DISSEMINATION OF DATA

7-1 Machine Readable Products

7-2 Survey Documentation in Reports

7-3 Release and Dissemination of Reports and Data Products

SUBJECT: MACHINE READABLE PRODUCTS

NCES STANDARD: 7-1

PURPOSE: Although NCES data are made available to the public through the use of online data tools, every NCES data collection is also made available as a public-use and/or restricted-use microdata machine-readable data file.

KEY TERMS: confidentiality, confidentiality edit, edits, imputation, metadata, reference year, response rates, survey, survey system, survey year, universe, and variance.

STANDARD 7-1-1: To ensure the utility of data files created by NCES staff and contractors, all NCES data files must be accompanied by easily accessible documentation that clearly describes the metadata necessary for users to access and manipulate the data. If on-line analysis tools are used to disseminate the data to the public, the metadata must also be available on-line.

STANDARD 7-1-2: Machine-readable products must be released in ASCII format. Machine-readable products include flat files, relational databases, and spreadsheets. Each record must contain a unique case identifier such as ID. Files with multiple records per case must also contain unique record type identifiers (e.g., record number, year of data). Data files must be in one of two acceptable formats:

1. Delimited, text quoted file format that is importable; or
2. Positional files where the locations of all variables are identified (i.e., file, record within file, and position within record).

> **GUIDELINE 7-1-2A:** Data producers are invited to provide additional data sets in alternate formats that may be helpful to users. For guidance on web-based formats, see the NCES public web publishing standards; request a copy by sending an e-mail to NCESwebmaster@ed.gov.

> **GUIDELINE 7-1-2B:** To facilitate the sharing and use of data elements, national and international standards organizations have produced drafts of several standards for the creation of metadata on data elements. Examples are the International Organization for Standards "Specification and Standardization of Data Elements" standard (ISO/IEC 11179) http://metadata-standards.org/11179/ and the more detailed American National Standards Institute "Metadata for the Management of Shareable Data" Standard (ANSI X3.285) (www.ansi.org). These standards continue to be refined. Data producers should determine what metadata standards are current at the time data files are prepared and produce associated metadata for their files that are in compliance with applicable standards.

STANDARD 7-1-3: A file description and record layout must be provided for each file. The file information/metadata header must include the following:

1. Title of the survey (survey name, part, and year as applicable);

2. Name(s) of each file;

3. Reference year for the data;

4. Version number and date of release;

5. Logical record length (in positional files) or number of variables on the file (delimited files);

6. Number of records per case or observation; and

7. Number of cases in the data file. For delimited files also include the delimiters (e.g., comma, space).

STANDARD 7-1-4: For each variable on the file, the file description must include the following:

1. Variable name;

2. Data type (alpha or numeric);

3. Record number (if multiple records per case);

4. Position within the record (beginning—end, or variable number if delimited) within the record, field length, and variable label; and

5. The survey question wording and response categories.

STANDARD 7-1-5: Data set naming conventions must be standardized and must conform to International Organization of Standardization (ISO Standards) (or more recent) standards for pressing a CD.

STANDARD 7-1-6: Jewel box covers and web links or URL links must identify the survey system (e.g., HS&B, CCD), component, survey year, and version number.

STANDARD 7-1-7: All variables must be clearly identified and described.

1. The description of variables must include the universe for the variable.

2. In the case of composite variables, the description must identify all survey items used to construct the variables and must include the algorithm used to construct the variables.

3. Upper and lower case labels that clearly describe the variables must be used.

4. For all categorical variables, each value must be associated with a frequency, a percentage of total cases, and a label for each category. In public-use and restricted-use file documentation, unweighted frequencies must be included (see Standard 4-2-10 for public-use files without confidentiality edits).

5. For all continuous variables, the distribution of values (e.g., minimum, maximum, mean, and standard deviation) must be provided.

GUIDELINE 7-1-7A: The American National Standards Institute (ANSI) International Committee for Information Technology Standards (INCITS), the Office of Management and Budget, and individual statistical agencies abbreviations and coding systems should be used where applicable. NCES standard definitions and codes should be used where applicable (see Standard 1-4).

GUIDELINE 7-1-7B: Variables names should be consistent across surveys within a survey system, within and across years.

GUIDELINE 7-1-7C: In a printable record layout file, line length should be specified so that it prints correctly without wrapping and without special modification (e.g., 72 characters, 12 point type).

