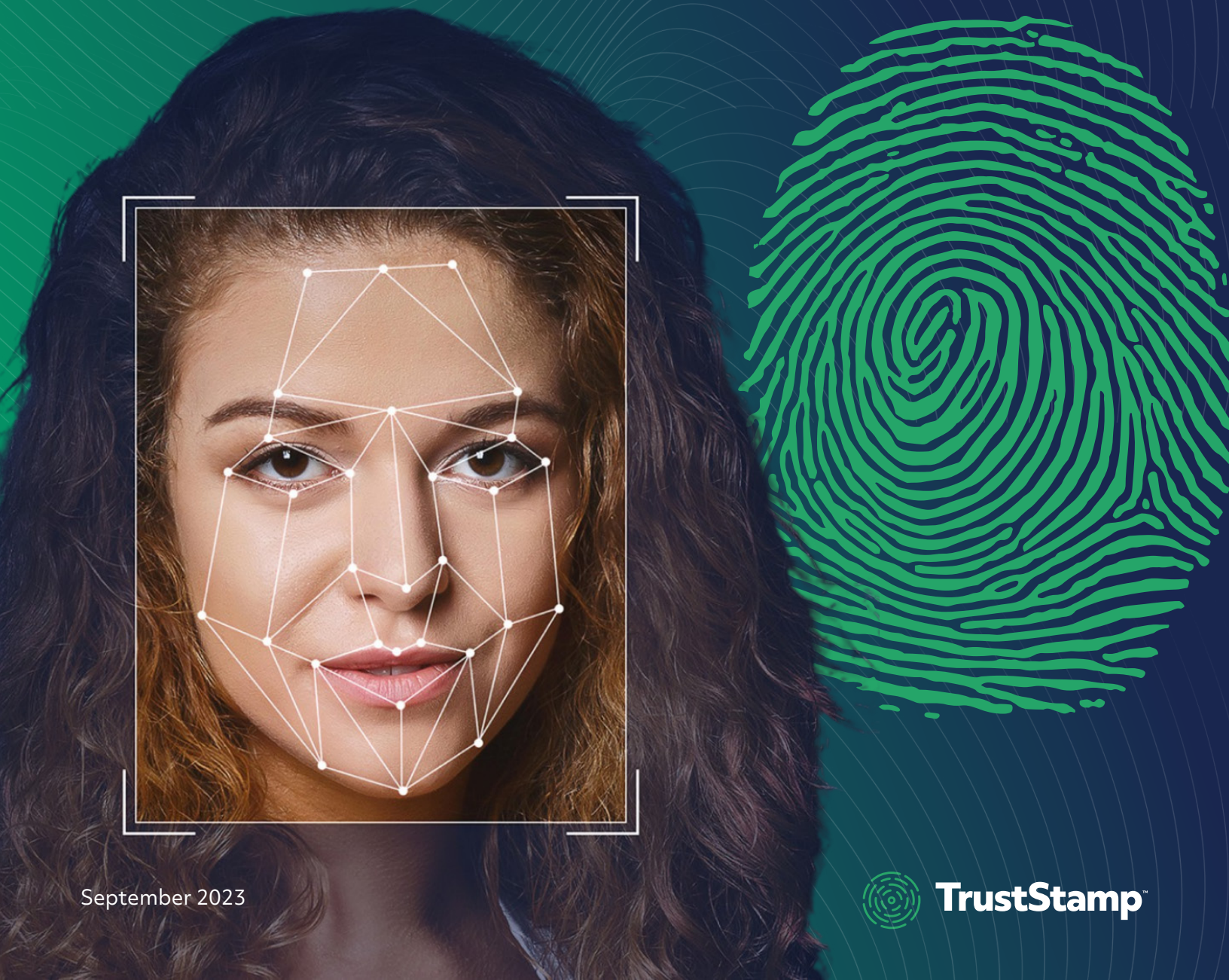
BIOMETRIC DATA & DIGITAL IDENTITY

Decoding Biometrics: Exploring the Differences Between Raw Biometrics, Templates, Vectors, and Tokens



