鼻音化符号	I
012A	Ī	I 加长音符号	I
012C	Ĭ	I 加短音符号	I
012E	Į	I 加元音鼻音化符号	I
0130	ı	I 加顿音符号	I
0131	ı	I 无点 (土耳其)	I
0132	Ĳ	IJ 连字	IJ
0134	Ĵ	J 加抑扬音调符号	J
0136	Ɔ	K 加变音符号	K
0139	Ĺ	L 加尖重音符号	L
013B	Ł	L 加变音符号	L
013D	Ł	L 加倒折符号	L
013F	Ł	L 加间隔符号	L
0141	Ł	L 加斜线符号	L
0143	Ń	N 加尖重音符号	N
0145	Ņ	N 加变音符号	N
0147	Ň	N 加倒折符号	N

统一码	国家字符	说明	建议的转写
014A	Ð	Eng	N
014C	Ō	O 加长音符号	O
014E	Ö	O 加短音符号	O
0150	Ö	O 加双尖重音符号	O
0152	Œ	OE 连字	OE
0154	Ŕ	R 加尖重音符号	R
0156	Ŗ	R 加变音符号	R
0158	Ř	R 加倒折符号	R
015A	Ś	S 加尖重音符号	S
015C	Ŝ	S 加抑扬音调符号	S
015E	Ş	S 加变音符号	S
0160	Š	S 加倒折符号	S
0162	Ţ	T 加变音符号	T
0164	Ť	T 加倒折符号	T
0166	Ƨ	T 加斜线符号	T
0168	Ũ	U 加鼻音化符号	U
016A	Ū	U 加长音符号	U
016C	Ů	U 加短音符号	U
016E	Ů	U 加小圆圈	U
0170	Ů	U 加双尖重音符号	U
0172	Ū	U 加元音鼻音化符号	U
0174	Ŵ	W 加抑扬音调符号	W
0176	Ŷ	Y 加抑扬音调符号	Y
0178	ÿ	Y 加分音符号	Y
0179	Ž	Z 加尖重音符号	Z
017B	Ž	Z 加顿音符号	Z
017D	Ž	Z 加倒折符号	Z
1E9E	ß	双 s (德国)	SS

B. 西里尔字符的转写

统一码	国家字符	建议的转写
0401	Ё	E (例外: 白俄罗斯语= IO)
0402	Ѣ	D
0404	Є	IE (例外: 乌克兰语首字, 则=YE)
0405	С	DZ
0406	І	I
0407	Ї	I (例外: 乌克兰语首字, 则=YI)
0408	Ј	J
0409	Љ	LJ
040A	Њ	NJ
040C	Ќ	K (例外: 在塞尔维亚语和前南斯拉夫马其顿共和国所讲语言= KJ)
040E	Ў	U
040F	Ѹ	DZ (例外: 在塞尔维亚语和前南斯拉夫马其顿共和国所讲语言= DJ)
0410	А	A
0411	Б	B
0412	В	V
0413	Г	G (例外: 白俄罗斯语、塞尔维亚语和乌克兰语=H)
0414	Д	D
0415	Е	E
0416	Ж	ZH (例外: 塞尔维亚语=Z)
0417	З	Z
0418	И	I (例外: 乌克兰语=Y)
0419	Й	I (例外: 乌克兰语首字, 则=Y)
041A	К	K
041B	Л	L
041C	М	M
041D	Н	N
041E	О	O
041F	П	P
0420	Р	R
0421	С	S
0422	Т	T
0423	У	U
0424	Ф	F

统一码	国家字符	建议的转写
0425	X	KH (例外: 塞尔维亚语和前南斯拉夫马其顿共和国语=H)
0426	Ц	TS (例外: 塞尔维亚语和前南斯拉夫马其顿共和国语=C)
0427	Ч	CH (例外: 塞尔维亚语=C)
0428	Ш	SH (例外: 塞尔维亚语=S)
0429	Щ	SHCH (例外: 保加利亚语=SHT)
042A	Ъ	IE
042B	Ы	Y
042D	Э	E
042E	Ю	IU (例外: 乌克兰语首字, 则=YU)
042F	Я	IA (例外: 乌克兰语首字, 则=YA)
046A	Ѹ	U
0474	V	Y
0490	Г	G
		G (例外: 在前南斯拉夫马其顿共和国所讲语言=GJ)
0492	F	
04BA	h	C

C. 阿拉伯字母的转写

统一码	阿拉伯字母	名称	机读区
0621	ء	hamza	XE
0622	آ	alef 上加 madda	XAA
0623	أ	alef 上加 hamza	XAE
0624	ؤ	waw 上加 hamza	U
0625	إ	alef 下加 hamza	I
0626	ئ	yeh 上加 hamza	XI
0627	ا	alef	A
0628	ب	beh	B

统一码	阿拉伯字母	名称	机读区
0629	ة	tehmarbuta	XTA/XAH ²
062A	ت	teh	T
062B	ث	theh	XTH
062C	ج	jeem	J
062D	ح	hah	XH
062E	خ	khah	XKH
062F	د	dal	D
0630	ذ	thal	XDH
0631	ر	reh	R
0632	ز	zain	Z
0633	س	seen	S
0634	ش	sheen	XSH
0635	ص	sad	XSS
0636	ض	dad	XDZ
0637	ط	tah	XTT
0638	ظ	zah	XZZ
0639	ع	ain	E
063A	غ	ghain	G
0640	-	tatwheel	(没有编码)
0641	ف	feh	F
0642	ق	qaf	Q
0643	ك	kaf	K
0644	ل	lam	L
0645	م	meem	M
0646	ن	noon	N

2. 一般使用 XTA, 当 teh marbuta 出现在名字成分的末尾时除外, 这时应使用 XAH。

统一码	阿拉伯字母	名称	机读区
0647	ه	heh	H
0648	و	waw	W
0649	ى	alefmaksura	XAY
064A	ي	yeh	Y
064B	◌َ	fathatan	(没有编码)
064C	◌ِ	dammatan	(没有编码)
064D	◌ِ	kasratan	(没有编码)
064E	◌َ	fatha	(没有编码)
064F	◌ُ	damma	(没有编码)
0650	◌ِ	kasra	(没有编码)
0651	◌◌◌	shadda	[双次] ³
0652	◌◌◌	sukun	(没有编码)
0670	◌◌◌	用作上标的 alef	(没有编码)
0671	أ	alefwasla	XXA
0679	ٹ	tteh	XXT
067C	ت◌◌	teh 加小圆圈	XRT
067E	پ◌◌	peh	P
0681	ح◌◌	hah 上加 hamza	XKE
0685	ح◌◌◌	hah 上加 3 个点	XXH
0686	چ	tch eh	XC
0688	ڈ	ddal	XXD
0689	د◌◌	dal 加小圆圈	XDR
0691	ڑ	rreh	XXR
0693	ر◌◌	reh 加小圆圈	XRR
0696	ر◌◌◌	reh 上下各加小圆圈	XXRX

3. Shadda 表示双次的意义：重复一个拉丁字符或序列，如 عبااا 即为 EBBAS，فضة 即为 FXDZXDZXA H。

统一码	阿拉伯字母	名称	机读区
0698	ژ	jeh	XJ
069A	ڙ	seen 上下各加小圆圈	XXS
069C	ڙڙڙ	seen 上下各加3个点	(没有编码)
06A2	ڦ	Feh 将点移至下方	(没有编码)
06A7	ڦ	qaf 上加1个点	(没有编码)
06A8	ڦ	qaf 上加3个点	(没有编码)
06A9	ڪ	keheh	XKK
06AB	ڪ	kaf 加小圆圈	XXK
06AD	ڦڦ	ng	XNG
06AF	گ	gaf	XGG
06BA	ن	noonghunna	XNN
06BC	ن	noon 加小圆圈	XXN
06BE	ھ	hehdoachashmee	XDO
06C0	ه	heh 上加 yeh	XYH
06C1	ه	heh goal	XXG
06C2	ه	heh goal 上加 hamza	XGE
06C3	ه	tehmarbuta goal	XTG
06CC	ی	farsiyeh	XYA
06CD	ی	yeh 加小尾巴	XXY
06D0	ی	yeh	Y
06D2	ے	yeh barree	XYB
06D3	ے	yeh barree 上加 hamza	XBE

7. 偏差

随着世界各国不断采用机读旅行证件，越发复杂的情况和偏差的增多使得有必要通过一种标准化机制报告一国偏离标准或通常做法的情况。偏差指机读旅行证件中包含未完全符合国际民航组织规范以及 ISO 和 RFC 规范性标准的元素。偏差一般出现在国家签名认证机构（CSCA）或证件签名者证书（DSCs）中。但是，有些国家还指出在其机读旅行证件内也存在与逻辑数据结构和机读区域相关的问题。本节的目的是详述签发国可借以公布其所存在偏差的机制。

虽然旅行证件可能会包含一些偏差，但在边境管理系统中它们还是可用的。而对于有效的证件，可以使用数年。因此，信赖方应确定自己的处理公布偏差的程序。

7.1 运作经验

很长一段时间以来，管理偏差的唯一做法是签发国通过外交手段提供一般性的建议。本节讲述了涉及大量机读旅行证件的各种偏差，可通过报告这些偏差协助边境管理部门确定旅行证件是有效的、伪造的或是被替换的。运作中出现的一些错误包括机读区、逻辑数据结构和公钥基础设施方面的偏差。

