



INSTITUTE of  
**Museum and Library**  
SERVICES

**Museum Universe  
Data File Documentation**

*November 2015*



# **Museum Universe Data File**

Fiscal Year 2015, Third Quarter Release

Documentation

November 2015

**Justin Grimes**

**C. Arturo Manjarrez**

**Kim A. Miller**

**Deanne W. Swan, PhD**

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**

Kathryn K. Matthew, PhD

*Director*

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](http://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](http://www.ims.gov)

The IMLS Internet address is <http://www.ims.gov>.

This publication is available only on the Web at <http://www.ims.gov/research/>

IMLS will provide an audio recording of this publication upon request.

For questions or comments, contact [Research@ims.gov](mailto:Research@ims.gov)

### **November 2015**

Cover Design: Ellen Arnold Losey

#### **Suggested Citation:**

Grimes, J., Manjarrez, C.A., Miller, K.A., & Swan, D.W. (2015). *Museum Universe Data File, FY 2015 Q3 Release: Documentation*. (IMLS-2015-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.

# **Museum Universe Data File** **FY 2015, 3<sup>rd</sup> Quarter Release:** **Data Documentation**

## **Introduction**

The Museum Universe Data File (MUDF) is a data resource that provides information about known museums and museum related organizations in the United States. The MUDF is updated and maintained by the Institute of Museum and Library Services (IMLS) and is publicly released on a regular basis.

This document provides technical information about the MUDF, a list of known museums and related organizations in the United States maintained by IMLS. This data file contains descriptive information about museums in the 50 states and the District of Columbia based on public records and administrative data.

## **Background**

The MUDF is a list of known museums and related organizations in the United States that is maintained by IMLS's Office of Planning, Research and Evaluation (OPRE). This file includes basic institutional identifying information of approximately 33,000 museums and related organizations, 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 has been 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), lists of grantees from private foundations, and data obtained from third party commercial vendors.

The MUDF is used to support IMLS activities. The data will be used for sampling purposes in the upcoming [Museums Count survey](#). Museums Count is planned to be a nationwide survey that will provide more detailed information about the museum sector, including information about collections, programs, services, visitations, and financial information. In 2013, IMLS made the MUDF publicly available to inform the public about museums and museum-related activities in the United States.

## **Purpose**

The primary purpose of the MUDF is to provide a maintained list of active museums in the United States. This information is useful to the museum sector, policymakers at the federal, state, and local levels, museum and public policy researchers, the public, journalists, and others.

## Intended Audience

The MUDF is intended for the general public. The data is provided at no cost in order to inform the public about museums and museum-related services and activities. The dataset is in the public domain, and there is no need to contact IMLS before using this data or incorporating this data into a separate system or application. IMLS does, however, encourage users to contact us to tell us about how you have used the data. Users should acknowledge credit either through citation or by referencing the Institute of Museum and Library Services as a data source.

## Data Quality Information

The information contained in this data file is derived from a variety of public sources, primarily information from tax-exempt nonprofit organizations from the IRS. There are inherent limitations in the use of such administrative information. Although steps have been taken to improve data quality, IMLS cannot attest to overall accuracy and quality of the information contained within. IMLS advises users to develop strategies for addressing these limitations. Specific caution should be used when using financial information or post-processed geographic information.

Users of the MUDF understand and accept the following:

- 1) This information is provided as “general reference” information.
- 2) Information is provided for the purpose of describing museums and museum-related organizations in the United States and may not be inclusive of all museum and museum-related organizations.
- 3) This information is provided by, and periodically updated by, the Institute of Museum and Library Services and is based on information from various administrative sources (e.g., IMLS grant records and IRS non-profit tax filings).
- 4) The Institute of Museum and Library Services does not guarantee the accuracy, relevance, timeliness or completeness of the information contained in the dataset. It is the responsibility of the user to check for and revise information as necessary.
- 5) The Institute of Museum and Library Services does not endorse or sponsor the organizations contained in this dataset nor the views they express or the products or services they provide.
- 6) The Information provided in this dataset is in the public domain.

If you have any questions or would like to report data quality issues, please contact IMLS at [research@imls.gov](mailto:research@imls.gov).

## Summary of Changes

This is the third public release of the Museum Universe Data File.

Major changes in this release include:

- Addition of university and college museums using a separate process with data from Integrated Postsecondary Education Data System (IPEDS). New elements include IPEDS (unique identifier if museum or museum-related organization is associated with post-secondary institution) and INSTNAME (name of post-secondary “parent” institution);
- Updated IRS 990 financial information from the most recent Business Master File May 2015 (TAXPER, INCOME, REVENUE, and INCOMECD);
- Additions and corrections based on public information from professional organizations such as International Planetarium Society;
- A systematic review of information contained in the file to remove inactive museums and museum-related organizations;
- Due to changes in geocoding services these variables will no longer be available: Country Code (CNTRYCD), Core based statistical area code (CBSACODE), Metropolitan/Micropolitan area flag (MICROF), Minor Civil Division (FIPSMIN), and FIPS Place Code (FIPSPLAC); and
- Addition of new variables such as CONGDIST (Congressional District) and BEAREG (Bureau of Economic Analysis Region).