STANDARD 7-1-8: Data file documentation must be complete for all data files. This includes an abstract or summary that cites the methodology report or technical notes associated with the survey and a description of survey methodology that is consistent with the NCES standard for survey system documentation (see Standard 3-4). In general, survey methodology documentation for data files must include the following:

1. Description of data collection methods;
2. Weighting and imputation procedures;
3. Description of editing, error resolution, and imputation flags;
4. Guidelines for processing the data;
5. The reference year for the data;
6. Unweighted frequency counts, and response rates;
7. Information on how to use replicate weights or PSUs and stratum for variance estimation; and
8. Procedures for using weights to produce estimates.

STANDARD 7-1-9: The following data element conventions must be used:

1. Numeric fields must contain only numbers or blanks. Reserve codes for numeric fields should be extreme negative values (e.g., lower than the lowest real value).
2. "0" must represent zeros. Blanks or " — " may not be used to represent 0s.
3. Unique values must be used to distinguish between legitimate skips and nonresponse.
4. Suppression symbols, if used, must be removed from numeric fields and stored in associated "flag" fields.
5. Separate record locations must be used for all data items.
6. Imputed data must be flagged in associated "flag" fields. Imputation methods must be identified in the flag. Blanks are not legitimate values for flags.

GUIDELINE 7-1-9A: When practical, numeric data fields containing continuous variables should be identical in length.

STANDARD 7-1-10: Retain all microdata products and documentation according to appropriate Federal records disposition and archival policy.

SUBJECT: SURVEY DOCUMENTATION IN REPORTS

NCES STANDARD: 7-2

PURPOSE: To provide the appropriate amount of documentation on the data, methodology, and other important aspects of a survey in each NCES report. Survey documentation in the report should enable the reader—even the nonstatistical user—to understand its contents, and the use and limitations of data, readily and clearly.

KEY TERMS: coverage, disclosure risk analysis, frame, instruments, key variables, pretest, probability of selection, and survey.

STANDARD 7-2-1: All NCES reports must include documentation that allows the reader to understand the nature and limitations of the results presented. The level of detail included will vary depending on the type of report. The general areas to be covered include executive summary, status of data, sample design, data collection, and data presentation. List 7-2-A outlines the types of documentation to be included in the various types of NCES reports. "C" for "Complete" indicates the full item is to be included. "B" for "Brief" indicates that a brief description should be included; and "†" indicates not applicable.

STANDARD 7-2-2: Sampling standard errors must be available for all estimates included in reports. Sampling standard errors (se's) or confidence intervals (CI's) for statistics in tables and graphs can be included in reports in their entirety. In which case, se's or CI's for each table or graph are reported either in a separate table in an appendix, or in columns accompanying the statistics being presented. Alternatively, especially for publications that are targeted to general audiences, a separate table of exemplar standard errors on key statistics may be presented in the technical appendix with the detailed standard error tables for all tables and graphs included in a report available on the web.

> **GUIDELINE 7-2-2A:** To caution users who might attempt to independently test for certain differences using the standard errors provided, a cautionary note should be provided with the standard errors, stating the following:
>
> > Some estimates may be correlated with each other. Generating statistical tests for such estimates solely with these standard errors implicitly assumes these covariances are zero and may be different from the actual significance test used in the report.

List 7-2-A. Checklist for documentation to be included in NCES reports	Compendia	First Look	Statistics in Brief	R&D Report, Statistical Analysis Report	Survey Technical Report
EXECUTIVE SUMMARY[1]					
History and purpose of the survey	†	†	†	B	B
Target population	B	†	†	B	B
Time and geographic coverage of the survey	B	†	†	B	B
Main findings	B	†	†	B	†
STATUS OF DATA					
Identification of data as preliminary, revised, or final	C	C	C	C	†
Schedule of revisions	†	C	C	C	†
Relationship of survey to previous surveys in same series	B	B	C	C	C
SAMPLE DESIGN					
Target population	B	B	B	C	C
Size of target population	†	B	B	C	C
Survey frame, including source of frame, reference date, and number of units	†	B	B	B	C
Units selected for sample at each stage	†	B	†	B	C
Number of sampling units at each stage[2]	†	B	†	B	C
Sample allocation procedure at each stage	†	B	†	B	C
Sample selection process at each stage	†	B	†	B	C
Total sample sizes[2]	†	B	B	C	C
Response rates and their derivations	†	B	B	B	C
Measures of size defined for sampling with probability proportional to size	†	†	B	B	C
Summary of sources of bias	†	B	B	B	C
DATA COLLECTION					
Nature of instruments used (e.g., the contents or kinds of data sought in major sections of the instrument(s) and number of questions in each major section)	†	B	B	B	C
Method(s) of administering the instrument(s)	B	B	B	B	C
Copies of interview scripts/forms/ questionnaire, or copies upon request	†	†	†	B	C
Quality control procedures used in data processing and results of their implementation	†	†	†	†	C
Results of pretest and independent evaluations	†	†	†	†	C
Problems, if encountered	B	B	B	B	C
Type of disclosure limitations used	†	B	B	B	C
DATA PRESENTATION					
Definitions of key variables, critical concepts and constructed variables	B	B	C	C	C
Supporting numbers for graphs[3]	C	C	C	C	C
Selected exemplar standard errors for tables and graphs[4]	B	B	B	B	C