尽管机读区已经使用了很多年，但仍存在一些众所周知的错误，近期的例子包括：

- 机读区的出生日期与视读区页面上的出生日期对不上。
- 机读区的国籍错误地报告成出生国，而非国籍。

在大多数情况下，机读区未符合规范的旅行证件将会被签发国召回。由于证件的签发和随后的补发之间存在一定的时间间隔，旅行者可能不得不使用有偏差的机读旅行证件。在这段时间里，公布偏差可缓解可能对旅行者造成的问题。⁴

对于逻辑数据结构和公钥基础设施方面的偏差，有些可以长时间不被发现，这是由于许多国家还未按照 Doc 9303 号的规定实行被动和主动认证。但是，大力提倡签发国公布偏差，以协助国际社会处理在采用机读旅行证件时遇到的技术问题。⁵

7.2 偏差列表做法

本节所述的做法旨在为签发国提供一种用于公布和分发旅行证件偏差列表的标准化方法。这一做法所依据的是在编制国家签名认证机构证书总表过程中所建立的原则（见 Doc 9303 号文件第 12 部分），根据这些原则，有关每个国家不合规情况的经签署的偏差列表将通过国际民航组织的公钥目录予以提供，或者由签发国通过网站或简易目录访问协议服务器予以提供。公钥目录旨在为分发与偏差管理相关的信息提供支持。

-
4. 本节不涉及仅影响单个证件或少量电子机读旅行证件的不合规的情况。签发国可自行决定是否要召回并补发单个证件。
 5. 如果存在与公钥基础设施证书相关的安全问题，正确的做法是如 Doc 9303 号文件第 12 部分所述的那样将其撤销。本节不宜提供更多的指导。

偏差分为四个具体方面：

- 密钥和证书；
- 逻辑数据结构（LDS）；
- 机读区（MRZ）；
- 芯片。

对于每一个类别，偏差将仅从一个层级上描述，如：

类别：	LDS
错误	DG2

更多的信息将通过各国提供的运行参数和/或在报告框架内设置的一个使通知国可添加任何所需说明性文字的自由文本域来提供。通知国可在自由的文本域内添加可获取更多信息的链接。对于证书方面的错误，签发机构可选择发放新的证书，但这并不是强制要求。

是否将不合规情况通知信赖方由签发国自己决定。在决定是否编制偏差列表时，各国应该考虑到，随着旅行者自助边检越来越普及，如果不能将与不合规的旅行证件相关的信息通知出去，可能会给旅行者造成延误和不便，而这将对签发国和整个边检过程产生不良影响。

偏差列表是报告涉及数以千计而非几个或几百个旅行证件的偏差的一种方式。对于少量的不合规的旅行证件，各国宜直接处理。

7.3 方法

7.3.1 偏差元素

机读旅行证件的组成部分包括纸片和无线射频辨识芯片等元素，每个元素都通过安全特征得到某种保护，这些安全特征可被界定，因而可在旅行证件的有效期内通过查验系统加以测试。在物理旅行证件上采用的安全特征既有显性的，也有隐性的。本节仅考虑机读区、逻辑数据结构和公钥基础设施内的偏差元素。

机读区是机读旅行证件数据页上的一个尺寸固定的区域，包含强制性和选择性数据，这些数据使用便于通过光学字符识别方法进行机读的格式加以编排。Doc 9303 号文件提供了与机读区相关的规范，包括：

- 目的；
- 限制；
- 转写；和
- 机读区行的数据结构。

机读区与规范的符合性定期通过查验系统进行测试，做法是将机读区的数据与对应的视读区页上的数据进行对比以及重新计算机读区的校验数位。

机读旅行证件无线射频辨识芯片上存储的数据的真实性和完整性由被动认证技术加以保护。这一安全机制以数字签名和公钥基础设施（PKI）为基础。

Doc 9303 号文件第 10 部分对机读旅行证件的逻辑数据结构进行了定义。虽然没有特定的测试可用于确定是否符合规范，但是存储在逻辑数据结构内的数据有一部分是机读旅行证件机读区或视读区页上数据的子集。因此，适用于机读区和视读区页的测试同样可用于测试机读区和视读区的数字化数据。逻辑数据结构的真实性可通过由查验系统准确实施被动认证程序来确定，而主动认证则由芯片进行。简要描述如下：

被动认证（PA）以数字签名为依据，包含以下公钥基础设施成分：

1. **国家签名认证机构（CSCA）**：每个国家设立一个国家签名认证机构作为其在电子机读旅行证件方面的国家级信任单位。国家签名认证机构为一个或多个（国家）证件签名者颁发公钥证书。此外，国家签名认证机构还发布包含所有被撤销证书的证书撤销列表（CRLs）。⁶
2. **证件签名者（DS）**：证件签名者对存储在机读旅行证件上的数据进行数字签名；该签名被存储在每份证件的证件安全对象上。

主动认证（AA）：实施主动认证时，每个芯片都包含有自己的主动认证密钥对。私钥被存储在芯片的安全存储器中，而公钥被存储在逻辑数据结构的第 15 数据组中。

7.3.2 发布偏差列表

偏差列表不得由国家签名认证机构直接发布，而是应由国家签名认证机构授权一个偏差列表签名者（Doc 9303 号文件第 12 部分）负责汇编、签署和发布偏差列表。关于偏差列表规范，参见 Doc 9303-12 号文件。

发布偏差列表所应采取的程序应反映在负责发布的国家签名认证机构所公布的认证政策中。

7.3.3 接收偏差列表

各接收国制定其关于接收偏差列表的政策，并确定在查验证件过程中处理偏差的方式。一般而言，这类政策属于保密信息。

是否允许存在偏差的机读旅行证件继续使用将由接收国自行决定。

6. 由于证书撤销列表是一种安全报告机制，且会不断重新发布，所以没必要报告其中的缺陷，本节也不宜涉及这些内容。

7.3.4 偏差的类别

7.3.4.1 密钥和证书

证书和密钥方面的偏差仅限于下列情况：

问题	说明
证书	描述至域或域的扩展
密钥	描述至域或域的扩展
主动认证	仅描述至错误/问题

注：如果报告国决定签发新的证书，则不得将证书纳入偏差列表中，但可以通过自由文本域对其加以说明。

7.3.4.2 逻辑数据结构（LDS）

逻辑数据结构方面的偏差仅限于下列情况：

问题	说明
数据组通用信息	描述至编码错误
数据组	描述至数据组
安全对象数据	描述至具体问题（如证件签名者证书）

7.3.4.3 机读区（MRZ）

机读区方面的偏差仅限于下列情况：

问题	说明
与视读区的匹配	描述至域
校验数位	描述至相关校验数位
信息编码错误	描述至机读区域

7.3.5 偏差类型的定义

偏差类别和相应参数可能会不断扩充，Doc 9303 号文件将使其保持在最新状态。

每一偏差由一个偏差描述元素加以描述。偏差用一个客体标识符即偏差类型来标示，并可通过参数对其进行更加详细的说明。域描述可包含更多的信息，如偏差的性质为何无法通过主要偏差类型加以充分描述。

偏差类型	参数	描述
证书/密钥偏差		
标识符—偏差—证书或密钥 (id-Deviation-CertOrKey)	无	未包括在下列更为详细的偏差范围之内的与证书或密钥相关的一般性偏差。
标识符—偏差—证书或密钥—证件签名者的签名 (id-Deviation-CertOrKey-DSSignature)	无	证件签名者证书上的签名有误。
标识符—偏差—证书或密钥—证件签名者编码 (id-Deviation-CertOrKey-DSEncoding)	证书域 (CertField)	证件签名者证书包含错误编码。
标识符—偏差—证书或密钥—国家签名认证机构编码 (id-Deviation-CertOrKey-CSCAEncoding)	证书域 (CertField)	国家签名认证机构证书包含编码错误。
标识符—偏差—证书或密钥—主动认证密钥泄密 (id-Deviation-CertOrKey-AAKeyCompromised)	无	主动认证密钥可能泄密，不应信赖。
逻辑数据结构偏差		
标识符—偏差—逻辑数据结构 (id-Deviation-LDS)	无	未包括在下列更为详细的偏差范围之内的与逻辑数据结构相关的一般性偏差。
标识符—偏差—逻辑数据结构—数据组格式错误 (id-Deviation-LDS-DGMalformed)	数据组 (Datagroup)	某一数据组的标志长度值编码被破坏。
标识符—偏差—逻辑数据结构—数据组散列错误 (id-Deviation-LDS-DGHashWrong)	数据组 (Datagroup)	安全对象数据中某一数据组的散列值有误。
标识符—偏差—逻辑数据结构—安全对象数字签名错误 (id-Deviation-LDS-SODSignatureWrong)	无	安全对象数据中所包含的签名有误。
标识符—偏差—逻辑数据结构—数据组通用信息不一致 (id-Deviation-LDS-COMinconsistent)	无	数据组通用信息和安全对象数据不一致。

