

## ICPSR Curation Levels

September, 2020

ICPSR developed three curation levels to specify the set of curation activities to be performed on a given dataset. The level of curation a dataset receives is determined by a combination of factors, including the format and size of the dataset and the preparation activities that the data creators performed on the dataset prior to deposit with ICPSR. It does not indicate data quality, but rather is an internal tool which indicates ICPSR staff effort and curation activities required to release the study with maximum potential for secondary use. “Time to Release” is an approximation, and actual time is contingent upon resource allocation and priorities set by the project/archive as well as the size and condition of the data.

### Description of ICPSR’s Curation Levels

All studies released by ICPSR, regardless of curation level, will receive a full disclosure risk review and will include a complete metadata record, a Digital Object Identifier (DOI) that persistently identifies the data, and an ICPSR Codebook.

- **Level 1 Curation:** This is ICPSR’s base level of curation. As with all ICPSR data, our curators conduct a disclosure risk review and create a study webpage with subject terms and relevant study information including a description, title, authors, and relevant notes. Level 1 curation also includes an ICPSR codebook and data files for all major statistical software packages. Level 1 releases sometimes constitute the first version of a dataset, with more intensive processing to follow in a later release.
- **Level 2 Curation:** This includes all the work done for Level 1 curation with the added benefit of improving usability. For Level 2 Curation, ICPSR staff makes sure variables are in the most appropriate format types (e.g., string, numeric, dates, etc.), missing values are identified and documented, acronyms and abbreviations are spelled out, spelling is checked and corrected, and labels are checked for completeness and readability. The data may also be prepared for online analysis.
- **Level 3 Curation:** This builds on the Level 2 Curation release and includes customizations of documentation and the addition of survey question text to variables so that it can be indexed and searched. This level is also used to handle non-tabular or non-numeric data such as GIS and qualitative data. This more intensive curation can be expected to take longer to complete than curation at the previous two levels.

**Curation Levels Comparison Table**

<b>Curation Level</b>	<b>Approximate Time to Release</b>	<b>Value-Add Work Performed on Data</b>	<b>Included in Study Release</b>
Level 1	6 - 8 weeks from start	Disclosure review, masking as needed to mitigate disclosure risk	Study homepage with subject terms and relevant study information, including description, title, authors, and ICPSR codebook; persistent DOI; full product suite (SPSS, SAS, Stata, R); and variables indexed for search purposes
Level 2	12-14 weeks from start	Everything from Level 1, plus more nuance in disclosure risk review mitigation, processing work to improve usability of the data, including ensuring variables are in the most appropriate format type (e.g. string, numeric, date, etc.), missing values are documented and identified, acronyms are spelled out, spelling is corrected, and labels are checked for completeness and readability.	Everything from Level 1, plus the option to add on online analysis of variables and codebooks that contain variable list groupings as bookmarks in the PDF codebook if variables are grouped for online analysis
Level 3	16 -18+ weeks from start	Everything from Level 2, as well as more intensive work such as creation of question text and processing more complicated forms of data (e.g. GIS and qualitative data)	Everything from Level 2, plus searchable question text.