Preservation Action Plan for Structured Data/Databases
National Archives and Records Administration (NARA)

Plan Date: 20220714
Template: 202105

Structured Data/Databases

Information that is accessed and updated through a database management system that has been organized, structured, and stored so that it can be manipulated, searched, and extracted for various purposes.

Significant Properties of Structured Data/Databases

As a general rule, appearance properties are not essential to databases. That being said, there is one that will always be core to this record type.

Behavior is generally the most important property to preserve in databases, followed by Structure. Database records require a documented data model describing what categories of data will be stored in which fields, columns or tags; data types (numeric, currency, alphabetic, name, date, address); and, if possible, controlled vocabulary. Documentation of the structure enables the retention of the database behavior.

Appearance

Name	Definition	Function Description
Character Encoding	The code used by computers can be: ● EBCDIC ● ASCII ● EBCPAC ● Binary ● Zone Decimal	The sequence of characters (letters, numbers, punctuation, and certain symbols) or coding that translate human readable or natural language characters to a specialized format for efficient transmission or storage. Assumption: Always has to exist and needs to be identified in order to open in a compatible format or to transform to another format, such as ASCII. Must meet Ingest requirements.

Text Properties	Properties such as font, text size, or color.	These have never been considered primary properties in the past. If at some point in time they are important to understanding the record, then they should be reconsidered as significant for databases.

Structure

Name	Definition	Function Description
Database Schema	Must include record layout, can also include any or all of code lists, table structure, data dictionary, linkage. (If any of these exist exists, they are significant.)	Manner in which individual data elements and data tables are organized, interrelated, and displayed.
Technical Metadata	Metadata describing the specific database format, software, software version, etc. This is generally automatically embedded in the file header.	Supports the ability to potentially recreate interactions with the data, such as queries or graphing, can be recreated.

Behavior

Name	Definition	Function Description
Manipulation Functionality	The ability to examine relationships within a table or between tables.	Once there is a defined schematic, the ability to manipulate exists.
Display Query	Presents results of query in a particular format. Like queries, this depends on use by the creator and expected use of data. If a report, see other categories such as textual records (.pdf, .doc) or web records for its component properties.	This does not refer to the GUI interface, but instead to the underlying query string(s). The variant of SQL used should be identified. Depending on the use of the data by the agency, and expected use of the data, any defined queries maintained by the creator may be necessary to preserve. The importance of maintaining is

	If geospatial, graph and plot are significant properties. See the plan for the geospatial records category for other functionalities that are significant.	based on appraisal questions and analysis: does the query demonstrate how the agency used/interpreted the data for decision making, legal use, or general accountability. Is it a reflection of how an agency made a decision? If yes, then determine if the results display was saved to some other format for reporting purposes. If the results were not saved, and the mechanism for decision making was the query itself, then the query template/construction becomes significant, and must be saved.

Context

Name	Definition	Function Description
Series	A group of related records that are normally used and filed as a unit because they relate to a similar activity/function. The relationship between records and the series of which they are a part must be preserved.	
Descriptive Metadata	May include but is not limited to: Caption/Title; Subject; Date; Event; Creator; Transaction.	Information contained within the record that refers to the intellectual content of material and aids discovery of such materials. Assumes that all descriptive metadata for a database is contained within the structure and content of the database itself.

Current NARA Transfer Guidance for Structured Data/Database Records
[Bulletin 2014-04](#)

- Preferred:
    - Comma Separated Value (CSV)
    - OpenDocument Format Spreadsheet (ODS)
    - ASCII Text

- Acceptable:
    - Microsoft Excel Office Open XML
    - Microsoft Excel 97 Binary Document Format

Current NARA Format(s) for Public Access and Reference for Database Records

Formats for Public Access are those made available online through the National Archives Catalog. Formats for Reference are defined as those made available to researchers upon direct requests for digital copies.

Formats Available for Public Access: Content created or delivered for public access in the Catalog is delivered primarily in the following file formats: PDF (Textual and Image), JPEG (Textual and Image), MP3 (Audio), and MP4 (Audio/Video) and ASCII (Datasets). Other file formats may be present depending on when they were added to the Catalog.

Format(s) Available for Reference: When available, records may be delivered to researchers in the formats in which they are preserved.

Some databases and datasets are made searchable at a row level through the Access to Archival Databases (AAD) tool and may be requested by researchers from NARA's Electronic Records Division (RRE).

Comments and Notes

In general, NARA accessions databases in formats defined in 36 CFR 1235 and NARA Bulletin 2014-04. NARA appraises the contents of database files and not the functionality for permanent retention in the National Archives. The Transfer Guidance drives the formats in which we preserve records, and ensure that content is preserved in formats that are sustainable and can be migrated over time to maintain accessibility.