偏差类型	参数	描述
机读区偏差		
标识符—偏差—机读区 (id-Deviation-MRZ)	无	未包括在下列更为详细的偏差范围之内的与机读区相关的一般性偏差。
标识符—偏差—机读区—错误数据 (id-Deviation-MRZ-WrongData)	机读区域 (MRZField)	机读区的某个域包含错误数据（如与视读区不一致），但可用导出的基本访问控制密钥打开芯片。 若基本访问控制密钥无法使用，则应该在偏差列表中另外加入标识符—偏差—芯片。
标识符—偏差—机读区—错误校验数位 (id-Deviation-MRZ-WrongCheckDigit)	机读区域 (MRZField)	机读区某一域的校验数位计算错误。
芯片偏差		
标识符—偏差—芯片 (id-Deviation-Chip)	无	芯片无法使用，如基本访问控制密钥错误、天线坏了或其他物理缺陷。

Doc 9303-10 号文件、Doc 9303-11 号文件和 Doc 9303-12 号文件载有国际民航组织目标标识符规范。偏差对象标识符的列表如下：

-- 偏差列表基础对象标识符

```
id-icao-mrtd-security-DeviationList OBJECT IDENTIFIER ::= {id-icao-mrtd-security 7}
```

```
id-icao-mrtd-security-DeviationListSigningKey OBJECT IDENTIFIER ::= {id-icao-mrtd-security 8}
```

-- 偏差目标标识符与参数定义

```
id-Deviation-CertOrKey OBJECT IDENTIFIER ::= {id-icao-DeviationList 1}
```

```
id-Deviation-CertOrKey-DSSignature OBJECT IDENTIFIER ::= {id-Deviation-CertOrKey 1}
```

```
id-Deviation-CertOrKey-DSEncoding OBJECT IDENTIFIER ::= {id-Deviation-CertOrKey 2}
```

```
id-Deviation-CertOrKey-CSCAEncoding OBJECT IDENTIFIER ::= {id-Deviation-CertOrKey 3}
```

```
id-Deviation-CertOrKey-AAKeyCompromised OBJECT IDENTIFIER ::= {id-Deviation-CertOrKey 4}
```

```
id-Deviation-LDS OBJECT IDENTIFIER ::= {id-icao-DeviationList 2}
```

```
id-Deviation-LDS-DGMalformed OBJECT IDENTIFIER ::= {id-Deviation-LDS 1}
```

```
id-Deviation-LDS-SODSignatureWrong OBJECT IDENTIFIER ::= {id-Deviation-LDS
3}

id-Deviation-LDS-COMInconsistent OBJECT IDENTIFIER ::= {id-Deviation-LDS 4}

id-Deviation-MRZ OBJECT IDENTIFIER ::= {id-icao-DeviationList 3}

id-Deviation-MRZ-WrongData OBJECT IDENTIFIER ::= {id-Deviation-MRZ 1}

id-Deviation-MRZ-WrongCheckDigit OBJECT IDENTIFIER ::= {id-Deviation-MRZ 2}

id-Deviation-Chip OBJECT IDENTIFIER ::= {id-icao-DeviationList 4}

id-Deviation-NationalUse OBJECT IDENTIFIER ::= {id-icao-DeviationList 5}
```

7.3.6 识别存在偏差的证件

可通过几种不同的方法识别受到偏差影响的证件：

- 通过用于为这些证件签名的证件签名者证书；证件签名者可通过以下方式加以识别：
 - 签发人的唯一甄别名以及证书的序列号（签发人和序列号（*issuerAndSerialNumber*）），
 - 用作识别证件签名者唯一标志的主体密钥标识符（*subjectKeyIdentifier*），或
 - 证件签名者证书散列（证书散列（*certificateHash*））；所使用的散列函数与偏差列表签名中所用的函数是一样的。
- 通过一系列签发日期（签发日期开始时间、签发日期结束时间（*startIssuingDate, endIssuingDate*））
- 通过证件号列表（证件号列表（*listOfDocNumbers*））。

对于偏差列表的发布者和接收者来说，每一种方法都有其优缺点。这些优缺点包括：

- 通过证件签名者加以识别的做法使得查验系统只有在实施被动认证后才能识别偏差。另外，通过证件签名者加以识别的做法可能太过粗略，无法准确地仅识别出有缺陷的证件，即偏差影响的可能仅是某一特定证件签名者所签名的一部分证件。
- 签发日期并非机读区的一部分，并且电子逻辑数据结构中一般不会包含签发日期。因此，这种方法不适用于自动处理。此外，在有些签发国，签发日期可能并不是对护照进行个性化处理的实际日期，而是申请日期，因此可能不够准确，无法仅识别出受到影响的证件。
- 如果证件号没有按顺序发放，则很难整理出证件号列表。如果许多证件都受到同一缺陷的影响，证件号列表会迅速变大到难以汇编的程度。

建议提供尽可能多的关于受影响证件的识别信息。如果给出了多种识别方法，则必须同时满足相关条件才能将证件确认为受到影响。信赖国应自行决定使用偏差列表条目中给出的哪些识别方法来识别受影响证件。

7.4 公布

偏差列表可以通过国际民航组织的公钥目录和/或由签发机关通过网站或简易目录访问协议服务器进行公布。偏差列表的一级分发点是公钥目录。

	偏差列表
一级分发点	公钥目录
二级分发点	网站/简易目录访问协议服务器

7.4.1 由签发国公布

偏差列表可通过签发机关的网站或简易目录访问协议服务器进行公布。

7.4.2 在公钥目录上公布

公钥目录可作为偏差列表的中央存储库。

公布偏差列表的程序如下所述：

1. 作为《公钥目录接口规范》和《公钥目录程序手册》中规定的证书上传常用过程的一部分，将偏差列表发送至可写公钥目录。
2. 国际民航组织公钥目录办公室按照《公钥目录程序手册》的规定验证所上传的偏差列表的签名。
3. 将有效的偏差列表移至只读公钥目录。
4. 分发国将决定是否将偏差列表公开发布，或者仅提供给公钥目录的成员国。

7.4.3 信赖方

为核实偏差列表，信赖方需通过带外传输收到签发国相应的国家签名认证机构证书。信赖方可自行决定如何处理在签发国偏差列表中有相应记录的机读旅行证件。

8. 参考文献（规范性）

- [ISO 1073-2] ISO 1073-2:1976, 光学识别字母数字字符集第 2 部分：字符集 OCR-B — 印刷图像的形状和尺寸
- [ISO 1831] ISO 1831:1980, 光学字符识别印刷规范
- [ISO 1664-2] ISO 11664-2:2007(E)/CIE S014-2/E: 2006, 国际照明委员会比色法标准光源
- [ISO 12233] ISO 12233: 摄影 — 电子静态图片成像 - 分辨率和空间频率响应
- [ISO 3166-1] ISO 3166-1:2013, 国家及其地区的名称代码 — 第 1 部分：国家代码
- [ISO 3166/MA] ISO 3166 维护机构 https://www.iso.org/iso/home/standards/country_codes.htm
- [ISO/IEC 7810] ISO/IEC 7810:2003, 识别卡 — 物理特性
- [ISO/IEC 39794-5] ISO/IEC 19794-5: 2019, 可扩展的生物特征数据交换格式 — 第 5 部分：人脸图像数据
- [ISO/IEC 7501] ISO/IEC 7501 多部分标准：机读旅行证件
- [ISO/IEC 10918-1] ISO/IEC 10918-1:1994, 信息技术 — 连续色调静态图像数字压缩和编码：要求和指南
- [ISO/IEC 15444-1] ISO/IEC 15444-1:2004, 信息技术 — JPEG 2000 图像编码系统：核心编码系统
- [ISO/IEC 15948] ISO/IEC 15948:2004, 信息技术 — 计算机图形和图像处理 - 便携式网络图形（PNG）：功能规范
- [ISO/IEC 14496-2] ISO/IEC 14496-2 信息技术 — 视听对象编码第 2 部分：视觉[MPEG4]
- [IEC 61966-2-1] IEC 61966-2-1: 多媒体系统和设备 — 颜色测量和管理 — 第 2-1 部分：颜色管理 — 红绿蓝缺省颜色空间 — 标准红绿蓝色彩空间
- [IEC 61966-8] IEC 61966-8:2001, 多媒体系统和设备 — 色彩测量和管理 — 第 8 部分：多媒体彩色扫描仪
- [TR-03121-3] BSI: 技术指南 TR-03121-3: 公共部门应用程序生物识别技术第 3 部分：应用程序配置文件和功能模块第 1 卷：电子护照和身份验证方案第 3.0.1. 2013 版

[RFC 3852] 加密报文语法 — 2004年7月

[RFC 5280] D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley, W. Polk, “互联网 X.509 公钥基础设施证书和证书撤销列表（CRL）简介”，2008年5月。

第 3 部分附录 A 校验数位计算方法示例（资料性）

示例 1：校验数位应用在日期域上

以 1952 年 7 月 27 日为例，日期使用数字格式，其计算方法为：

日期：	5	2	0	7	2	7							
加权：	7	3	1	7	3	1							
步骤 1（乘法运算）	乘积：	35	6	0	49	6	7						
步骤 2（乘积之和）		35	+	6	+	0	+	49	+	6	+	7	= 103
步骤 3（除以模数）		$\frac{103}{10} = 10$ ，余数为 3											

