INTERNATIONAL STANDARD

ISO/IEC 19794-14

Second edition
2022-10

Information technology — Biometric data interchange formats —

Part 14: DNA data

Technologies de l'information — Formats d'échange de données biométriques —

Partie 14: Données ADN

Reference number
ISO/IEC 19794-14:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see https://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

This second edition cancels and replaces the first edition (ISO/IEC 19794-14:2013), which has been technically revised. It also incorporates the Amendment ISO/IEC 19794-14:2013/Amd. 1:2016.

The main changes are as follows:

— Clause 6 and Annex A have been technically revised to enable the standardized interchange of DNA profile search results;

— Annex B has been technically revised to reflect the revised data interchange format;

— New Annexes E, F and G have been added.

A list of all parts in the ISO/IEC 19794 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

Forensic molecular genetics has evolved from a rapidly developing field with changing technologies into a highly recognized and generally accepted forensic science. Forensic genetics using deoxyribonucleic acid (DNA) profiling comprises a number of important applications. Examples are the investigation of biological stains to obtain evidence for the presence of an alleged perpetrator at a crime scene by comparing the genetic profiles from crime scene samples of human origin, to those available at DNA databases administered by law enforcement agencies. These also include the identification of unknown corpses in the context of both natural death and crime, immigration, paternity testing and disaster victim identification (DVI).

This document is based on DNA data from forensic DNA typing techniques that are commonly used, namely short tandem repeat (STR) profiling and other DNA typing techniques that are standardized by scientific bodies for the purpose of discriminating between individuals.

The purpose of this data interchange format is to enable the exchange of DNA data from different systems, not to impose any constraints on the specific DNA typing system/technique to be used. Where existing DNA data exchange formats have been referenced in the preparation of this document, these formats are listed as references.

Standard profiling systems exploit the non-coding parts of DNA that are referred to as "junk DNA". The coding regions, which are richer in information pertaining to specific genetic traits of an individual, are deliberately avoided in order to maintain the privacy and civil rights of the donor. In addition, national data protection and privacy legislation can impose special security safeguards, such as (but not limited to) encryption of data transfers and/or storage.

This document supports XML (Extensible Markup Language) encoding, to support a spectrum of user requirements. Annex A specifies the schema against which XML-encoded DNA data XML documents are required to validate. It also contains a sample DNA data XML document. Annex B addresses the conformance testing methodology. Annex C lists some examples of DNA analysis kits. Annex D lists the names of DNA loci. Annex E lists interoperability test data for kinship searching in the form of pedigrees. In Annex F, there is a description of interoperability tests at Level 3 (semantics). By means of the sample inclusion and comparison rules listed in Annex G, a target can be identified among a number of candidates.

Information technology — Biometric data interchange formats —

Part 14: DNA data

1 Scope

This document specifies a data interchange format for the exchange of deoxyribonucleic acid (DNA) data for person identification or verification technologies that utilize human DNA. Consideration of laboratory procedures is out of scope of this document.

This document provides the ability for DNA profile data to be exchanged and used for comparison (subject to privacy regulations) with DNA profile data produced by any other system that is based on a compatible DNA profiling technique and where the data format conforms to this document.

This document is intended to cover current forensic DNA profiling or typing techniques that are based on short tandem repeats (STRs), including STRs on the X chromosome (X-STRs) the Y chromosome (Y-STRs), as well as mitochondrial DNA. A single DNA profile for a subject can contain data resulting from more than one of these different DNA techniques. This document enables data from multiple DNA techniques to be presented in a single DNA profile for a given subject.

This document has been prepared in light of ongoing efforts to reduce human involvement in the processing (enrolment and comparison) of DNA. In anticipation of the data format requirements for automated DNA techniques, this document describes a format for both processed and raw (electrophoretic) DNA data. A normative XML schema definition (XSD) is provided in Clause A.1 for the syntax of DNA data XML documents. In Clause A.2, there is a sample DNA data XML document.

This document is not intended for any other purposes than exchange of DNA for biometric verification and identification of individuals. In particular, it is not intended for the exchange of medical and other health-related information.

This document also specifies elements of conformance testing methodology, test assertions and test procedures as applicable to this document. It establishes test assertions pertaining to the structure of the DNA data format (Type A Level 1 as defined in ISO/IEC 19794-1:2011/Amd. 1:2013) and test assertions pertaining to internal consistency of the values contained within each field (Type A,ind Level 2 as defined in ISO/IEC 19794-1:2011/Amd. 1:2013). This document also specifies test assertions pertaining to the content of DNA data XML documents (Level 3 as defined in ISO/IEC 19794-1:2011/Amd. 1:2013). The successful completion of Level 1 and Level 2 is a prerequisite for carrying out the tests at Level 3.

