

Memorandum

Department of Veterans Affairs

Date: **FEB 26 2014**

From: Acting Assistant Secretary for Policy and Planning (008)

Subj: Open Data – Quarterly Data Asset Collection Requirements (VAIQ 7423360)

To: Under Secretaries, Assistant Secretaries, and Other Key Officials

1. In support of Open Data efforts, this memorandum provides guidance regarding the Department's forthcoming data asset collection efforts.
2. Issued in May 2013 (and periodically enhanced since then), the Office of Management and Budget's (OMB) Open Data Policy calls for each Federal agency to manage data as an asset and make data available to the public when allowable. OMB requires each Federal agency to establish, populate, and maintain an enterprise-wide data asset inventory by November 30, 2014. To ensure progress, OMB established the following quarterly data asset collection reporting deadlines: February 28, 2014; May 31, 2014; August 30, 2014; and November 30, 2014.
3. This effort is being discussed on a weekly basis during the Department's Open Data Integrated Project Team meetings and during the monthly Data Governance Council meetings. Members from your organizations actively participate in both forums.
4. Please review and concur on the attached documents by ~~24 February 2014~~ *27 February 2014*.
5. If you have any questions, please have a member of your staff contact Mr. Joe Salvatore, Director, Reports and Information Service at (202) 266-4685 or Joe.Salvatore@va.gov.


Robert D. Snyder

Attachments

OVERVIEW

Quarterly Data Asset Collection Effort

To ensure that each organization effectively contributes to the Department of Veteran Affairs (VA) quarterly data asset collection efforts, the following memorandum provides guidance regarding the Department's forthcoming data asset collection efforts.

The memorandum is supported by the following eight attachments.

Attachment 1	Data Assets
Attachment 2	Operating Guidance
Attachment 3	Metadata Fields
Attachment 4	Bureau Codes
Attachment 5	Program Codes
Attachment 6	Frequently Asked Questions
Attachment 7	Key Definitions
Attachment 8	Key References

The major output of this effort will be the generation of the Department's enterprise-wide data asset inventory. A subset of the inventory, known as the public data asset listing, will be available to the public. The inventory is no more than a comprehensive listing of select data assets all associated characteristics known as metadata.

Under no circumstance does metadata include personally identifiable information (PII) or protected health information (PHI). Nor will any databases or data sets containing PII or PHI be uploaded.

The maximum number of metadata fields that each data steward or point of contact will need to provide for each data asset is 29. Some types of data assets to include statistical or analytical reports will require far less. Typical metadata fields include items such as title, description, and data steward/point of contact information. For those organizations that inputted metadata into Data.gov, the metadata for Open Data is virtually the same.

Only Federal employees will have access to the entire enterprise-wide data asset inventory. The public will only see those entries classified with a "public" access level via the submission of a hyperlink. This subsection of the inventory is known as the public data asset listing.

Note: This activity incorporates all data assets submitted via the now-decommissioned Data.gov effort.

Please do not hesitate to contact Mr. Joe Salvatore if your staff has questions or concerns.

E-mail: joe.salvatore@va.gov
Telephone: (202) 266-4685

ATTACHMENT 1

Data Assets

In order to ensure that all Administrations and Staff Offices approach the OMB reporting requirements in a consistent fashion, the following narrative explores the meaning of a data asset for VA purposes. Additionally, the document provides an outline of high-level requirements and boundaries for the collection efforts.