步骤 4 校验数位为余数 3。因此，该日期和其校验数位应写作 **5207273**。

示例 2：校验数位应用在证件号码域上

以号码 AB2134 为例，在一个 9—字符的固定长度域中（例如护照号）对其编码，其计算方法为：

样本数据元素：	A	B	2	1	3	4	<	<	<
赋予的数字值：	10	11	2	1	3	4	0	0	0
加权：	7	3	1	7	3	1	7	3	1
步骤 1（乘法运算）乘积：	70	33	2	7	9	4	0	0	0
步骤 2（乘积之和）	$70 + 33 + 2 + 7 + 9 + 4 + 0 + 0 + 0 = 125$								
步骤 3（除以模数）	$\frac{125}{10} = 12$ ，余数为 5								

步骤 4 校验数位是余数 5。因此，该号码和其校验数位应写作 **AB2134<<<5**。

复合校验数位计算示例

所有机读旅行证件的复合校验数位计算方法都是一样的。但在进行计算时，不同类型证件的数字位置和数量则有所不同。为完整起见，在此各举一例。

样本数据元素:	7	1	2	2	<	<	<	<	<	<
赋予的数字值	7	1	2	2	0	0	0	0	0	0
加权:	1	7	3	1	7	3	1	7	3	1
步骤 1 (乘法运算) 乘积:	7	7	6	2	0					

样本数据元素:	<	<	<	<	<
赋予的数字值	0	0	0	0	0
加权:	7	3	1	7	3
步骤 1 (乘法运算) 乘积	0	0	0	0	0

步骤 2 (乘积之和)	91	+	6	+	3	+	7	+	12	+	5	+	56	+	27	+	0	+	49	+	
步骤 2 (乘积之和)	0	+	0	+	0	+	0	+	0	+	0	+	0	+	0	+	0	+	0	+	
步骤 2 (乘积之和)	0	+	0	+	0	+	0	+	0	+	9	+	4	+	0	+	21	+	1	+	
步骤 2 (乘积之和)	14	+	21	+	9	+	35	+	0	+	7	+	7	+	6	+	2	+	0	+	
步骤 2 (乘积之和)	0	+	0	+	0	+	0	+	0	+	0	+	0	+	0	+	0	+	0	+	
步骤 2 (乘积之和)	= 392																				
步骤 3 (除以模数)	$\frac{392}{10} = 39$, 余数为 2																				

步骤 4 校验数位是余数 2。因此，机读区中行数据和其复合校验数位可以写成如下形式：
3407127M9507122YT0<<<<<<<<<<<<<<2。

示例 5: TD2 机读旅行证件复合校验数位的计算方法

以下列机读区的下行数据为例，对复合校验数位进行编码，其计算方法为：
下机读行（字符位置 1-35）：

HA672242<6YT05802254M9601086<<<<<<<<

样本数据元素:	H	A	6	7	2	2	4	2	<	6
赋予的数字值:	17	10	6	7	2	2	4	2	0	6
加权:	7	3	1	7	3	1	7	3	1	7
步骤 1 (乘法运算) 乘积:	119	30	6	49	6	2	28	6	0	42

样本数据元素:	5	8	0	2	2	5	4	9	6	0
赋予的数字值:	5	8	0	2	2	5	4	9	6	0
加权:	3	1	7	3	1	7	3	1	7	3
步骤 1 (乘法运算) 乘积:	15	8	0	6	2	35	12	9	42	0

样本数据元素:	1	0	8	6	<	<	<	<	<	<
赋予的数字值:	1	0	8	6	0	0	0	0	0	0
加权:	1	7	3	1	7	3	1	7	3	1
步骤 1 (乘法运算) 乘积:	1	0	24	6	0	0	0	0	0	0

样本数据元素:	<
赋予的数字值:	0
加权:	7
步骤 1 (乘法运算) 乘积:	0

步骤 2 (乘积之和) 119 + 30 + 6 + 49 + 6 + 2 + 28 + 6 + 0 + 42 +

步骤 2 (乘积之和) 15 + 8 + 0 + 6 + 2 + 35 + 12 + 9 + 42 + 0 +

步骤 2 (乘积之和) 1 + 0 + 24 + 6 + 0 + 0 + 0 + 0 + 0 + 0 +

步骤 2 (乘积之和) 0

步骤 2 (乘积之和) = 448

步骤 3 (除以模数) $\frac{448}{10} = 44$, 余数为 8

步骤 4 校验数位为余数 8。因此，机读区下行数据和其复合校验数位可以写成如下形式：

HA672242<6YT05802254M9601086<<<<<<<8

第 3 部分附录 B 机读旅行证件中阿拉伯文字的 转写（资料性）

B.1 阿拉伯文字

阿拉伯文字是阿拉伯文所用的文字，阿拉伯文是摩洛哥、阿曼等 24 个国家的官方语文。还有其他语文也使用阿拉伯文字，特别是伊朗的波斯文、阿富汗的普什图文和达里文、巴基斯坦的乌尔都文，以及库尔德文、亚述文、豪萨文和维吾尔文等许多其他语文。过去，它曾用于中亚的一些语文，如塔吉克文和乌兹别克文。

阿拉伯文字是草写体，一个字母的形状是多变的，这取决于它是独立的一个词（孤立的）、位于词首（开头）、词中（中间）还是词尾（末尾）。例如，ب (beh) 这一字母在位于 بكر (Bakr) 的词首时形状变为 ب — 注意阿拉伯文的阅读顺序为从右往左，因此，第一个字母在右侧。这里，我们不关注字母的不同形状（字形），只关注由独立的字母形状表示的基本字母代码。

阿拉伯文和使用阿拉伯文字的其他语文在书写时通常仅使用辅音。因此，محمد (默罕默德) 这个名字的书写形式仅包含四个辅音，大概等于拉丁语的“Mhmd”。翻译人员可自己决定是否加上元音，以实现语音上的对应。阿拉伯文还可“附加母音符”，如果需要通过添加母音符号 (“harakat”) 改变发音的话。但是，harakat 通常省略。

标准的阿拉伯文字包括 32 个辅音、18 个元音和双元音以及 3 个其他符号。此外，在用于阿拉伯文以外的语文时，阿拉伯文字中还有 100 多个本国特殊字符，但有些已经过时，不再使用。

B.2 机读旅行证件中的阿拉伯文字

B.2.1 视读区

视读区有一个针对姓名的强制性域（参见 Doc 9303 号文件第 4 部分至第 7 部分中与各种尺寸的机读旅行证件相关的规范）。Doc 9303 号文件第 3 部分的第 3.1 段写到：

“当强制性数据元素用不使用拉丁字母的语文表示时，还应提供转写。”

因此，当姓名用阿拉伯文字进行书写时，应增加一个拉丁字符表示法。Doc 9303 号文件将这种表示法称为“转写”，但这通常是一种对等译音，更确切地说应该将其称为“音译”。

例如：

用阿拉伯文字表示的姓名¹： **ابو بكر محمد بن زكريا الرازي**

音译为拉丁字符是：**Abū Bakr Mohammed ibn Zakarīa al-Rāzi**

首先应注意的是，Doc 9303 号文件第 3 部分的第 3.2 段允许签发国决定是否在视读区中使用变音符号（如 **al-Rāzi** 中的 **ā**）。

其次，应注意的是这种特定的拉丁字符音译方式只是许多可能的音译方式中的一种。例如：以下不同表示方式即是对 **محمد** 适用的变体：

- | | | |
|---------------|--------------|---------------|
| 1. Muhammad | 2. Moohammad | 3. Moohamad |
| 4. Mohammad | 5. Mohamad | 6. Muhamad |
| 7. Muhamad | 8. Mohamed | 9. Mohammed |
| 10. Mohemmed | 11. Mohemmed | 12. Muhemmed |
| 13. Muhamed | 14. Muhammed | 15. Moohammed |
| 16. Mouhammed | | |

在有些国家，一般会将末尾的“d”替换为“t”，这样 **محمد** 就有 32 种不同的表示方式。

所用的音译方案取决于使用阿拉伯文字的国家的语文和地方口音（波斯文、普什图文和乌尔都文等非阿拉伯文语文也使用阿拉伯文字）、说拉丁文的人所使用的语文，以及所用的音译方案。

B.2.2 机读区

Doc 9303 号文件本部分第 4 节对机读区进行了描述。

机读区以适用于各类机读旅行证件的标准格式提供了一套基本的数据元素，所有的接收国，不管其国家的文字或习惯如何，都可以使用这一格式。机读区的数据经格式化后，全世界具有标准能力的机器都可以阅读，因此，机读区的数据表示法不同于视读区的数据表示法。国家特有字符一般只出现在使用这些字符的国家的计算机处理系统中，不能全球通用。因此，这种字符不得出现在机读区。

在机读护照中，机读区的姓名域包含 39 个字符位置，且仅可使用 OCR-B 的 A-Z 子集和填充符。因此，机读区中不得使用阿拉伯字符，必须使用“对应的”OCR-B 字符来表示。

在机读区将以阿拉伯文字表示的姓名转化为拉丁字符，而所使用的字符被限制为 OCR-B 的 A-Z 和填充符，这种做法是有问题的。此外，允许使用基于语音的音译将会引致的不确定性意味着数据库搜索可能变得徒劳无益。

例如，在上文所举的例子中：

用阿拉伯文字书写的姓名：**ابو بكر محمد بن زكريا الرازي**

1. Abū Bakr al-Rāzi 是大约 1100 年前波斯的一位伟大的科学家和医生。在波斯文 (Farsi) 中，他的名字的最后两个波斯字母通常拼为“yeh” (ی)، 但为了避免引起混乱，我们使用的是“yeh” (ي) 这一标准的阿拉伯文拼写方式。

