



INSTITUTE of
Museum and Library
SERVICES

**Museum Universe
Data File Documentation**

May 2014



Museum Universe Data File

Documentation

May 2014

Justin Grimes

C. Arturo Manjarrez

Kim A. Miller

Deanne W. Swan

Institute of Museum and Library Services

C. Arturo Manjarrez

Director, Office of Planning, Research, and Evaluation

Institute of Museum and Library Services

Institute of Museum and Library Services

Susan H. Hildreth

Director

About the Institute of Museum and Library Services

The Institute of Museum and Library Services is the primary source of federal support for the nation's 123,000 libraries and 35,000 museums. Our mission is to inspire libraries and museums to advance innovation, lifelong learning, and cultural and civic engagement. Our grant making, policy development, and research help libraries and museums deliver valuable services that make it possible for communities and individuals to thrive. To learn more, visit www.ims.gov and follow us on Facebook and Twitter.

As part of its mission, IMLS conducts policy research, analysis, and data collection to extend and improve the Nation's museum, library, and information services. IMLS research activities are conducted in ongoing collaboration with state library administrative agencies; national, state, and regional library and museum organizations; and other relevant agencies and organizations. IMLS research activities are designed to provide consistent, reliable, complete, and accurate indicators of the status and trends in library and museum services; and to report timely, useful, and high-quality data to Congress, the states, other policymakers, practitioners, data users, and the general public.

Contact Information

Institute of Museum and Library Services
1800 M Street NW, 9th Floor
Washington, DC 20036-5802
202-653-IMLS (4657)

www.ims.gov If you have questions about this data file please email research@ims.gov.

Suggested Citation:

Grimes, J., Manjarrez, C.A., Miller, K.A., & Swan, D.W. (2014). *Museum Universe Data File: Documentation*. (IMLS-2014-MUDF-01). Institute of Museum and Library Services: Washington, DC.

Unless specifically noted, all information contained herein is in the public domain and may be used and reprinted without special permission. Citation of this source is suggested.

Introduction

This documentation provides detailed information about the Museum Universe Data File (MUDF), a list of known museums in the United States maintained by the Institute of Museum and Library Services (IMLS). This list contains descriptive information about museums in the 50 states and the District of Columbia based on public records and administrative data. The Museum Universe Data File is actively maintained and released publicly twice a year.

This is the second release of the Museum Universe Data File. Unlike the previous Museum Universe Data File (FY 2013), this release is markedly different in several ways. This release of the Museum Universe Data File (FY 2014 Q3) is based on more current data sources than the previous release (FY 2013) and employs a different approach.

This change in process helped to improve the currency and quality of information provided. This new process enabled the introduction of several new data items such as recent income and revenue as well as additional metadata regarding the underlying data sources. Moreover, this release introduces a new unique identifier for museums, which will provide a common identifier for identifying museums in the United States.

Background

The Museum Universe Data File is a list of known museums in the United States that is maintained by Institute of Museum and Library Services (IMLS) Office of Planning, Research and Evaluation. Included in this file are the basic institutional identifying information of over 35,000 museums including aquariums, arboretums, botanical gardens, art museums, children's museums, general museums, historic houses and sites, history museums, nature centers, natural history and anthropology museums, planetariums, science and technology centers, specialized museums, and zoological parks. Data for this listing is collected and aggregated from a variety of sources including IMLS administrative data and records, Internal Revenue Service (IRS) nonprofit data (e.g., 990, 990-EZ, 990-N), grantee lists from private foundations and data obtained from third party commercial vendors.

The Museum Universe Data File is used to support IMLS efforts and activities. This data will be used for sampling purposes in the upcoming [Museums Count survey](#). Museums Count is a nationwide survey that will provide more detailed information about the museum sector including information about collections, programs, services, visitations, as well as financial information. In 2013, the Institute of Museum and Library Services first made this resource publicly available to better inform the public about museums and museum related activities in the United States. This information is useful to the museum sector, policymakers at the federal, state, and local levels, the American Alliance of Museums and its members or customers, museum and public policy researchers, the public, journalists, and others.