Data Asset:

The term “data asset” can mean different things to different people. As defined by OMB’s Project Open Data, a data asset is essentially “a collection of data elements or datasets that make sense to group together.” For VA purposes, a data asset is a work product generated by a VA employee or VA-affiliated entity which possesses one or more of the following attributes:

- Contains a collection of record-level data (data about an individual) and/or aggregated data (a form of information) generated in support of the Department’s mission;
- Structured in composition and done so in a purposeful and meaningful way;
- Encased in one of many file formats to include but not limited to .doc, .xls, .csv, or .xml;
- Presented in one of many mediums to include but not limited to electronic, digital, or paper;
- Composed primarily of numerics (record-level data or aggregated information) and/or numerics with descriptive narratives such as in a report;
- Physically housed either within VA’s domain (Intranet, data warehouse, bookshelf etc.) or in the public domain (Internet);
- May or may not have a data dictionary or system of records notice (SORN);
- May or may not be releasable in part or in whole under provisions of the Freedom of Information Act (FOIA) and/or the Privacy Act;
- May require permanent information security safeguards and dissemination restrictions;
- Can consist exclusively of internal VA data/information or partial VA data/information merged with another agency’s records; and
- Was not generated in its entirety by another Federal agency or other government entity.

ATTACHMENT 1 *continued*
Data Assets

Data Asset Inventory:

The enterprise-wide data asset inventory consists of two major components: a public data asset listing and a restricted data asset listing. A data asset's inclusion in either listing is based upon the access level (public, restricted public, or non-public) which is determined, assigned, and submitted into the inventory's metadata by the Administration and Staff Office of jurisdiction. For the public data asset listing, only those data assets which are or could be made publicly available to all without restrictions are included. Please reference Attachment 7 for complete definitions of these access levels.

Collection Boundaries:

As noted in the above, a data asset can have many characteristics. That said, VA is not seeking to include all types of data assets in the enterprise-wide data asset inventory at this time. To assist Administrations and Staff Offices in their efforts, the below narratives and tables contain guiding parameters.

a. Public Data Asset Listing

The public data asset listing should only contain vetted and official data assets generated by each Administration and Staff Office and assigned a "public" access level. These products may or may not be sourced entirely with agency information. Data assets generated by VA-affiliated entities (i.e., contracted vendors) are permissible so long as approved for release by the organization.

Included Data Assets	Excluded Data Assets
Non-reidentifiable statistical data	Record-level data
Statistical reports	Personally Identifiable Information (PII) data
Analytical reports	Protected Health Information (PHI) data
Geographic and Information System (GIS) maps (no latitude and longitude data)	Sensitive data
Master non-reidentifiable aggregated/statistical files	Classified data
Official data and information (fully vetted)	National security data
	Re-identifiable statistical data
	GIS products with latitude and longitude data
	Extracts or snapshots of record-level data
	Non-official data and information (draft, pre-decisional, or not vetted)
	Created exclusively by another Federal agency or other government entity
	Any data asset placed in the restricted public data asset listing

ATTACHMENT 1 *continued*
Data Assets

b. Restricted Data Asset Listing

The restricted data asset listing should only contain vetted and official data assets generated by each Administration and Staff Office and either assigned a “restricted public” or “non-public” access level. These products may or may not be sourced entirely with agency information. Data assets generated by VA-affiliated entities (i.e., contracted vendors) are permissible so long as approved for release by the organization.

Included Data Assets	Excluded Data Assets
Record-level data	Non-reidentifiable statistical data
PII data	Statistical reports
PHI data	Analytical reports
Sensitive data	Extracts or snapshots of record-level data
Re-identifiable data	Non-official data and information (not fully vetted)
GIS products with latitude and longitude data	Any data asset placed in the public data asset listing
Master data files	Created exclusively by another Federal agency or other government entity.
Official data and information (fully vetted)	

ATTACHMENT 2

Operating Guidance

To ensure that each organization effectively contributes to the Department's quarterly data asset collection efforts, the following guidance is put forward as reference.

Team Organization

1. Select an organizational lead (s) to spearhead your organization's efforts.
Note: VAIQ 7415689 solicited points of contact leads from each Administration and Staff Office.
2. Issue internal communications outlining the forthcoming data collection effort.
3. Form an internal committee consisting of representatives from various business lines as well as Privacy Act, Freedom of Information Act (FOIA), and Information Security Officers.
4. Identify an organizational data asset coordinator, if different from organizational lead.