在机读区将其音译成拉丁字符的一种方式:

ABU<BAKR<MOHAMMED<IBN<ZAKARIA<AL<RAZI

但是, 仅“Mohammed”这个名字机读区就可能有至少 32 种不同的表示方法。“Zakaria”可以写作“Zakariya”; “ibn”可以写作“bin”; “al”可以写作“el”。仅这些不同的表达方式就可以产生 256 种备选结果。

与此形成对比, 使用 Buckwalter 表(见下文)对例如上述名字محمد中的四个阿拉伯字符进行音译的结果是“mHmd”。在这种情况下, 每个阿拉伯字符对应一个拉丁字符。不考虑语音上的对应。

用 Buckwalter 表对上述姓名进行完整转写的结果是:

Abw<bAkr<mHmd<bn<zkryAY<AlrAzY

遗憾的是, Buckwalter 表会用到小写形式(a-z)和特殊字符(‘,|,>,\$,<,},*,_~,~), 因此, 不适合用于机读区(见 <http://www.qamus.org/transliteration.htm>)。

B.3 对视读区的建议

B.3.1 视读区内字符的音译

如上文所述, Doc 9303 号文件第 3 部分第 3.1 段要求在使用非拉丁文字的本国文字时在视读区加入“转写”。与此项要求相关的 Doc 9303 号文件第 3 部分第 3.4 段特别说明了姓名方面的要求。

“转写”和“音译”这两个词容易引起混淆。“转写”是指严格地对非拉丁文字进行逐字翻译。“音译”是一种不那么严格的表示方法, 一般以语音为基础(在说出姓名时所“发出的音”)。当然, 有时候用一种语文发出的音在另一种语文中并没有与之相对等的发音, 并且要取决于目标语文的情况, 例如, “ch”、“sh”和“th”在英文、法文和德文中的读法是不一样的。可以比较一下数学家和诗人 عمر خيام 的姓名的英文音译“Omar Khayyam”和德文音译“Omar Chajjam”。

有许多种“音译”方案:

- 德国标准化协会: DIN31635 (1982 年)
- 德国东方学会 (1936 年)
- 国际标准化组织: ISO/R233 (1961 年)、ISO233 (1984 年) [3]、ISO233-2 (1993 年)
- 英国标准协会: BS4280 (1968 年)
- 联合国地名专家组 (UNGEGN): UN (1972 年) [4]
- Qalam (1985 年)
- 美国图书馆协会—国会图书馆: ALA-LC (1997 年) [1]
- 伊斯兰百科全书 (新版): EI (1960 年) [2]

一些国家在出生或公民登记簿中用阿拉伯文字和拉丁文字记录其公民的姓名, 而拉丁文字姓名是经核准的对阿拉伯文字姓名的音译。这些国家似宜继续在视读区纳入经核准的拉丁文字音译字符。

建议

如上文所述，Doc 9303 号文件第 3 部分的第 3.1 段和第 3.4 段要求在视读区提供对等的拉丁字符，因此，签发国可自行决定是采取基于语音的音译，还是沿用机读区的转写做法（如下所述）。

B.3.2 音译方案

一些音译方案见下文所示：

统一码	阿拉伯字母	名称 ²	DIN 31635	ISO 233	UN GEGN	ALA-LC	EI	
0621	ء	hamza	'	'	'	'	'	
0622	آ	alef 上加 madda	'ā	'â	ā	ā	Ā	
0627	ا	alef	Ā	'				
0628	ب	beh	B	b	b	b	B	
0629	ة	teh marbuta	h,t	ṭ	h,t	h,t	a,at	
062A	ت	teh	T	t	t	t	<u>T</u>	
062B	ث	theh	<u>T</u>	ṭ	th	th	<u>Th</u>	
062C	ج	jeem	Ĝ	ğ	j	j	<u>Dj</u>	
062D	ح	hah	ḥ	ḥ	ḥ	ḥ	ḥ	
062E	خ	khah	ḫ	ḫ	kh	kh	<u>Kh</u>	
062F	د	dal	D	d	d	d	D	
0630	ذ	thal	<u>D</u>	<u>d</u>	dh	dh	<u>Dh</u>	
0631	ر	reh	R	r	r	r	R	
0632	ز	zain	Z	z	z	z	Z	
0633	س	seen	S	s	s	s	S	
0634	ش	sheen	Š	š	sh	sh	Sh	
0635	ص	sad	Ş	ş	ş	ş	ş	
0636	ض	dad	Ḍ	Ḍ	Ḍ	Ḍ	Ḍ	
0637	ط	tah	ṭ	ṭ	ṭ	ṭ	ṭ	
0638	ظ	zah	Ẓ	Ẓ	Ẓ	Ẓ	Ẓ	
0639	ع	ain	'	'	'	'	'	
063A	غ	ghain	Ĝ	ğ	gh	gh	<u>Gh</u>	
0640	-	tatwheel	[图形填充符，无音译]					
0641	ف	feh	F	f	f	f	F	
0642	ق	qaf	Q	q	q	q	q	
0643	ك	kaf	K	k	k	k	K	
0644	ل	lam	L	l	l	l	L	
0645	م	meem	M	m	m	m	M	
0646	ن	noon	N	n	n	n	N	
0647	ه	heh	H	h	h	h	H	
0648	و	waw	W	w	w	w	W	
0649	ى	alef maksura	Ā	ỳ	y	y	Ā	
064A	ي	yeh	Y	y	y	y	Y	

2. 统一码和 ISO/IEC 10646 中给出的名称。

统一码	阿拉伯字母	名称 ²	DIN 31635	ISO 233	UN GEGN	ALA-LC	EI
064B	◌َ	fathatan	An	á'	a	an	
064C	◌ِ	dammatan	Un	ú	u	un	
064D	◌ِ	kasratan	In	í	i	in	
064E	◌َ	fatha	A	a	a	a	A
064F	◌ُ	damma	u	u	u	u	U
0650	◌ِ	kasra	i	i	i	i	I
0651	◌◌◌	shadda	[双次]	—	[双次]	[双次]	[双次]
0652	◌◌◌	sukun		◌◌◌			
0670	◌ْ	用作上标的 alef	ā	ā	ā	ā	Ā

其他国家特有字符:

067E	پ	peh	p			p	P
0686	چ	tch eh	č			ch,zh	Č
0698	ژ	jeh	ž			zh	<u>Zh</u>
06A2 ⁴	ف	feh 点移至下方	f	f		q	
06A4	ف	veh	v			v	
06A5	ف	feh 下加 3 个点	v			v	
06A7 ⁴	ق	qaf 上加 1 个点	q	q		f	
06A8 ³	ق	qaf 上加 3 个点	v			v	
06AD	گ	ng	G			g	G
06AF	گ	gaf	G			g	G

B.4 机读区内字符的转写

B.4.1 机读区内欧洲语文的转写

考虑一下欧洲语文中的国家字符的情况很有必要。Doc 9303 号文件第 3 部分第 6 节 — “建议各国使用的转写方法” 包含了一个表格，题为：基于拉丁语的多国字符的转写。

大多数国家字符的变音符号都被省去，以便可以用在机读区中。有 9 个字符用特殊方式进行了处理，例如，“Ñ” 这一字符在机读区中可转写为“NXX”，从而保留其在数据库搜索中的唯一性和重要性。

例如：

用欧洲国家文字表示的姓名：**Térèsa CAÑON**

转写到机读区是 **CANXXON<<TERESA**

虽然机读区的表示方式看上去不太美观（并且可能招致抱怨），但这种做法是为了便于机读，使得原来的姓名可以通过数据库搜索等手段来查找。因此，机读区把这一姓名定为 **CAÑON**，从而与 **CANON** 区分开来。

B.4.2 统一码的使用

在内部处理上，计算机使用编码方案来表示不同语文的字符。常用的一种编码方案是统一码，它与 ISO/IEC 10646 标准几乎相同（下文的表格中使用了统一码的字符索引）。

阿拉伯文字所有字符的表示方式都可以在统一码中找到。统一码字符索引通常表示为一个四位的十六进制数（十六进制以 16 作为基数，用数字 0-9 和字母 A-F 表示 16 个可能的数字）。所有阿拉伯字符都在第 06 行，因此，06 就是数字的前两位（即 06XX）。

例如：

ابو بكر محمد بن زكريا الرازي

可用统一码将其编码为：

ابو	Alef (ا) - Beh (ب) - Waw (و) => 0627 + 0628 + 0648
بكر	Beh (ب) - Kaf (ك) - Reh (ر) => 0628 + 0643 + 0631
محمد	Meem (م) - Hah (ح) - Meem (م) - Dal (د) => 0645 + 062D + 0645 + 062F
بن	Beh (ب) - Noon (ن) => 0628 + 0646
زكريا	Zain (ز) - Kaf (ك) - Reh (ر) - Yeh (ي) - Alef (ا) => 0632 + 0643 + 0631 + 064A + 0627
الرازي	Alef (ا) - Lam (ل) - Reh (ر) - Alef (ا) - Zain (ز) - Yeh (ي) => 0627 + 0644 + 0631 + 0627 + 0632 + 064A