Caveats for Using these Data

The information contained in this data file comes from a variety of sources, primarily public information. There are inherent limitations in the use of such administrative information. Although steps were taken to improve data quality the Institute of Museum and Library Services cannot attest to overall accuracy and quality of the information contained within. Examples of possible data accuracy and quality issues include incorrect or misspelled contact information, duplicate entities, incorrect or inaccurate discipline classification, the listing of inactive or non-museum entities, or missing information about active museums.

Please report any data inaccuracies or data quality issues to research@imls.gov

Acknowledgement

This release's data collection effort was done in coordination with Neville Vakharia from Westphal College of Media Arts and Design at Drexel University. The updated Museum Universe Data File is part of a broader cooperative agreement with Drexel University to development a new set of online tools and resources for museum leaders, researchers, and students entering the museum and cultural sectors.

Description of Data Elements

Below is a list and description of data elements available in the Museum Universe Data File (May 2013).

MID – Unique museum identifier

NAME – Name of institution

ALTNAME – Alternative name of institution

ADDRESS – Address institution, Street address

CITY – Address institution, City

STATE - Address institution, State

ZIP – Address institution, Postal zip code plus zip code extension

ZIP5 - Address institution, Postal zip code

ZIP4 – Address institution, Four-digit postal zip code extension

PHONE – Institution phone number

WEBURL – Institution website address

DISCIPL –Museum discipline or type

Code	Discipline
ART	Art Museums
BOT	Arboretums, Botanical Gardens, & Nature Centers
CMU	Children's Museums
GMU	Uncategorized or General Museums
HSC	Historical Societies, Historic Preservation
HST	History Museums
NAT	Natural History & Natural Science Museums
SCI	Science & Technology Museums & Planetariums
ZAW	Zoos, Aquariums, & Wildlife Conservation

EIN - Federal Employer Identification Number (EIN) number

NTEECC – National Taxonomy of Exempt Entities Core Codes, a classification system used by the Internal Revenue Service (IRS) and the National Center for Charitable Statistics (NCCS) to classify nonprofit

organizations. Additional information including a full list of NTEE organizational codes and definitions can be found at the following site: <http://nccs.urban.org/classification/NTEE.cfm>. Definitions for terms in the National Taxonomy of Exempt Entities (NTEE) and Nonprofit Program Classification System (NPC) are based, in large part, on their counterparts in the AIRS/211 LA County Taxonomy of Human Services, Information and Referral Federation of Los Angeles County, Inc.

TAXPER - Tax period of the latest return filed (YYYYMM). Tax Information comes from Internal Revenue Service (IRS) Business Master File, March, 2014.

INCOME –Gross receipts from most recent Internal Revenue Service (IRS) 990 Form. Income is a computer generated amount by the Internal Revenue Service (IRS). Information comes from Internal Revenue Service (IRS) Business Master File, March, 2014. Information comes from Internal Revenue Service (IRS) Business Master File, April, 2014.

REVENUE – Total revenue from most recent Internal Revenue Service (IRS) 990 Form. Information comes from Internal Revenue Service (IRS) Business Master File, March, 2014.

INCOMECD - Income Codes relate to the amount of income shown on the most recent Form 990 series return filed by the organization. Information comes from Internal Revenue Service (IRS) Business Master File, March, 2014.

Code	Description
0	\$0
1	\$1 to \$9,999
2	\$10,000 to \$24,999
3	\$25,000 to \$99,999
4	\$100,000 to \$499,999
5	\$500,000 to \$999,999
6	\$1,000,000 to \$4,999,999
7	\$5,000,000 to \$9,999,999
8	\$10,000,000 to \$49,999,999
9	\$50,000,000 to greater

LOCALE4 – National Center for Education Statistics (NCES) Urban-Centric Locale Codes classification; based on geocoded address of institution. Additional information can be found at the following site: https://nces.ed.gov/ccd/rural_locales.asp

Code	Description
1	City – Territory inside an urbanized area and inside a principal city

2	Suburb – Territory inside an urbanized area but outside a principal city
3	Town – Territory inside an urban cluster that is outside an urbanized area
4	Rural – Census-defined rural territory that is outside an urbanized area or urbanized cluster.

AAMREG – museum region, regions are determined by the American Alliance of Museums (AAM). Additional information can be found at the following site: <http://www.aam-us.org/>.