Pre-Collection Activities

1. Comprehend definitions for data asset and Mosaic Effect. Reference Attachment 1, *Data Assets*, and Attachment 7, *Key Definitions*.
2. Identify possible locations of known data assets holdings (e.g., data warehouses, FOIA reading rooms, SMART database, intranet and internet webpages, etc.)
3. Examine system of records notices (SORN) and privacy impact assessments (PIAs) for listings of databases and datasets.
4. Secure familiarity with required variables and associated definitions. Reference Attachment 3, *Metadata Fields*.
5. Ensure accessibility to both the forthcoming Open Data website (URL to be disseminated under separate cover) and centralized server location for all metadata documents (e.g., data dictionaries, system of records notices, and the occasional supporting technical documents classified as public). This location will permit the public to access the documents via a generated hyperlink/uniform resource locator (URL).
6. Establish internal ability to generate metadata hyperlinks/URLs for the public data asset listing.
7. Utilize VA's data asset collection tool via the following Veterans Health Administration's (VHA)-maintained SharePoint site.

<https://vaww.dwh.cdw.portal.va.gov/Regions/NDS/Open%20Data%20Initiative/Forms/AllItems.aspx>.

8. Have the organizational data asset coordinator (s) secure edit capability by sending an e-mail to Robert Street at robert.street2@va.gov or Susan Hickey at susan.hickey@va.gov.

ATTACHMENT 2 *continued* **Operating Guidance**

Collection Activities

1. Recommend initially supplementing the VHA SharePoint site tool with available data asset information located in existing compendiums (e.g., VHA's Corporate Monograph, National Center for Veterans Analysis and Statistics' website content tracker etc.).
2. Commence search for additional data assets and input only if an entry is not present.
3. Analyze output of the Office of Information Technology's (OIT) forthcoming search.
4. Assign data stewards/point of contact for each entry and have that individual/organization responsible for filling in remaining content and validating existing content.

Note: All collected data assets are not official inventory entries, but merely nominations, which will be included once final approval is secured.

Post-Collection Analysis

1. Have designated Privacy Officer and other designated personnel review the listed data assets nominations - particularly those items classified with a "public" access level - on a running basis for privacy and information security concerns to include findings of PII and PHI as well as the re-identification of de-identified data via the mosaic effect. Reasons for not releasing the data asset to the public will be placed in "Access Level Comment" metadata field. Reference Attachment 3.
2. Conduct necessary level of reviews within organization.

Note: Following each organizational layer of review, a designated approval flag or indicator should be applied.

Final Internal Review

1. Submit cleared data assets to organizational lead for review by the Open Data Integrated Project Team (IPT).
2. Await clearance and approval from Open Data IPT.
3. Receive e-mail approval from the Open Data IPT Chair.

ATTACHMENT 2 *continued* **Operating Guidance**

Dissemination

1. VHA – Create JSON file and submit to OIT.
2. OIT Architecture, Strategy, and Design (ASD) - Upload the approved quarterly VA data asset inventory, convert to JSON, and upload to OMB MAX.
3. OIT ASD and OPIA – Take the public data asset listing section of data asset inventory, generate a separate JSON file, and post to www.va.gov/data.json.

Documentation Adjustments

1. Office of Policy and Planning (OPP) only – Make adjustments to narratives and the six points of contact on an as needed basis.
2. OIT and Office of Policy and Intergovernmental Affairs (OPIA) only – Upload any changed narrative text or points of contact to digital strategy page.

ATTACHMENT 3 Metadata Fields

For VA purposes, the VHA SharePoint Site breaks down the 29 metadata fields into two distinct categories: “required” fields and “optional” fields. This breakout provides alignment with OMB’s three metadata breakouts as follows.