B.5 关于机读区的建议

B.5.1 影响机读区转写的因素

Doc 9303 号文件第 3 部分的第 4.1 段指出：“……机读区可为视读区的信息提供验证，并可用来为数据库查询提供搜索字符。”第 4.1 段还指出“机读区的数据经格式化后，全世界具有标准能力的机器都可以阅读”，以及“机读区的数据表示法不同于视读区的数据表示法。”但是，第 4.2 段指出，“机读区的数据必须是既可机读，又可视读的”。

这样做的目的是在机读区中将阿拉伯姓名转写为对应的拉丁字符，确保姓名只有一种可能的表示方式。这对于避免歧义以及使数据库和敏感人物列表搜索尽可能准确以确保识别结果的可靠性是有必要的。同时，机读区必须以尽可能容易辨认的方式来表示视读区所显示的姓名，以便使机读区的数据是可视读的，方便用于进行先进的旅客处理和类似目的。

B.5.2 现有的转写方案

现有几种转写方案：标准阿拉伯文技术转写系统（SATTS）、Buckwalter 和 ASMO 449。这些方案见下文所示：

统一码	阿拉伯字符	名称	SATTS	Buckwalter	ASMO 449
0621	ء	hamza	E	'	A
0622	آ	alef 上加 madda	(无)		B
0623	أ	alef 上加 hamza	(无)	>	C
0624	ؤ	waw 上加 hamza	(无)	&	D
0625	إ	alef 下加 hamza	(无)	<	E
0626	ئ	yeh 上加 hamza	(无)	}	F
0627	ا	alef	A	A	G
0628	ب	beh	B	b	H
0629	ة	teh marbuta	?	p	I
062A	ت	teh	T	t	J
062B	ث	theh	C	v	K
062C	ج	jeem	J	j	L
062D	ح	hah	H	H	M
062E	خ	khah	O	x	N
062F	د	dal	D	d	O
0630	ذ	thal	Z	*	P
0631	ر	reh	R	r	Q
0632	ز	zain	;	z	R
0633	س	seen	S	s	S
0634	ش	sheen	:	\$	T
0635	ص	sad	X	S	U
0636	ض	dad	V	D	V
0637	ط	tah	U	T	W
0638	ظ	zah	Y	Z	X
0639	ع	ain	"	E	Y
063A	غ	ghain	G	g	Z
0640	-	tatwheel	(无)	_	0x60
0641	ف	feh	F	f	A
0642	ق	qaf	Q	q	B
0643	ك	kaf	K	k	C
0644	ل	lam	L	l	D
0645	م	meem	M	m	E
0646	ن	noon	N	n	F
0647	ه	heh	?	h	G
0648	و	waw	W	w	H
0649	ى	alef maksura	(无)	Y	I
064A	ي	yeh	I	y	J

统一码	阿拉伯字符	名称	SATTS	Buckwalter	ASMO 449
064B	◌َ	fathatan	(无)	F	K
064C	◌ِ	dammatan	(无)	N	L
064D	◌ِ	kasratan	(无)	K	M
064E	◌َ	fatha	(无)	a	N
064F	◌ُ	damma	(无)	u	O
0650	◌ِ	kasra	(无)	i	P
0651	◌ِ	shadda	(无)	~	Q
0652	◌َ	sukun	(无)	o	R
0670	◌َ	用作上标的 alef	(无)	`	(无)

仔细查看表格可以看出，这些方案所用的拉丁字符超出了 A-Z 的范围，因此，从根本上说不适用于机读区中。

ASMO 449 方案在分配拉丁字符时比较随意，而 Buckwalter 方案将一些有对等发音的字符写的很相似。

标准阿拉伯文技术转写系统未对 heh (ه) 和 teh marbuta (ة) 或最后的 yeh (ي) 和 alif maksura (ي) 进行区分，并且这一系统无法对 alif madda (ا) 进行转写。

B.5.3 其他需考虑的方面

在使用所建议的转写方案时，不能不考虑到机读旅行证件的运作环境。特别是，机读区的姓名应该在外形和格式上尽可能地接近其他来源所提供的姓名。航空公司所使用的并通过旅客信息预报 (API) 方案提交给移民局的旅客姓名记录 (PNR) 就是一个例子。虽然机读区中的转写几乎总是无法与视读区中的音译（以及旅客姓名记录等基于语音的其他表示方法）完全一致，但此处所建议的方案试图使两个区中的姓名看起来是明显相似的。

为此，与基于拉丁语的多国字符转写表中的做法一样，使用了“X”这一字符作为“换码”字符，区别在于这里只用到一个“X”，并且用在其要修饰的字符之前，而非字符之后（如“XTH”对“NXX”）。在每个“X”之后跟上一个或两个字符来代表一个阿拉伯字母。这样使用“X”是可行的，因为现有的阿拉伯音译和转写方案中并不存在“X”。

[“X”在阿拉伯字符和拉丁字符转写中的不同用法不太可能造成混淆。为了可以顺利地进行逆向转写，必须明确源文字，且最好以签发国为基础。]

在一些转写条目中，在开头的“X”之后会加上另一个“X”：例如，alef 上加 madda ا 表示为“XAA”，alef wasla ا 表示为“XXA”。采用这一方法主要是为了避免引入会造成机读区不易肉眼阅读的其他字符。

这一做法的目的是让从现有系统中查看机读区原始数据的操作人员忽略任何“X”字符。所得到的姓名应该与其他来源所提供的姓名相类似。机读区原始数据还将缺少元音，而视读区中的音译和旅客姓名记录等其他来源一般会包含元音。但是，如果操作人员被告知元音未被包括在内，那么他们就应该认为机读区数据是基于语音的一种合理音译。

转写还将不包括“太阳字母”之前冠词的同化 (sandhi)，因为这本质上是一种语音现象，因此拼写可能与视读区的基于语音的音译对不上（例如，在视读区中，“AL-RAZI”可能写为“AR-RAZI”）。

“shadda”（用于表示将字母写两次的标志）意味着带有该标志的字符在机读区中将被重写一次（两次）。搜索算法应该考虑到“shadda”并不会总是存在。

B.5.4 标准阿拉伯文的转写建议方案

以 Buckwalter 转写表为基础，并考虑到各种音译方案（第 B.3.2 段）所列出的常见的基于语音的对等字符，形成了仅使用拉丁字符 A-Z 的一个转写建议方案。由于有用“X”表示不同字母的先例（第 B.5.3），这里将“X”这一字符用作一个“换码”字符，用于表示跟在“X”之后的一个或两个字符代表一个阿拉伯字母。

统一码	阿拉伯字母	名称	机读区	说明
0621	ء	hamza	XE	
0622	آ	alef 上加 madda	XAA	B.5.5.1
0623	أ	alef 上加 hamza	XAE	B.5.5.2
0624	و	waw 上加 hamza	U	B.5.5.3
0625	إ	alef 下加 hamza	I	B.5.5.4
0626	ئ	yeh 上加 hamza	XI	B.5.5.5
0627	ا	alef	A	
0628	ب	beh	B	
0629	ة	teh marbuta	XTA/XAH	B.5.5.6
062A	ت	teh	T	
062B	ث	theh	XTH	
062C	ج	jeem	J	
062D	ح	hah	XH	B.5.5.7
062E	خ	khah	XKH	
062F	د	dal	D	
0630	ذ	thal	XDH	
0631	ر	reh	R	
0632	ز	zain	Z	
0633	س	seen	S	
0634	ش	sheen	XSH	
0635	ص	sad	XSS	
0636	ض	dad	XDZ	
0637	ط	tah	XTT	
0638	ظ	zah	XZZ	
0639	ع	ain	E	
063A	غ	ghain	G	
0640	ـ	tatwheel	(注 1)	B.5.5.8
0641	ف	feh	F	
0642	ق	qaf	Q	
0643	ك	kaf	K	
0644	ل	lam	L	
0645	م	meem	M	
0646	ن	noon	N	
0647	ه	heh	H	B.5.5.7
0648	و	waw	W	
0649	ى	alef maksura	XAY	B.5.5.9
064A	ي	yeh	Y	

统一码	阿拉伯字母	名称	机读区	说明
064B	◌َ	fathatan	(注 1)	B.5.5.10
064C	◌ِ	dammatan	(注 1)	B.5.5.10
064D	◌ِ	kasratan	(注 1)	B.5.5.10
064E	◌َ	fatha	(注 1)	B.5.5.10
064F	◌ِ	damma	(注 1)	B.5.5.10
0650	◌ِ	kasra	(注 1)	B.5.5.10
0651	◌ِ	shadda	(双次)	B.5.5.11
0652	◌ِ	sukun	(注 1)	B.5.5.12
0670	◌ِ	用作上标的 alef	(注 1)	B.5.5.13
0671	أ	alef wasla	XXA	B.5.5.14

以下两个字母通常用在外国人名中：

06A4	ف	veh	V	
06A5	پ	feh 下加 3 个点	XF	

注 1: 没有编码

B.5.5 关于转写表的说明

B.5.5.1 Alef 上加 madda

ALA-LC 罗马化表[1]未对 alef 上加 madda (◌ِ) 进行表示。但是, Interpol [5]和 Dr Hoogland [6]都建议将其转写为 XAA。

