# Code Sets

## What is a Code Set?

A code set is a list of permissible values for a given data element, often referred to as reference, pick, permissible and/or valid value lists. For example, if one wanted to store a reference to a state in a database record, that value should be a valid state.

In a code set each value should be accompanied by a description and/or additional information that unambiguously describes that value, e.g. each state in the list should be accompanied by a description and synonyms. Continuing the state example, if a state code set contained the value "VA", additional information should be spelled out, i.e., Virginia.

## Why are Code Sets Important?

It is vital that the Agency identify and centrally implement broadly applicable code sets to not only obtain a shared understanding of the data we use/collect/store, but to strive for robust data quality, consistency and standardization as well as reducing redundancy and costs.

Since code sets are sources of permissible values for data elements, code sets can contribute greatly to robust data quality and consistency. The use of code sets has been a common practice for application and database developers for a very long time for things like forms, data validation and many other common application/database development scenarios.

## Code Set Management at EPA

The Office of Mission Support (OMS) has a division focused on defining and making available code sets which have wide applicability to both EPA and our partners, including states, tribes, territories, industry and the public. EPA develops code sets and obtains them from Federal government authoritative sources when applicable, for instance zip codes from the United States Postal Service. The OMS Office of Information Management (OIM)/Data Management and Standards Division (DMSD) makes all their hosted code sets digitally available through shared services and API's, including "specialty" code sets, such as facility lists, chemical lists, North American Industry Classification System (NAICS) codes, Source Classification

### Are There Other Definitions for Code Set?

Yes! In the field of IT/IM, developers often refer to bodies of application programming code as code sets. When one encounters or uses the term "code set", one must be careful to be very clear as to what is being referenced. The term is broadly understood to be authoritative lists (as described above), but it may be wise to use terms like reference and/or pick lists – familiar terms that many folks also use. When speaking of application code, it is best to explicitly state "application code" or "programming code." The focus of this overview is on lists of permissible values.

Codes (SCC), tribal identifier codes and a host of others. See the "*Additional Information*" section below to learn more about DMSD's available code sets and how to obtain/access/update them.

## What New Code Sets Will Be Available?

The DMSD is expanding its offerings and web services to include some very commonly needed code sets, which have not been encoded and made available to the enterprise to date – such as state, county, country and zip codes. In the near term, DMSD will also provision a few more "specialty" code sets – Federal Information Processing Standards State/County codes (FIPS)[1], Federal Agency Codes and an expansion of information contained in the tribal code set.

## Additional Information

For more information or new code set recommendations, contact Angelina Feldman (feldman.angelina@epa.gov, (202) 566-0701).

---

[1] FIPS are five-digit codes, which uniquely identifies United States counties. The first two digits are the state code, followed by a three digit county code.