VA Metadata Categories	OMB Metadata Categories	OMB Metadata Definitions
Required (18)	Common core required	A critical metadata field which is mandatory for all data assets regardless of the designated access level.
	Common core required, if applicable fields.	A critical metadata field which is mandatory should the data asset have the outlined metadata properties. For example, a statistical report will not have a data dictionary, but a database regardless of the designated access level should have a completed one. As such, a URL will be entered for the database’s data dictionary and the statistical report entry will be blank.
Optional (11)	Expanded	An optional metadata field which provides additional descriptive information about the data asset.

Note: All data assets will not have SORNs or data dictionaries. But when available, those fields are required.

Definitions for each metadata category and metadata field follow. For reference, the former Data.gov names have been provided. *Note:* Some business lines may elect to capture additional information above and beyond the OMB requirements below.

REQUIRED FIELDS:

#	Field	Definition	OMB Tag	Data.gov Tag
1	Title	Human readable name of the asset. This should be in plain English and include sufficient detail to facilitate search and discovery.	title	Title
2	Description	Human readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether an asset is of interest. No more than six sentences.	description	Description
3	Tags	Keywords that help users discover your dataset. Terms used by technical and non-technical users	keyword	Keywords

		should be included.		
4	Last Update	Most recent date on which the datasets was changed, updated, or modified.	modified	Date updated
5	Publisher	Name of the publishing entity.	publisher	Agency Name
6	Contact Name	Contact person's first name and last name.	contactPoint	Contact Name
7	Contact E-mail	Contact person's e-mail address.	mbox	Contact E-Mail address
8	Unique Identifier	A unique identifier for the data asset as maintained within an agency catalog or database. For instance, ASI is the designed unique identifier for VHA's Addiction Severity Index. The standard protocol is VA-<Administration Name or Staff Office Acronym>-<Program Name>-00X where 00X starts with 001.	identifier	User Generated ID
9	Public Access Level	The degree to which this data asset could be made publicly available, regardless of whether it has been made available. Options are: "Public" (data asset is or could be made available to all without restricts); "Restricted Public" (data asset is available under certain use restrictions); and "Non-Public" (data asset is not available to members of the public). Reference Attachment 7 for detailed definitions.	accessLevel	n/a
10	Access Level Comment	An explanation which supports the access level designation for the particular data asset. That is, why was the data asset classified as "non-public" or "restricted public" rather than "public." Also include instructions for how to access a restricted file, if applicable.	accessLevelComment	n/a

11	Bureau Code	The federal agency bureau code from OMB Circular A-11, Appendix C. Format is XXX:XX. Reference the bureau codes located in Attachment 4.	bureauCode	n/a
12	Program Code	The OMB-defined code related to the data asset. Format XXX:XXX. Reference the program codes located in Attachment 5.	programCode	n/a
13	Download URL	The URL which provides direct access to the downloadable distribution of a data asset. This only applies to data assets classified with a "public" access level.	accessURL	Access Point
14	Format	The file format or API type of the distribution. This must describe the exact files available to access URLs using Multipurpose Internet Mail Extension types. Examples include but are not limited to the following: .csv; and .xml. Do not use .doc.	format	Media Format
15	Endpoint	The endpoint for an API. This represents the endpoint of web service to access a data asset.	webService	Access Point
16	License	The license each data asset (dataset, API, and select other) is published with.	license	Dataset license agreement URL
17	Spatial	The range of spatial applicability of a data asset (dataset, API, and select other). This could include a spatial region such as a bounding box or a named place.	spatial	Geographic scope
18	Temporal	The range of temporal applicability of a data asset or the start and end date for applicability for the data. For example, a FY2011-FY2012 report would be recorded as 2011/2012.	temporal	Period of coverage

OPTIONAL FIELDS:

#	Field	Definition	OMB Tag	Data.gov Tag
19	Release Date	The date of the data asset’s formal issuance.	issued	Date Released
20	Frequency	The frequency with which the data asset is published. Options include but are not limited to daily, weekly, bi-weekly, monthly, bi-monthly, and annually.	accrualPeriodicity	Frequency
21	Language	The language of the data asset. All entries should adhere to the RFC 5646 standard. Example: A language tag is comprised of either one or two parts, the language subtag (such as “en” for English; “sp” for Spanish and “wo” for Wolof) and the regional subtag (such as “US” for United States, GB for Great Britain, and MX for Mexico), separated by a hyphen. Format is language subtag-regional subtag. Regional sub-tags should only be provided when needed to distinguish a language tag from another one (such as American versus British English).	language	n/a
22	Data Quality	A determination as to whether the data asset (dataset, API, and select other) meets VA’s information quality guidelines. Response is a “Yes” or “No.”	dataQuality	Data Quality
23	Distribution	Holds multiple download URLs for datasets composed of multiple files and/or file types.	distribution	n/a
24	Category	The main thematic category of the data asset (dataset, API, and select other). If multiple categories are listed, a comma must separate each category.	theme	Subject Area
25	Data Dictionary	The URL to the data dictionary for the designated data asset. Since not all data assets have data dictionaries, this applied to those datasets, APIs, and the like which are classified with a “public” access level. All data assets must be certified by the business owner as 100 percent accurate, complete, and validated. If not, the data dictionary cannot be uploaded. Note: Documentation other than a data dictionary can be referenced using the “Related Documents” field in the Expanded Section. Important - Please be very judicious, strategic, and ever mindful	dataDictionary	Data Dictionary

		of information security/privacy issues when providing additional documents.		
26	Related Documents	In addition to the earlier provided data dictionary and document download URLs, this opportunity permits URLs for related documentation such as technical information about the data asset (dataset, API, and select other), developer documentation, etc. The provided URL must be enclosed within strings. Additionally, multiple URLs must be separated with a comma. Important - Please be very judicious, strategic, and ever mindful of information security/privacy issues when providing this documentation.	references	Reference for Technical Documentation
27	Homepage URL	The alternative landing page used to redirect user to a contextual, agency-hosted "homepage" for the data asset (dataset, API, and select other) when selecting this resource from the Data.gov user interface.	LandingPage	n/a
28	Primary IT Investment UII	For linking a dataset with an IT Unique Investment Identifier (UII). UII's are OMB Exhibit 300 Line Items for Major and Non-Major Investments. There are a total of 24 Major Investments and 9 Non-Major Investments. The Major UII's are broken into 4 categories: Medical or Health, Benefits and Memorials, Corporate, and Interagency. The codes will be listed in the SharePoint site.	PrimaryITInvestmentUII	n/a
29	System of Records	If the data asset is a system and is designated as a system of records under the Privacy Act of 1974, the URL to the System of Records Notice or SORN needs to be provided. Some but not all agency SORNs are posted to the following VA webpage: http://www.rms.oit.va.gov/sor_records.asp <i>Note:</i> If the provided SORN was published more than three years ago and/or has incurred changes since publication, please commence efforts to refresh the SORN.	systemOfRecords	n/a

ATTACHMENT 4 Bureau Codes

For each data asset, there is a metadata field that deals with bureau codes. (VA's equivalent of a bureau is either an Administration or Staff Office.) These codes were established by OMB and are used to partially evaluate participation and progress with the data asset inventory effort.

VA Bureau Codes

OMB Nomenclature	Agency Code	Bureau Code
Department of Veterans Affairs	029	00
Veterans Health Administration	029	15
Benefits Programs	029	25
Departmental Administration	029	40

Reference: OMB Circular A-11, Appendix C, Page 9

http://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/app_c.pdf.

VA Bureau Codes with Supporting OMB-Identified Programs

According to OMB records, the designated sub-components of each OMB nomenclature are outlined below. For those Staff Offices not identified below, please default to "Department of Veterans Affairs" nomenclature.