B.5.5.2 Alef 上加 hamza

ALA-LC 罗马化表[1]未对 alef 上加 hamza (◌ِ) 进行表示。但是, Interpol [5]建议将其转写为 XAE。

B.5.5.3 Waw 上加 hamza

ALA-LC 罗马化表[1]未对 waw 上加 hamza (◌ِ) 进行表示。由于 waw 上加 hamza 通常音译为“U”，因此这里使用 U 进行表示。

B.5.5.4 Alef 下加 hamza

ALA-LC 罗马化表[1]未对 alef 下加 hamza (◌ِ) 进行表示。这里所使用的转写方式是 I, 因为该拉丁字母并未在别处使用, 而且 alef 下加 hamza 通常出现在 إبراهيم (Ibrahim) 等名字的头, 这种情况中的 alef 下加 hamza 通常音译为“I”。

B.5.5.5 Yeh 上加 hamza

ALA-LC 罗马化表[1]未对 yeh 上加 hamza (◌ِ) 进行表示。这里所使用的转写方式是 XI, 因为 yeh 上加 hamza 用于 فانز (Faiz) 等名字中, 而这种情况中的 yeh 上加 hamza 通常音译为“I”。

B.5.5.6 Teh marbuta

Teh marbuta (ة) 在 ALA-LC 罗马化表[1]中表示为 H 或 T 或 TAN，具体取决于相关语境。Dr Hoogland [6] 建议将其表示为 XTA。这里，teh marbuta 有两种不同的转写方式：一般情况下都使用 XTA，teh marbuta 出现在名字成分的末尾时除外，这时使用的是 XAH。这是因为女性姓名通常使用 teh marbuta 来修饰男性姓名，如 فاطمة (Fatimah)。搜索算法应该将这两种可能性纳入考虑。

B.5.5.7 Hah and heh

根据 Interpol [5]的建议，将 hah (ح) 和 heh (ه) 的转写方式进行了对调。Hah 现在表示为 XH，而 heh 表示为 H。

B.5.5.8 Tatwheel

Tatwheel (ـ) 是一个图形符号，未进行转写。

B.5.5.9 Alef maksura

根据 Dr Hoogland [6]的建议，现在将 alef maksura (ع) 转写为 XAY。其他字符转写为 XY_，这便使得之前的 XY 变得不协调。

B.5.5.10 短元音 fatha、damma、kasra、fathatan、dammatan 和 kasratan

选择性的短元音 (haracat) 一般不用于姓名中，因此未对其进行转写。

B.5.5.11 Shadda

Shadda (ّ) 表示其下面的辅音发两次音，因此它的转写方式是将对应的字符写两次。搜索算法应该注意，shaddah 是选择性的，字符有时候会写两次，有时则不会。

注意 الله (Allah) 这种特殊情况。

B.5.5.12 Sukun

Sukun (◌ْ) 表示未将元音包括在内，它是选择性的，因此未对其进行转写。

B.5.5.13 用作上标的 alef

未对用作上标的 (‘) (“元音-剑号-alef”) 进行转写。

B.5.5.14 Alef wasla

根据 Interpol [5]的建议，现在将 alef wasla (أ) 转写为 XXA。其他字符转写为 XA_，这便使得之前的 XA 变得不协调。Dr Hoogland [6]也建议将其转写为 XXA。

B.5.6 其他语文的转写建议方案

讲波斯文的国家有伊朗（波斯文）、阿富汗（达里文）、塔吉克斯坦和乌兹比克斯坦。

讲普什图文的国家有阿富汗和巴基斯坦西部。

讲乌尔都文的国家有巴基斯坦和印度。

统一码	阿拉伯字母	语文	名称	机读区
0679	ٹ	乌尔都文	tteh	XXT
067E	پ	波斯文、乌尔都文	peh	P
067C	ت	普什图文	teh 加小圆圈	XRT
0681	ح	普什图文	hah 上加 hamza	XKE
0685	ځ	普什图文	hah 上加 3 个点	XXH
0686	چ	波斯文、乌尔都文	tch eh	XC
0688	ڈ	乌尔都文	ddal	XXD
0689	د	普什图文	dal 加小圆圈	XDR
0691	ڑ	乌尔都文	rreh	XXR
0693	ر	普什图文	reh 加小圆圈	XRR
0696	ړ	普什图文	reh 上下各加一个点	XRX
0698	ژ	波斯文、乌尔都文	jeh	XJ
069A	ښ	普什图文	seen 上下各加一个点	XXS
06A9	ک	波斯文、乌尔都文	keheh	XKK
06AB	ک	普什图文	kaf 加小圆圈	XXK
06AD	ځ		ng	XNG
06AF	گ	波斯文、乌尔都文	gaf	XGG
06BA	ں	乌尔都文	noon ghunna	XNN
06BC	ن	普什图文	noon 加小圆圈	XXN
06BE	ھ	乌尔都文	heh doachashmee	XDO
06C0	ہ	乌尔都文	heh 上加 yeh	XYH
06C1	و	乌尔都文	heh goal	XXG
06C2	و	乌尔都文	heh goal 上加 hamza	XGE
06C3	ۋ	乌尔都文	teh marbuta goal	XTG
06CC	ی	波斯文、乌尔都文	farsi yeh	XYA ⁴
06CD	ی	普什图文	yeh 加小尾巴	XXY
06D0	ی	普什图文	yeh	Y ⁵
06D2	ے	乌尔都文	yeh barree	XYB
06D3	ۛ	乌尔都文	yeh barree 上加 hamza	XBE

4. “farsi yeh” (ی) 这一字母在功能上等同于标准的“yeh” (ی)，但在单独使用和用在末尾时在形状上等同于标准的“alef maksura” (آ)، 因此，可以转写为“Y”或“XAY”。数据库匹配算法应该将这一点纳入考虑。

5. “Pashto yeh” (ی) 这一字母在功能上等同于标准的“yeh” (ی)。

B.5.7 标准阿拉伯文的转写示例

上文的例子

ابو بكر محمد بن زكريا الرازي

在机读区中可编码为:

ابو	Alef (ا) - Beh (ب) - Waw (و) => ABW
بكر	Beh (ب) - Kaf (ك) - Reh (ر) => BKR
محمد	Meem (م) - Hah (ح) - Meem (م) - Dal (د) => MXHMD
بن	Beh (ب) - Noon (ن) => BN
زكريا	Zain (ز) - Kaf (ك) - Reh (ر) - Yeh (ي) - Alef (ا) => ZKRYA
الرازي	Alef (ا) - Lam (ل) - Reh (ر) - Alef (ا) - Zain (ز) - Yeh (ي) => ALRAZY

即 ABW<BKR<MXHMD<BN<ZKRYA<ALRAZY

这种转写方式的优点是:

1. 用阿拉伯文字表示的姓名总是可以转写为同一种拉丁表示方法。这意味着更有可能实现数据库匹配;
2. 该过程是可以逆向操作的—可以复原到用阿拉伯文字表示的姓名。

复原用阿拉伯文字表示的姓名:

ABW	A=Alef (ا) - B=Beh (ب) - W=Waw (و) => ابو
BKR	B=Beh (ب) - K=Kaf (ك) - R=Reh (ر) => بكر
MXHMD	M=Meem (م) - XH=Hah (ح) - M=Meem (م) - D=Dal (د) => محمد
BN	B=Beh (ب) - N=Noon (ن) => بن
ZKRYA	Z=Zain (ز) - K=Kaf (ك) - R=Reh (ر) - Y=Yeh (ي) - A=Alef (ا) => زكريا
ALRAZY	A=Alef (ا) - L=Lam (ل) - R=Reh (ر) - A=Alef (ا) - Z=Zain (ز) - Y=Yeh (ي) => الرازي

略去 *harakat* 和其他变音符号的理由是, 这些符号是选择性的, 基本上不会用到。因此, 这些符号的处理方式应该与欧洲国家字符上的变音符号 (如 é、è、ç) 相同, 这些变音符号是作为发音之用的。

同样地, *harakat* 是选择性的这一特点会对数据库的准确匹配性产生不利影响。

机读区	阿拉伯字母名称	阿拉伯字母	统一码
XAA	alef 上加 madda	آ	0622
XAE	alef 上加 hamza	أ	0623
XAH	teh marbuta (另见 XTA)	ة	0629
XAY	alef maksura	ى	0649
XBE	yeh barree 上加 hamza	ے	06D3
XC	tcheh (波斯文、乌尔都文)	چ	0686
XDH	thal	ذ	0630
XDO	heh doachashmee	ھ	06BE
XDR	dal 加小圆圈 (普什图文)	ډ	0689
XDZ	dad	ڌ	0636
XE	hamza	ء	0621
XF	feh 下加 3 个点	فِ	06A5
XGG	gaf (波斯文、乌尔都文)	گ	06AF
XGE	heh goal 上加 hamza (乌尔都文)	هَ	06C2
XH	hah	ح	062D
XI	yeh 上加 hamza	يَ	0626
XJ	jeh (乌尔都文)	جَ	0698
XKE	hah 上加 hamza (普什图文)	هَخ	0681
XKH	khah	خ	062E
XKK	keheh (波斯文、乌尔都文)	ک	06A9
XNN	noon ghunna (乌尔都文)	ن	06BA
XNG	ng	نگ	06AD
XRR	reh 加小圆圈 (普什图文)	رَ	0693
XRT	teh 加小圆圈	تَ	067C
XRX	reh 上下各加 1 个点 (普什图文)	رِ	0696
XSH	sheen	ش	0634
XSS	sad	س	0635
XTA	teh marbuta (see also XAH)	ة	0629
XTG	teh marbuta goal (乌尔都文)	هَ	06C3
XTH	theh	ث	062B
XTT	tah	ط	0637
XXA	alef wasla	آ	0671
XXD	ddal (乌尔都文)	ڌ	0688
XXG	heh goal (乌尔都文)	هَ	06C1
XXH	hah 上加 3 个点 (普什图文)	هَخ	0685
XXK	kaf 加小圆圈 (普什图文)	کَ	06AB
XXN	noon 加小圆圈 (普什图文)	نَ	06BC
XXR	rreh (乌尔都文)	رَ	0691
XXS	seen 上下各加 1 个点 (普什图文)	سِ	069A
XXT	tteh (乌尔都文)	تَ	0679
XXY	yeh 加小尾巴 (普什图文)	ی	06CD
XYA	farsi yeh (波斯文、乌尔都文)	ی	06CC
XYB	yeh barree (乌尔都文)	ے	06D2
XYH	heh 上加 yeh (乌尔都文)	هَ	06C0
XZZ	zah	ظ	0638