† Not applicable.

[1] Required if report is longer than 15 pages.

[2] Estimates must be rounded to a minimum of units of 10 for all unweighted counts produced from restricted data files.

[3] Numbers not included in graphics in the report must be cited to an existing report.

[4] Standard errors are required for all sample survey estimates in technical reports.

NOTE: "C" for Complete indicates the item should be included. "B" for Brief indicates that a brief description should be included.

SUBJECT: RELEASE AND DISSEMINATION OF REPORTS AND DATA PRODUCTS

NCES STANDARD: 7-3

PURPOSE: To ensure that all NCES products are disseminated in ways which help to promote the widespread use of NCES data, and to increase the awareness of NCES data among potential users.

KEY TERMS: <u>On-line analysis tool (OAT)</u>.

STANDARD 7-3-1: All NCES products must be disseminated to reach intended and potential users.

    GUIDELINE 7-3-1A: To ensure that the contents of a product reflect the needs of intended users, authors should consider user needs early in the publication development process.

    GUIDELINE 7-3-1B: In designing a publication or product, the author should consider the web presentation of the final product.

    GUIDELINE 7-3-1C: Innovative ways to disseminate NCES data should be explored. Presentations at annual meetings, seminars on specific publications, training on the use of data bases, outreach to external groups, and special research efforts using NCES data should be encouraged.

    GUIDELINE 7-3-1D: NCES should have strategies in place to collect user feedback on the utility of its products and solicit recommendations for making NCES data more useful.

STANDARD 7-3-2: When a publication is approved by IES for release, but prior to the release, the NCES contact should make any required notifications about the release to the appropriate people. Notification of the release should include the title of the report; a brief summary of the findings, including any that may be policy-sensitive; the scheduled release date; and contact information for the NCES staff member who can answer questions about the report.

STANDARD 7-3-3: NCES products should utilize a variety of dissemination techniques, as outlined in Table 7-3-A. All publications must be produced in PDF format, and all mandatory publications must also be produced in HTML format.

    GUIDELINE 7-3-3A: Efforts should be made to produce other publications in HTML format as well.

STANDARD 7-3-3: Staff responsible for NCES products requiring minor revisions must prepare an errata sheet for a level 1 review (see <u>Standard 6-1</u>). Staff responsible for NCES products that require major revisions must prepare a revised report for a level 1 review. Reissued revised reports must carry the original NCES number followed by "rev." When minor revisions approved for an errata sheet are incorporated in a web release, the NCES number must be followed by "rev."

Table 7-3-A.
NCES Products:
Required Product
Formats

Web Tools/Product:	Print	Web Products			Web Tools			
	GPO	pdf file	html	ASCII/ Excel/ data base file	Locator	Peer tool: Public access	Peer tool: Limited access	Glossary search
Standard products: Compendium	R	R	R					
Directory					R	R		
NCES Handbook		R	O					
Updated indicator		R	R					
Pre-release data				R			R	
Statistical Analysis Report		R	O					
R&D Report		R	O					
Technical/Methodological Report		R	O					
First Look		R	O					
Statistics in Brief		R	O					
Data File (including CD ROM/DAS/WEB)				R		O	O	
Data file Documentation/User's Manual (must accompany data file)		R	O					
Video/Data								
Conference Report		R	O					
Working Paper		R						
Non-data Videotape (e.g., conference, Commissioner's statements)		R						
Brochure/Pamphlet	O	O						
Newsletter	O	O						
Co-op Product (e.g., FORUM, NPEC)		R						
Questionnaire		R	O					O
Glossary		R	O					O

_____

R Must be produced for the document/product.

O Optional format for the document/product

10