## Data Collection and Processing

The MUDF was initially compiled in FY 2013 through the combination and aggregation of several data sources that contained organization information about museums and museum-related organizations in the United States. Data for this listing was collected and aggregated from a variety of sources, including IRS nonprofit data (e.g., 990, 990-EZ, 990-N), IMLS administrative grant data, lists of grantees from private foundations, and data obtained from third party commercial vendors.

Since the initial release, additional information has been added through the inclusion of other public sources of information about museums and museum-related organizations from professional organizations, such as the International Planetarium Society (IPS) and American Public Gardens Association (APGA), and federal government agencies, such as National Oceanic and Atmospheric Administration (NOAA), the National Archives and Records Administration (NARA), and the National Park Service (NPS), and through user-contributed feedback. To improve quality, additional steps were taken to help standardize, validate,

and correct the information provided. These steps included automated validation of data elements as well as human review of data fields.

Supplemental geographic information has been added, where possible, to each record through a commercial geocoding service. Geocoding information is derived from the geocoded address information listed in the data file (e.g. G STREET, G CITY, G STATE, G ZIP). The geocoded address is derived from the physical address, if available. If no physical address was available, the geocoded address is based on the administrative address or IRS nonprofit tax information. The quality of geocoded information can vary greatly depending on a variety of factors (e.g., the quality and accuracy of address information, the quality of underlying geocoding service, the propensity of postal box addresses). Caution should be used when using information added during post-processing of the file.

The MUDF provides the following supplemental geographic information based on geocoded information:

- Longitude (LONGITUDE)
- Latitude (LATITUDE)
- FIPS State Code (FIPSST)
- FIPS County Code (FIPSCO)
- Census Tract (TRACT)
- Census Block (BLOCK)
- BEA Region (BEAREG)
- AAM Region (AMREG)
- Locale (LOCALE4)

## **Additional Information about Data Sources**

The MUDF used a variety of data sources in its compilation. This section provides additional information about data sources used to generate this release of the MUDF.

### ***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., 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 which are required to file Form 990-N to maintain their exempt status (e.g., organizations that have gross receipts of

less than \$50,000). Beginning in 2008, the IRS introduced a new form for these small nonprofits: Form 990-N, also 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 source<sup>1</sup> as any organization that (1) filed between fiscal year 2009 and fiscal year 2013 and (2) identified itself 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
- 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 collects information about museums as part of the agency's grantmaking 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 used to cross-validate records from other data sources. For some records, additional information from the administrative data was appended.

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

Data from the Foundation Center contained information about grants from 1,000 of the largest foundations in the United States. It included all of the grants of \$10,000 or more that were awarded by independent, corporate, and grantmaking 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

---

<sup>1</sup> 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," or "HISTORIC SITE" in the organizational name.



used to cross-validate records from other existing data sources. For some records additional information was appended to the data file. 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 more than 65 million points of interest in 50 countries and 23 million places in the United States. This data is aggregated and procured from several sources, to include third party provided contributions, public data sources, and 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 and longitude coordinates, type of organization, and website (when available). Factual's Resolve service was used for entity resolution to cross-validate records from other data sources and to append additional records that were identified as museums, but which not listed in any existing source (e.g., IRS, grant data, or 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.

### ***Academic/Postsecondary Data***

Using information from the FY 2013 Integrated Postsecondary Education Data System (IPEDS) survey, IMLS collected additional data about museums that were associated with academic institutions which were not previously collected or properly identified using other data sources. This process involved a systematic review of academic institution websites to identify associated museum and museum-related organizations. IPEDS data was used to identify approximately 7,700 academic institutions for initial review. This was later limited to approximately 4,800 institutions that met certain a criterion (i.e., active, degree granting, Title IV institutions that offer at least an associate's degree or higher). Using a standard protocol for identification and inclusion, approximately 2,700 active museum and museum-related organizations were identified and included in this release of the MUDF.

### ***Information from Professional Organizations and User-Contributed Data***

IMLS continued to collect user-contributed feedback from previous releases. These included reported errors or corrections provided by the public. User-contributed information is reviewed by staff and incorporated into future releases. In addition to user-contributed information, additional information was provided by professional organizations, such as the International Planetarium Society (IPS), the American Public Gardens Association (APGA), and the Association of Nature Center Administrators (ANCA).

# Museum Universe Data File (MUDF), FY 2015 Q3

This section contains the list and description of data elements available in the Museum Universe Data File, Fiscal Year 2015, Third Quarter release (MUDF FY 2015 Q3).