September 2023

TrustStamp

Physiological biometric systems, which use unique physical attributes of individuals for identification, compare data derived from direct measurement of various parts of the human body. These systems either measure or convert data from the physical characteristics of individuals for identification. This summary distinguishes the different types of physiological data representations, which are: (1) raw biometric samples, (2) biometric templates or biometric vectors, and (3) biometric tokens. Let's distinguish between these:

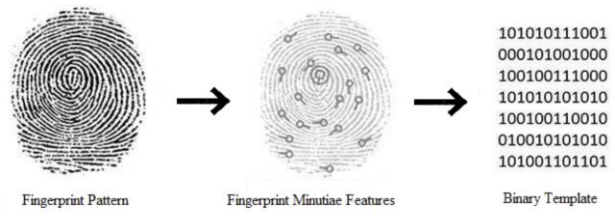
1. Raw Biometric Input

Raw biometric input is the initial unprocessed data collected directly from a biometric device, such as a fingerprint scanner, facial recognition camera, or iris scanner, among others. It represents an individual's unaltered and unique physiological or behavioral attributes, which can be anything from the detailed image of a fingerprint, a high-resolution image of a face, or the nuanced patterns of an iris. This raw data, being quite complex and voluminous, is typically processed and converted into a smaller, more manageable format, known as a biometric template or vector. The transformation from raw biometric input to a template is a critical step in biometric systems, enabling efficient storage, comparison, and identification tasks while preserving the uniqueness of the original data.

2. Biometric Templates

A biometric template is a digital representation of an individual's distinct biometric characteristic or trait. This data representation is created during the biometric enrollment process, where the software extracts distinctive data points from the raw biometric input to build a template. For instance, in a fingerprint biometric system, the pattern of ridges and valleys on the finger would be measured and captured. A neural network is typically used to transform the measurements into a digital representation that combines the measurements in a way that preserves the unique characteristics of the biometric sample.

This template is then stored and used for future matching and identification processes. See Figure 1.



A typical pattern of a fingerprint template

Biometric Vectors/Biometric Embeddings/Biometric Feature Vectors

The terminology used to describe biometric identifiers can vary from biometric provider to biometric provider. To help clarify the terminology, biometric templates can also be called biometric vectors, biometric embeddings, and biometric feature vectors. Biometric embeddings and vectors refer to the output of a neural network encoder after processing a raw biometric input. These terms are synonymous because biometric templates are often represented as a vector or embedding, such that two vectors or embeddings can usually be compared. For example, it is very common to have a face neural network encoder produce a biometric embedding (or a vector) with 256 measurements that can be treated as a point in 256-dimensional space. Typically a probe vector that is derived or extracted from a biometric sample is compared or matched against stored vectors in the gallery in an identification operation.

