Developers guidance

Understanding types of health and care data

Downloaded on November 19th, 2026

This is required guidance

It is legally required and it is an essential activity.

This Guide covers:

- England

From:

- Health Research Authority (HRA)

Last reviewed: 29 August 2025

Last updated: 29 August 2025



Reviewed by: Health and Care IG Panel

Two types of health and care data can be distinguished to help you determine when the relevant legal and regulatory frameworks apply:

1. Data that relates to identified or identifiable individuals. **Confidential patient and service-user information** includes information that identifies or could be used to identify the individual (such as name and address), relating to, or in connection with, an identified or identifiable individual's past or present use of services (NHS or adult social care). This broad definition is in recognition of the importance of maintaining trust in health and care services, so that all individuals can be reassured in fully engaging with these services that their confidential information will only be used in ways that they reasonably expect.
2. Data that does not or no longer relates to identified or identifiable individuals **(anonymous data)**, such that the process of rendering the data anonymous means that the laws that apply to the modified data no longer apply to those receiving it.

Important note: data can be rendered anonymous in different ways. One way is by applying a machine-learning model to a real-world dataset. For example, real-world derived synthetic data can be generated by creating average results from a large set of data. This type of synthetic data can be distinguished from synthetic data that is totally made up with no relationship to anyone living or dead (hereafter called 'artificial data').

When data is generated to statistically mimic real-world data that it replaces, an assessment should be carried out regarding the likelihood of individuals being re-identified from the synthesised data. If necessary, additional safeguards may be needed to make sure that any re-identification risks (or other privacy risks) are sufficiently remote so that it may be considered anonymous.

Creating statistically realistic synthetic data may require processing some real data first to establish model parameters. Where that real data relates to identified or identifiable individuals, its use must comply with data protection law. Read more here How should we assess security and data minimisation in AI? | ICO

Get more information on anonymisation:

ICO guidance on anonymisation

 Downloaded on 19/11/2026, for the latest information, check the website as printed documents may be outdated.