The conformance testing methodology specified in this document does not establish:

— tests of other characteristics of biometric products or other types of testing of biometric products (e.g. acceptance, performance, robustness, security);

— tests of systems not claimed to conform to the requirements of this document.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 2382-37, *Information technology — Vocabulary — Part 37: Biometrics*

ISO 3166-1, *Codes for the representation of names of countries and their subdivisions — Part 1: Country code*

ISO 3166-2, *Codes for the representation of names of countries and their subdivisions — Part 2: Country subdivision code*

ISO/IEC 19794-1:2011, *Information technology — Biometric data interchange formats — Part 1: Framework*

ISO/IEC 19794-1:2011/Amd. 1:2013, *Information technology — Biometric data interchange formats — Part 1: Framework — Amendment 1: Conformance testing methodology*

ISO/IEC 19794-1:2011/Amd. 2:2015, *Information technology — Biometric data interchange formats — Part 1: Framework — Amendment 2: Framework for XML encoding*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 2382-37 and ISO/IEC 19794-1 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1 Terms related to basic DNA concepts

3.1.1
deoxyribonucleic acid
DNA
complex molecule found in virtually every cell in the body that carries the genetic information from one generation to another

3.1.2
chromosome
structure within the cell that bears the genetic material as a linear strand of DNA

Note 1 to entry: In humans, each cell normally contains 23 pairs of chromosomes, for a total of 46. 22 of these pairs, called autosomes, look the same in both males and females. The 23rd pair, the sex chromosomes, differs between males and females. Sex chromosomes in males are different in size and are called X and Y. Sex chromosones in females are identical in size and both are called X.

3.1.3
Y chromosome
organized structure of the DNA molecule containing male-specific DNA only

3.1.4
non-coding part of DNA
chromosome regions not genetically expressed, i.e. not known to provide for any functional properties of an organism

3.1.5
locus
unique physical location on the DNA molecule

Note 1 to entry: The plural of locus is loci.

3.1.6
allele
member of two or more alternative forms of a DNA sequence found at a particular locus

3.1.7
tri-allelic pattern
locus that shows an occasional detection of three alleles in single-source samples

Note 1 to entry: Tri-allelic patterns can show unbalanced peak heights (Type I: The sum of heights of two of the peaks is equal to the third) or balanced peak heights (Type II: The peaks of the three alleles are of a similar height).

3.1.8
chimera
<genetic> individual having two different sets of DNAs with the code to make two separate individuals

Note 1 to entry: Otherwise said, this is a single individual composed of cells with more than one distinct genotype.

3.1.9
homozygote
individual having the same (or indistinguishable) alleles at a particular locus due to the inheritance of the same allele from each parent

Note 1 to entry: A heterozygote is an individual having two different alleles at a particular locus.

3.1.10
short tandem repeat
STR
short sequence of DNA that is repeated numerous times in direct succession

Note 1 to entry: The number of repeated units can vary widely between individuals and this high level of variation makes STRs particularly useful for discriminating between individuals.

Note 2 to entry: STR analysis is one of the most useful methods in forensic genetics for comparing specific loci on DNA from two or more samples.

3.1.11
autosomal STR
aSTR
STR region found only in autosomal chromosomes in the nucleus of the cell

3.1.12
X-STR
STR region found in female-specific DNA on the X chromosome only

3.1.13
Y-STR
STR region found in male-specific DNA on the Y chromosome only

Note 1 to entry: Y-STR can be used to trace paternal lineages as it is male specific and only inherited from fathers to their sons.

3.1.14
mitochondrial DNA
mtDNA
small circular DNA molecules located in structures used to provide energy to the cell (mitochondria)

Note 1 to entry: Mitochondria often are called the powerhouse of the cell. Their small size and abundant nature make them particularly useful when examining small or much-damaged biological material.

Note 2 to entry: The mitochondria, and thus mitochondrial DNA, are passed only from mother to offspring through the egg cell. It can be used to trace maternal lineages as it is only inherited from one's mother.

3.2 Terms related to DNA profiling

3.2.1
DNA profiling
DNA typing
technique used by scientists to discriminate between individuals by examining variations in their DNA

3.2.2
allelic ladder
artificial mixture of the common alleles present in the human population for a particular STR marker that is used during a DNA profiling process (capillary electrophoresis) in parallel with the sample of interest for accurate allele call determination

3.2.3
electropherogram
graphic representation of results of a DNA profiling process (capillary electrophoresis) with the X axis displaying the observed alleles and the Y axis recording the relative amount of DNA detected based on the relative fluorescent unit collected during analysis

Note 1 to entry: Electropherograms can be transmitted as image files if this is needed from partner DNA laboratories for validation of DNA profiles.