Code	Name	Description
1	New England	Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island, Vermont
2	Mid-Atlantic	District of Columbia, Delaware, Maryland, New Jersey, New York, Pennsylvania
3	Southeastern	Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia
4	Midwest	Iowa, Illinois, Indiana, Michigan, Minnesota, Missouri, Ohio, Wisconsin
5	Mount Plains	Colorado, Kansas, Montana, North Dakota, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, Wyoming
6	Western	Alaska, Arizona, California, Hawaii, Idaho, Nevada, Oregon, Utah, Washington

LATITUDE - Latitude of institution's address; this field consists of two integers and six decimal places, with an explicit decimal point, decimal degree format, World Geodetic System Datum 1984; determined by geocoding service.

LONGITUDE - Longitude of institution's address; this field consists of a negative sign, three integers and six decimal places, with an explicit decimal point, decimal degree format, World Geodetic System Datum 1984; determined by geocoding service.

FIPST - Federal Information Processing Standards (FIPS) State Code (US Census 2010); based on geocoded address of institution.

FIPSCO - Federal Information Processing Standards (FIPS) County Code (US Census 2010); based on geocoded address of institution.

TRACT - US Census Tract (2010); based on geocoded address of institution.

BLOCK - US Census Block (2010); based on geocoded address of institution.

FIPSMIN - Federal Information Processing Standards (FIPS) Minor Civil Division (U.S. Census 2010); based on geocoded address of institution.

FIPSPLAC –Federal Information Processing Standards (FIPS) Place code (U.S. Census 2010); five-digit place; based on geocoded address of institution.

CBSACODE – Core Based Statistical Area (CBSA) code, U.S. Office of Management and Budget (2013); based on geocoded address of institution.

METRODIV –Metropolitan Statistical Area Division code, U.S. Office of Management and Budget (2013); based on geocoded address of institution.

MICROF – Micropolitan Area flag, U.S. Office of Management and Budget (2013); based on geocoded address of institution; 0 – No, 1- Yes

CNTRYCD – Country code, ISO 3166

IRS990_F – IRS 990 Data source flag; 0 – No, 1- Yes

IMLSAD_F – IMLS administrative data source flag; 0 – No, 1- Yes

FCT3P_F– Third party (Factual) source flag; 0 – No, 1 - Yes

PFND_F– Private grant foundation data source flag; 0 – No, 1 - Yes

Data Collection and Processing

The Museum Universe Data File (May 2014) was generated through the combining and aggregating of several data sources that contain organization information about museums in the United States. Information for the data sources were integrated by linking records through common identifiers. Once sources were combined, staff reconciled records using an established protocol to improve data quality and reduce duplication. Additional geographic information was added during post processing through a geocoding service.

Data Integration, Record Linkage, and Record Reconciliation

Data from each of the sources was cleaned and standardized into a common format. Records from each data source were linked together using common identifiers when possible (e.g., EIN). Due to size and quality, the primary data source was determined to be the IRS nonprofit data.

The previous Museum Universe Data File (May 2013) was validated against existing IRS nonprofit and third party data sources. Records that were not found in current data sources were removed. Remaining records were appended and combined with existing IRS nonprofit data. This process was repeated for the remaining sources (e.g, grant data). Once the sources were appended or combined, duplicate records were identified using a pattern matching process based on the name and address fields. Staff reconciled these potential duplicates using an established protocol. In situations where similar information existed for duplicate sources priority was given first to IRS nonprofit data, when possible, then to the most recent record.

Classification of museum discipline type was determined based on a crosswalk of the classification schemes associated with the data sources used to compile the MUDF. This classification scheme is derived from the National Taxonomy of Exempt Entities (NTEE) taxonomy. Organizations were first classified based on their NTEEC code. If the NTEEC code was missing, invalid, or listed as Museum (A50) then amore granular classification from another data source was used, if it existed. For example, IMLS administrative records use an internal code for classifying museums by discipline. When no classification data was available, organizations were classified based on keywords. For example, if the word “planetarium” appeared in the name of the organization it was classified as “Science & Technology Museums & Planetariums.” Due to limited information and quality issues with classification information from available sources, discipline type should be considered an approximation.