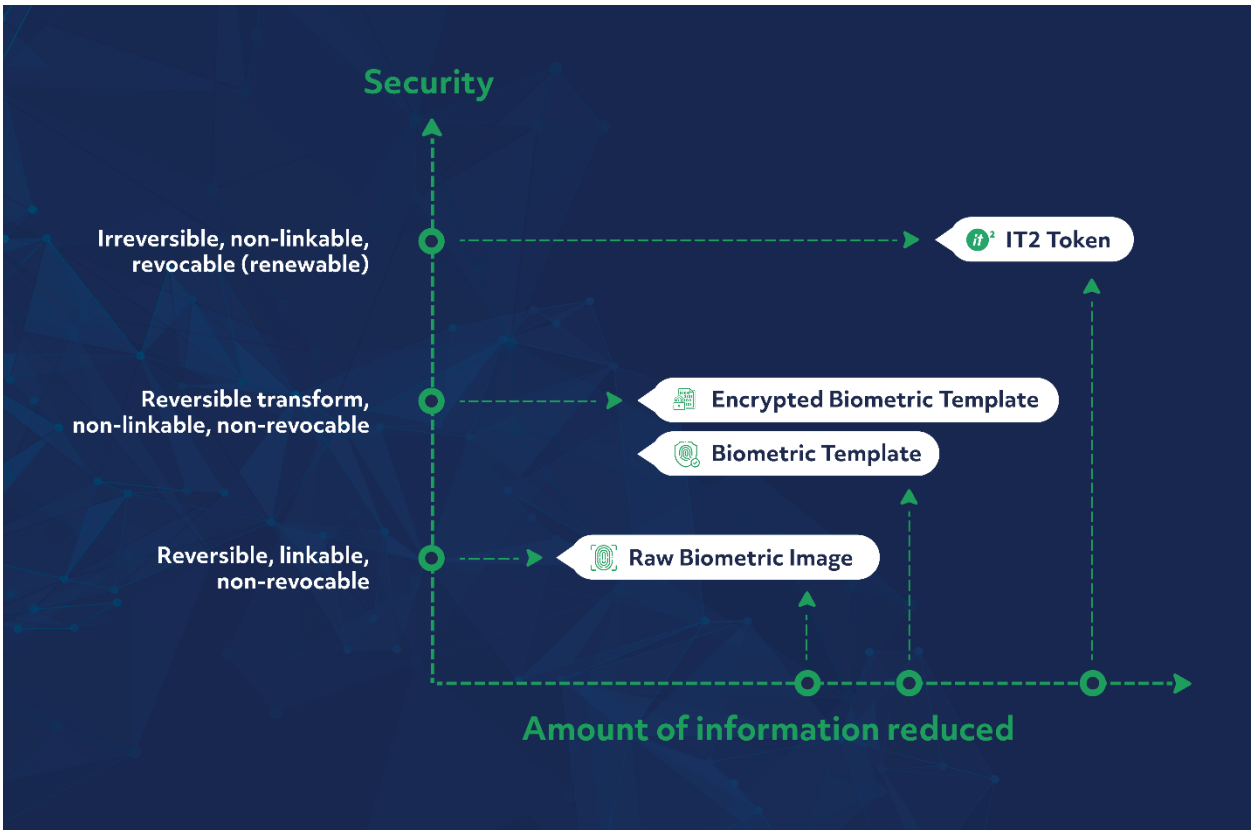
3. Biometric Tokens

A biometric token is another representation of biometric data. At Trust Stamp, we further convert biometric embeddings into a biometric token called the Irreversibly Transformed Identity Token ("IT2"). It is obtained by applying a biometric template protection scheme to a biometric template. According to ISO/IEC 24745 Biometric Information Protection, there are two types of biometric information protection schemes, namely, (1) cancellable biometrics and (2) biometric cryptosystem.

Trust Stamp's IT2 is a type of cancellable biometric that generates a data output that improves the security and searchability of the biometric data. With the IT2, even if a token value is compromised, it cannot be reversed to extract the original biometric

data. Outside of the biometric technology lexicon, this type of transformation is also known as proximity hashing.

Moreover, Trust Stamp's Biometric Token technology offers a critical security feature by reducing the amount of information from the biometric. The IT2 preserves enough essential information to perform accurate biometric verifications while rendering it nearly impossible to reconstruct the original biometric template, thereby amplifying the security measures in place.



Summary

Trust Stamp's tokens offer a range of benefits for privacy, security, and performance:

- Makes it even harder to regenerate the biometric data as compared to biometric templates or vectors. In some jurisdictions, tokens can be treated as pseudonymous data at lower risk than personal data.
- Tokens generated for one organization cannot be compared to tokens generated for another. In the event of a data breach at "Company 1", tokens used by "Company 2" would be unaffected.
- Tokens can be revoked/cancelled if required (e.g., following a data breach).
- Works efficiently at scale due to its smaller size when compared to templates and vectors, enabling a biometric to be compared against a large watchlist to detect frauds such as synthetic identity fraud.