3.2.4
DNA profile
set of alphanumeric values describing the molecular structure at a group of loci identified in an individual's DNA

Note 1 to entry: A DNA profile is referred to as DNA fingerprint, DNA type or genetic fingerprint in other documents.

3.2.5
forensic DNA profile

DNA profile that represents a set of identification characteristics from non-coding parts of an analysed human DNA sample

3.2.6
mixed stain
biological stain that contains body fluids or tissues from more than one individual

EXAMPLE Contaminated sample, DNA sample taken from a swabbing of a surface of a drinking vessel or cigarette that has been shared

3.2.7
mixed DNA profile
DNA profile generated from a mixed stain

Note 1 to entry: In many cases where a sample consists of a stain or body fluid deposits of multiple individuals, the mixed DNA cannot be isolated when the sample is acquired.

Note 2 to entry: Where the profile of one or more of the mixed DNA sample contributors is known, the mixture can be separated into its contributing DNA profiles. One of the processes is called mixture deconvolution. This involves analysing the mixture DNA profile and exploiting the probabilistic and genetic hereditary properties of DNA to separate the profiles.

3.2.8
fully designated locus
locus of which all positions are reliably typed

Note 1 to entry: The locus status of a fully designated locus is "Normal".

3.2.9
partial locus
locus at which not all the alleles show up

3.2.10
partial DNA profile
DNA profile with partial loci or in which not all the loci targeted show up

EXAMPLE If 13 loci were targeted and only 9 could be reported, that would be termed a partial DNA profile.

Note 1 to entry: A DNA profile can be partial at the profile level, partial at the locus level or both.

3.2.11
DNA mobile processing unit
fully-functional DNA laboratory that is mobile

3.2.12
rapid DNA instrument
self-contained device that carries out a fully-automated DNA analysis of a DNA sample

3.3 Terms related to DNA databases

3.3.1
Interpol DNA Database
central forensic DNA database to which all Interpol member states can submit forensic DNA profiles of unsolved crimes, criminals, missing persons or unknown human remains through their National Interpol Bureaus, both with classic DNA profile storage or search requests and through online DNA profile data transfers from their national DNA databases for automated searching

Note 1 to entry: Interpol runs also a separate Missing Persons DNA database (I-Familia) using family DNA comparison to identify unknown human remains.

3.3.2
Interpol Standard Set of Loci
ISSOL
set of STR loci defined by the Interpol DNA Monitoring Expert Group, which recommends for use as common DNA loci for forensic DNA analyses in all forensic DNA kits and with minimum loading criteria to input a profile in the Interpol DNA Database to enable worldwide comparability of STR profiles and thus uniform crime fighting worldwide by usage of forensic DNA technology

3.3.3
Interpol DNA Monitoring Expert Group
advisory board with senior experts from Interpol member states for creation of recommendations on the use of DNA in criminal and missing person investigations including creation of Interpol DNA profile interchange standards and forms as well as rules for the Interpol DNA Database

3.3.4
Prüm DNA Database Network
decentralized database network system originally developed by some EU member states, in which biometric data, such as forensic DNA profiles, can be compared online and in real time with DNA profile search queries between the Prüm partner states

Note 1 to entry: The Prüm network has not only been implemented in a legally binding manner by all EU member states through EU legal acts but has also been extended through bilateral and multilateral state agreements to become a globally functioning Prüm data network system for biometric online data exchange (e.g. Western Balkan states).

3.3.5
European Standard Set of loci
ESS
set of STR loci defined by the ENFSI DNA Working Group which is recommended for use as minimum and common DNA loci for forensic DNA analyses in all forensic DNA kits and with minimum loading criteria to input a profile in the Prüm DNA Database Network to enable European comparability of STR profiles

3.3.6
ENFSI DNA Working Group
working group that supports the aims and objectives of ENFSI in the area of DNA casework analysis including definition of quality and STR loci standards for possible international forensic DNA cooperation

3.3.7
request
message containing one or more DNA profiles to be searched or stored or updated in or removed from a DNA profile database

3.3.8
response
message containing one or more answers depending on request message

Note 1 to entry: Match results, non-match results, error messages, notification of storage or deletion or update.

3.4 Terms related to DNA profile comparison and interpretation of results

3.4.1
power of discrimination
potential power of a genetic marker or set of markers to differentiate between any two individuals chosen at random

3.4.2
reference DNA profile
DNA profile of an identified person

3.4.3
target DNA profile
DNA profile contained in a request for comparison against a DNA profile database

3.4.4
exact match
outcome of a DNA search engine when all allele values of the compared loci are the same in two DNA profiles

3.4.5
rare allele value
allele value present in low frequency at a specific population and, therefore, much more significant than other alleles for identification purposes

3.4.6
wildcard
symbol substituting a rare allele value at a locus and matching any value at the corresponding locus in a DNA profile

Note 1 to entry: An asterisk is commonly used as a wildcard.