B.7 计算机程序

B.7.1 阿拉伯字符转写为机读区字符

这一用 Python 进行编写的程序是将阿拉伯字符（用统一码表示）转换为机读区格式的一个示例。

阿拉伯字符被包含在“Arabic souce.txt”这样一个文件中，而对应的机读区数据被写入“MRZ output.txt”这样的文件中。

```
*****
# # -*- coding: iso-8859-15 -*-

import unicodedata
import encodings.utf_8_sig
import codecs

# TRANSLITERATE
def Arabic_to_MRZ(unicode_string):
    transform = {0x20: '<', 0x21: 'XE', 0x22: 'XAA', 0x23: 'XAE', 0x24: 'U',
                 0x25: 'I', 0x26: 'XI', 0x27: 'A', 0x28: 'B', 0x29: 'XAH',
                 0x2A: 'T', 0x2B: 'XTH', 0x2C: 'J', 0x2D: 'XH', 0x2E: 'XKH',
                 0x2F: 'D', 0x30: 'XDH', 0x31: 'R', 0x32: 'Z', 0x33: 'S', 0x34: 'XSH',
                 0x35: 'XSS', 0x36: 'XDZ', 0x37: 'XTT', 0x38: 'XZZ', 0x39: 'E',
                 0x3A: 'G', 0x41: 'F', 0x42: 'Q', 0x43: 'K', 0x44: 'L',
                 0x45: 'M', 0x46: 'N', 0x47: 'H', 0x48: 'W', 0x49: 'XAY',
                 0x4A: 'Y', 0x71: 'XXA', 0x79: 'XXT', 0x7E: 'P', 0x7C: 'XRT',
                 0x81: 'XKE', 0x85: 'XXH', 0x86: 'XC', 0x88: 'XXD', 0x89: 'XDR',
                 0x91: 'XXR', 0x93: 'XRR', 0x96: 'XRX', 0x98: 'XJ', 0x9A: 'XXS',
                 0xA4: 'XV', 0xA5: 'XF', 0xA9: 'XKK', 0xAB: 'XXK', 0xAD: 'XNG',
                 0xAF: 'XGG', 0xBA: 'XNN', 0xBC: 'XXN', 0xBE: 'XDO', 0xC0: 'XYH',
                 0xC1: 'XXG', 0xC2: 'XGE', 0xC3: 'XTG',
                 0xCC: 'XYA', 0xCD: 'XXY', 0xD0: 'Y', 0xD2: 'XYB', 0xD3: 'XBE'}

    name_in = unicode_string
    name_out = ""
    for c in name_in:
        # check for shadda (double)
        if ord(c) == 0x51:
            name_out = name_out + char
        else:
            if ord(c) in transform:
                char = transform[ord(c)]
                name_out = name_out + char
    print name_out
    return name_out

#
# MAIN - Arabic to MRZ
#

# open input and output files
```

```

fin = encodings.utf_8_sig.codecs.open('Arabic source.txt', 'r') #b', 'utf-8-sig', 'ignore', 1)
fout = open('MRZ output.txt', 'w')

# loop through the input file

try:
    for arabic_name in fin:
        MRZ_name = Arabic_to_MRZ(arabic_name)
        fout.write(MRZ_name)
        fout.write('\n')
finally:
    fin.close()
fout.flush()
fout.close()

*****

```

B.7.2 机读区字符转写为阿拉伯字符

这一用 Python 进行编写的程序是将机读区字符转换为阿拉伯字符（用统一码表示）的一个示例。

机读区字符被包含在“MRZ souce.txt”这样一个文件中，而对应的阿拉伯数据被写入“Arabic output.txt”这样的文件中。

```

*****

# # -*- coding: iso-8859-15 -*-

import unicodedata
import encodings.utf_8_sig
import codecs

# TRANSLITERATE
def MRZ_to_Arabic(ascii_string):
    transform = { '<': 0x20, 'XE': 0x21, 'XAA':0x22, 'XAE': 0x23, 'U': 0x24,
                  'I': 0x25, 'XI': 0x26, 'A': 0x27, 'B': 0x28, 'XAH': 0x29,
                  'T': 0x2A, 'XTH': 0x2B, 'J': 0x2C, 'XH': 0x2D, 'XKH': 0x2E,
                  'D': 0x2F, 'XDH': 0x30, 'R': 0x31, 'Z': 0x32, 'S': 0x33, 'XSH': 0x34,
                  'XSS': 0x35, 'XDZ': 0x36, 'XTT': 0x37, 'XZZ': 0x38, 'E': 0x39,
                  'G': 0x3A, 'F': 0x41, 'Q': 0x42, 'K': 0x43, 'L': 0x44, 'M': 0x45,
                  'N': 0x46, 'H': 0x47, 'W': 0x48, 'XAY': 0x49, 'Y': 0x4A, 'XXA': 0x71,
                  'XXT': 0x79, 'P': 0x7E, 'XRT': 0x7C, 'XKE': 0x81, 'XXH': 0x85,
                  'XC': 0x86, 'XXD': 0x88, 'XDR': 0x89, 'XXR': 0x91, 'XRR': 0x93,
                  'XRX': 0x96, 'XJ': 0x98, 'XXS': 0x9A, 'XV': 0xA4, 'XF': 0xA5,
                  'XKK': 0xA9, 'XXK': 0xAB, 'XNG': 0xAD, 'XGG': 0xAF,
                  'XNN': 0xBA, 'XXN': 0xBC, 'XDO': 0xBE, 'XYH': 0xC0,
                  'XXG': 0xC1, 'XGE': 0xC2, 'XTA': 0x29, 'XTG': 0xC3, 'XYA': 0xCC,
                  'XXY': 0xCD, 'I': 0xD0, 'XYB': 0xD2, 'XBE': 0xD3}
    name_in = ascii_string

```

```
name_out = ""
# if this character is not X, does it appear by itself in the table?
search_string = ""
last_string = ""
iloop = 0
while iloop < len(name_in):
    search_string = search_string + name_in[iloop]
    if search_string in transform:
        if search_string <> last_string:
            name_out = name_out + chr((transform[search_string]))
            #insert shadda if double found
        else:
            name_out = name_out + chr(0x51)
    if search_string <> '<':
        name_out = name_out + chr(0x06)
    else:
        name_out = name_out + chr(0x00)
    #remember last string
    if search_string <> '<':
        last_string = search_string
    else:
        last_string = ""
    #clear the search string once found
    search_string = ""
    iloop = iloop + 1
print name_out
return name_out

#
# MAIN - MRZ to Arabic
#

# open input and output files

fin = open('MRZ source.txt', 'r')
fout = open('Arabic output.txt', 'wb') #b', 'utf-8-sig', 'strict', 1)
fout.write(encodings.utf_8_sig.codecs.BOM)

# loop through the input file

try:
    for MRZ_name in fin:
        Arabic_name = MRZ_to_Arabic(MRZ_name)
        Arabic_name = Arabic_name + chr(0x0D) + chr(0x00) + chr(0x0A) + chr(0x00)
        fout.write(Arabic_name)
finally:
    fin.close()
fout.flush()
fout.close()
```

```
*****
```

B.8 参考文献（资料性）

- [1] ALA-LC 罗马化表：非罗马文字转写方案，Randal K. Berry（编辑），美国国会图书馆，1997年。
- [2] 《伊斯兰百科全书》，新版，莱顿，1960年。
- [3] ISO 233:1984：《文件 - 将阿拉伯字符转写为拉丁字符》，国际标准化组织，1984年12月15日。
- [4] 《联合国地名罗马化系统：使用现状报告》，由联合国地名专家组罗马化系统工作组编制，第2.1版，2002年6月。
- [5] 创新公共服务小组对发布日期为2008年2月15日的文件 - 《机读旅行证件中阿拉伯文字的转写 - 技术报告 - 第2.3版》的意见，国际刑警组织，里昂，2008年3月17日。
- [6] 私人信件，Jan Hoogland 博士，奈梅亨大学阿拉伯语系，荷兰，2008年3月23日。
- [7] 对《机读旅行证件中阿拉伯文字的转写 - 技术报告 AMA 13052008》的意见，Mr. Abdalla M. Askar，阿联酋身份管理局。

ISBN 978-92-9275-347-4



9 789292 753474