# Principles and Best Practices in Data Versioning for all Data Sets Big and Small

## The Challenge:

The growing demand for reproducibility of research results, and increased availability of data being accessed through remote services mean it is becoming more and more important for researchers to be able to cite the exact extract of the data set that was used to underpin their research. However, while the means to identify datasets using persistent identifiers have long been in place, systematic data versioning practices are currently not available.

**Produced by:** Principles and best practices in data versioning for all data sets big and small

## What is the solution?

This document provides data versioning principles and recommendations based on the analysis of 38 collected use cases from a variety of organisations and domains and other resources on data versioning.

## What is the impact?

Provides recommendations for standard practices in the versioning of research data, adding a central element to the systematic management of research data at any scale which in turn enhances reproducibility and enables the attribution of any person or organisation that contributed to the development or funding of any version of a dataset .

**Find out more about the Principles and best practices in data versioning for all data sets big and small**

August 2020