Note 2 to entry: Two different patterns can match in a wildcard search.

3.4.7
microvariant
allele containing an incomplete repeat unit or appearing to have values beyond a specified range

Note 1 to entry: Many STR markers are composed of a specific sequence of four nucleotides (called nucleus, core or repeat unit). The sequence of nucleotides is repeated in tandem a number of times which varies. When one of the repeat units is incomplete (e.g. shows three nucleotides instead of four), the allele is called a microvariant.

3.4.8
mismatch
outcome of a DNA search engine when only one difference, which involves a wildcard or a microvariant, is found in a comparison of two DNA profiles

3.4.9
near match
outcome of a DNA search engine when only one of all allele values of the compared loci is different in two DNA profiles

3.4.10
match
outcome of a DNA search engine that is either an exact match, a near match or a mismatch

3.4.11
non-match
outcome of a DNA search engine other than exact match, near match or mismatch

3.4.12
match quality
level of agreement between two DNA profiles

EXAMPLE The following match quality levels can be distinguished:

— Q1: exact match

— Q2: near match (only one potential difference involving a wildcard)

— Q3: near match (only one difference, which involves a microvariant)

— Q4: mismatch (only one difference other than wildcards or microvariants)

Note 1 to entry: Some DNA search engines use a likelihood ratio to quantify the match quality.

3.4.13
match count
number of identical loci found in comparison of two DNA profiles

3.4.14
adventitious match
match that happens by chance instead of having the same source or being linked by kinship

Note 1 to entry: In the case of DNA testing, not having enough distinguished characteristics (e.g. due to a partial DNA profile) can lead to adventitious matches. DNA search engine matches therefore always need forensic verification/validation for possible detection of adventitious matches.

3.4.15
candidate
DNA profile found in a DNA profile database satisfying the defined matching criteria against the target DNA profile

3.4.16
hit
candidate confirmed by a DNA examiner

Note 1 to entry: A "no-hit" is a candidate rebutted by a DNA examiner, for example, detected adventitious match.

Note 2 to entry: Validation is required to be carried out in line with forensic quality management requirements (e.g. accreditation standards).

4 Abbreviated terms

AABB	American Association of Blood Banks
BDB	biometric data block
BIR	biometric information record
CBEFF	Common Biometric Exchange Formats Framework
CE	capillary electrophoresis
CODIS	Combined DNA Index System
CRS	Cambridge Reference Sequence
DLR	DNA loci reference
DVI	disaster victim identification
ENFSI	European Network of Forensic Science Institutes
FSA	fragment sequence analysis
GLP	Good Laboratory Practice
GPS	global positioning system
HV	hypervariable regions of mitochondrial DNA
ILAC	International Laboratory Accreditation Cooperation
ISFG	International Society of Forensic Genetics
IUPAC	International Union of Pure and Applied Chemistry
IUT	implementation under test
ICS	implementation conformance statement
NA	not available
NGS	next-generation sequencing
NIST	National Institute of Standards and Technology
ORI	originating agency identifier
PCR	polymerase chain reaction
POC	point of contact

QA	quality assurance
rCRS	revised Cambridge Reference Sequence
RDBMS	relational database management system
SNP	single-nucleotide polymorphism
SQL	Structured Query Language
UTC	Coordinated Universal Time
WGS	World Geodetic System
XML	Extensible Markup Language
XSD	XML schema definition

5 Conformance

An XML document conforms to this document if it satisfies the format requirements with respect to its structure, relations among its fields and relations between its fields and the underlying input that are specified within Clause 6 and Clause A.1.

Biometric data interchange format conformance tests conform to this document if they satisfy all the normative requirements set forth in Annex B.

Implementations are not required to conform to all possible aspects of this document, but only to those that are claimed to be supported by the implementation in an implementation conformance statement (ICS), filled out in accordance with ISO/IEC 19794-1:2011/Amd. 1:2013 and Table B.1 of this document.

6 DNA data format specification

6.1 Overview

XML documents encoding DNA data shall validate against the XML schema definition in Clause A.1. In conformance to ISO/IEC 19794-1, a DNA data XML document may or may not be embedded in an appropriate CBEFF (Common Biometric Exchange Formats Framework) compliant biometric information record (BIR).

There are two kinds of fields (in XML also known as elements): simple and combined. A simple field contains only one simple data object, and a combined field contains one or more fields that can be simple or combined. Simple and combined fields are implemented by the XML mechanisms “simple type” and “complex type”, respectively.

The structure of a DNA data XML document is depicted in Figure 1.