## ***Data File Formats***

The public-use MUDF for the FY 2015 Q3 release (filename: **mudf15q3pub**) is available in three formats: SAS (.sas7bdat), SPSS (.sav), and comma-delimited (.csv). Comma-delimited files can be opened in most spreadsheet programs. The data file has a total of 33,072 records and 44 variables. Each museum or museum-related organization consists of one record.

## ***Data Elements and Definitions***

**MID** – Unique Museum Universe Data File (MUDF) identification number

**COMMONNAME** – Common name of organization

**LEGALNAME** – Legal name of organization

**ALTNAME** – Alternate name of organization

**AKADBA** – Alternate name of organization (also known as [aka]; doing business as [dba])

**ADSTREET** – Administrative address of organization, street address

**ADCITY** – Administrative address of organization, city

**ADSTATE** – Administrative address of organization, state

**ADZIP** – Administrative address of organization, ZIP+4

**ADZIP5** - Administrative address of organization, ZIP

**PHSTREET** – Physical address of organization, street address

**PHCITY** – Physical address of organization, city

**PHSTATE** – Physical address of organization, state

**PHZIP** - Physical address of organization, ZIP+4

**PHZIP5** - Physical address of organization, ZIP

**PHONE** – Phone number of organization

**WEBURL** – Web address of organization

**DISCIPL** – Museum discipline code of organization

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 of organization

**NTEEC** - 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, May, 2015.

**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, May, 2015.

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

**INCOME** - Gross receipts from most recent Internal Revenue Service (IRS) 990 Form. Income is a computer generated amount by the IRS. Information comes from IRS Business Master File, May, 2015. For more information, please visit <https://www.irs.gov/charities-non-profits/exempt-organizations-business-master-file-extract-eo-bmf>

**REVENUE** - Total revenue from most recent Internal Revenue Service (IRS) 990 Form. Information comes from Internal Revenue Service (IRS) Business Master File, May, 2015.

**NAICS** – North American Industrial Classification System (NAICS), a classification system used by the federal government to classify business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. NAICS was developed under the auspices of the Office of Management and Budget (OMB), and adopted in 1997 to replace the Standard Industrial Classification (SIC) system. For more information about NAICS see: <http://www.census.gov/eos/www/naics/>

**IPEDS** – Unique identifier from Integrated Postsecondary Education Data System, which is the primary source for postsecondary education data (if applicable). For more information see National Center for Education Statistics <http://nces.ed.gov/ipeds/>

**INSTNAME** – Name of parent or affiliated academic institution (if applicable)

**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](https://nces.ed.gov/ccd/rural_locales.asp)

Code	Description
1	<b>City</b> – Territory inside an urbanized area and inside a principal city
2	<b>Suburb</b> – Territory inside an urbanized area but outside a principal city
3	<b>Town</b> – Territory inside an urban cluster that is outside an urbanized area
4	<b>Rural</b> – 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

**BEAREG** – Bureau of Economic Analysis regions. For more information see U.S. Department of Commerce Bureau of Economic Analysis <http://www.bea.gov/regional/docs/regions.cfm>

Code	Name	Description
1	New England	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
2	Mid-East	Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania
3	Great Lakes	Illinois, Indiana, Michigan, Ohio, Wisconsin
4	Plains	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
5	Southeast	Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia
6	Southwest	Arizona, New Mexico, Oklahoma, Texas
7	Rocky Mountains	Colorado, Idaho, Montana, Utah, Wyoming
8	Far West	Alaska, California, Hawaii, Nevada, Oregon, 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 two integers and six decimal places, with an explicit decimal point, decimal degree format, World Geodetic System Datum 1984; determined by geocoding service.

**FIPSST** - 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.

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

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

**FULLFIPS** - Federal Information Processing Standards (FIPS) 15-digit Census Block FIPS Code based on geocoded address of latitude and longitude using FCC Block Conversions API

AABBBCCCCCDEEE

A = State (2-digit FIPS code)

B = County (3-digit FIPS code)

C = Tract (6-digit FIPS code)

D = Block Group (1-digit FIPS code)

E = Block (3-digit FIPS code)

**CONGDIST** – Congressional District. ANSI code based on the location of the administrative entity/outlet. Legislatively defined subdivisions of the state for the purpose of electing representatives to the House of Representatives of the U.S. Congress.

**GSTREET** – Geocoded address of organization, street address

**GCITY** – Geocoded address of organization, city

**GSTATE** – Geocoded address of organization, state

**GZIP** - Geocoded address of organization, ZIP+4

**GZIP5** - Geocoded address of organization, ZIP

**BMF15\_F** – Internal Revenue Service's Business Master File May 2015 flag. This flag indicates that this record was found in the most recent IRS 990 Business Master File data. The IRS Business Master File (BMF) contain descriptive information for all active organizations (public charities, private foundations, etc) that have registered for tax-exempt status for the IRS.

0 – No, EIN not found in current Business Master File

1 – Yes, EIN found in current Business Master File