

# Introduction to the CDS

## Overview

The Clinical Dataset Structure (CDS) is a simple and intuitive way to organize and describe clinical research datasets such that they are readily interoperable and easily reusable by humans and machines. The CDS is especially designed to optimize AI-readiness of datasets.

## Why we need the CDS

During clinical research studies, multiple modalities of data are typically collected from study participants such as survey answers, blood test results, electrocardiogram data, magnetic resonance imaging (MRI), retinal images, etc. There exist some standards that specify how to structure the data and metadata for certain modalities such as the [Brain Imaging Data Structure \(BIDS\)](#) for MRI data, the [SPARC Data Structure \(SDS\)](#) for neuromodulation-related data, and the [Observational Medical Outcomes Partnership \(OMOP\)](#) Common Data Model (CDM) for observational clinical data. There is, however, no consensus on how to structure data and metadata from multiple modalities together into a consistently structured dataset. As a result, datasets from different clinical research studies are currently structured and documented differently which has two major impacts:

1. Datasets are not readily interoperable, meaning that it is difficult to combine datasets from different studies together. This limits, from instance, our ability to train AI/ML models on combined data from different studies.
2. Datasets are not easily reusable, meaning that it is difficult to understand and reuse data collected by someone else. This limits secondary analysis and reuse of data by external researchers not originally involved in a given study.

The Clinical Dataset Structure (CDS) is standard established to address these limitations by providing a simple and intuitive way for structuring clinical research data and associated metadata.

## Development of the CDS



The CDS is developed as part of the [AI-READI project](#) funded by the [NIM Bridge2AI Common Fund](#). We refer to the [Governance section](#) for information regarding the continuous development and

maintenance of the CDS.

## Acknowledgment

This project is funded by the NIH under award number 1OT20D032644. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

 March 26, 2024  September 13, 2023

 [Sanjay Soundarajan](#), [Sanjay Soundarajan](#), [slugb0t](#)

 Read the Docs

 latest