Geocoding

As part of the post-processing of the data files, supplemental geographic information was added, where possible, to each record. The following supplemental geographic information was provided

- Longitude (LONGITUDE)
- Latitude (LATITUDE)
- FIPS State Code (FIPSST)

- FIPS County Code (FIPSCO)
- FIPS Place Code (FIPSPLAC)
- Census Tract (TRACT)
- Census Block (BLOCK)
- Core based statistical area code (CBSACODE)
- Metropolitan/Micropolitan area flag (MICROF)
- Minor Civil Division (**FIPSMIN**)
- Locale (LOCALE4)

Information about data sources

Data for this listing is collected and aggregated from a variety of sources including Internal Revenue Service (IRS) nonprofit data (e.g., 990, 990-EZ, 990-N), IMLS administrative grant data, grantee lists from private foundations, and data obtained from third party commercial vendors.

Internal Revenue Service 990 Data (FY 2009-2013)

The Internal Revenue Service (IRS) Business Master File contains descriptive information about all active and registered nonprofit organizations that filed for tax exempt status with the IRS (i.e., those organizations that filed Form 990, 990-EZ, or 990-PF).

In addition to the Business Master File, the IRS also maintains data about small exempt organizations that are not required to file Form 990, 990-EZ, or 990-PF but are required to file Form 990-N to maintain their exempt status, e.g., organizations that have gross receipts less than \$50,000. Beginning in 2008, the IRS introduced a new form for these small nonprofits: Form 990-N [called the Electronic Notice (e-Postcard) for Tax-Exempt Organizations Not Required To File Form 990 or 990-EZ].

Museums were identified in this data source¹ as any organization that filed between fiscal year 2009 and fiscal year 2013 and identified themselves as a museum according to the National Taxonomy of Exempt Entities code:

- A50 Museum, Museum Activities
- A51 Art Museums
- A52 Children's Museums
- A53 Folk Art Museums
- A54 History Museums

¹ Due to the lack of NTEECC codes in 990-N data museum organizations were selected to be any organization that contained any of the following keywords: "MUSEUM," "ZOO," "ZOOLOGICAL," "AQUARIUM," "ARBORETUM," "BOTANICAL GARDEN," "OBSERVATORY," "CONSERVATORY," "PLANETARIUM," "NATURE CENTER," "NATURAL HISTORY," "HISTORICAL SOCIETY," "HISTORICAL HOUSE," "HISTORIC HOUSE," "HISTORIC SITE" in the organizational name.

- A56 Natural History, Natural Science Museum
- A80 Historical Organizations
- A82 Historical Societies & Historic Preservation
- C41 Botanical Gardens, Arboreta and Botanical Organizations
- D50 Zoo, Zoological Society

IMLS Administrative Records (IMLS, 2009-2014)

The Institute of Museum and Library Services (IMLS) collects information about museums as part of the agency's mission. Organizational information about museums was selected for those organizations identified as museums that appeared in IMLS administrative systems between FY 2009 to FY 2014. This information was then used to cross validate existing records from existing data sources; additional information was appended.

Private Foundation Grant Data (Foundation Center, 2009-2013)

Foundation Center data contains information about grants from 1,000 of the largest foundations in the United States. It includes all of the grants of \$10,000 or more awarded by independent, corporate, and grant making operating foundations, as well as unrestricted and donor-advised fund grants, when available, for community foundations.

Organizational information procured from the Foundation Center was used to select museums that received funds between 2009 and 2013. This information was used to cross validate existing records from other existing data sources and additional information was appended.

For more details about Foundation Center data see: <http://data.foundationcenter.org/about.html>.

Third Party Commercial Sources (Factual)

Factual is a commercial data aggregator and supplier that maintains a database of over 65 million points of interest in 50 countries; 23 million places in the United States. This data is aggregated and procured from several sources: third party provided contributions, public data sources, volunteered geographic information, as well as Factual's proprietary machine learned process for identifying data across the web. This data contains information about a place's name, address, phone number, latitude/longitude, type of organization and when available website.

Factual's Resolve service was used for entity resolution to cross validate existing records from existing data sources and to append any additional records that were identified as museums but not listed in any existing source (IRS, grant data, private foundation data). Museums were identified to be any organization matching museum specific categories in Factual's [place taxonomy](#). This taxonomy maps to existing North American Industry Classification System (NAICS) and Standard Industrial Classification (SIC) codes.