Veterans Health Administration (029:15)

- Medical Services
- Medical Support and Compliance
- Medical Facilities
- Medical and Prosthetic Research
- Joint DoD-VA Medical Facility Demonstration Fund
- Medical Care Collections Fund
- Medical Care Collections Fund (Legislative prop.)
- Canteen Service Revolving Fund
- Medical Center Research Organizations

Benefits Programs (029:25)

- Compensation And Pensions
- C&P Legislative proposal
- Readjustment Benefits
- Readjustment Benefits (Legislative proposal)
- Veterans Insurance and Indemnities
- Filipino Veterans Equity Compensation Fund
- Service Disabled Veterans Insurance Fund
- Veterans Reopened Insurance Fund
- Servicemembers' Group Life Insurance Fund
- Veterans Housing Benefit Program Fund
- Housing Direct Loan Financing Account
- Housing Guaranteed Loan Financing Account

ATTACHMENT 4 *continued*
Bureau Codes

Housing Liquidation Account
Vocational Rehabilitation Loans Program Account
Native American Direct Loan Financing Account
Transitional Housing Direct Loan Financing Account
Vocational Rehabilitation Direct Loan Financing Account
Post-Vietnam Era Veterans Education Account
National Service Life Insurance Fund
United States Government Life Insurance Fund
Veterans Special Life Insurance Fund

Departmental Administration (029:40)

Construction, Major Projects
Construction, Minor Projects
Grants for Construction of State Extended Care Facilities
Grants for Construction of Veterans Cemeteries
Veterans Employment and Infrastructure Enhancement Transfer
Fund
General Administration
General Operating Expenses, VBA
Office of Inspector General
Information Technology Systems
National Cemetery Administration
Supply Fund
Franchise Fund

ATTACHMENT 5 Program Codes

For each data asset, there is a metadata field that deals with program codes. These codes were established by OMB and are used to partially evaluate participation and progress with the data asset inventory effort.

In the instance where an existing Administration or Staff Office program isn't listed or doesn't properly correlate to the specific OMB nomenclature, please default to the organizational program code.

Reference: Performance.gov.

http://goals.performance.gov/sites/default/files/images/FederalProgramInventory_FY13_MachineReadable_091613.xls

Program Codes by Administration/Staff Office with Mail Routing Symbol

Office of the Secretary (00)

OMB Nomenclature	Program Code
Office of the Secretary	029-081

Board of Veterans' Appeals (01)

OMB Nomenclature	Program Code
Board of Veterans' Appeals	029-082

Office of the General Counsel (02)

OMB Nomenclature	Program Code
Office of General Counsel	029-083

Veterans Health Administration (10)

OMB Nomenclature	Program Code
Inpatient Care	029-040
Ambulatory Care	029-041
Mental Health Service, General Outpatient Care	029-042
Mental Health, Intensive Recovery-Oriented Program	029-043
Mental Health, Inpatient Care	029-044
Mental Health Residential Rehabilitation Treatment Programs (MH RRTP)	029-045
Readjustment Counseling	029-046

Pharmacy	029-047
VHA Research and Development	029-048
Spinal Cord Injury and Disorders (SCI/D) Service	029-049
Blind Rehabilitation Service (BRS)	029-050
Dental Care	029-051
Prosthetics and Sensory Aids Service (PSAS)	029-052
Women Veterans Health Care	029-053
Community Living Centers (CLCs)	029-054
Community Nursing Homes	029-055
State Veterans Home Program	029-056
Adult Day Health Care	029-057
Home Based Primary Care	029-058
Homemaker and Home Health Aide Care	029-059
Hospice and Palliative Care	029-060
Program of All Inclusive Care of the Elderly (PACE)	029-061
Respite Care	029-062
Skilled Home Health Care	029-063
Program of Comprehensive Assistance for Family Caregivers	029-064
Telehealth Care	029-065
Civilian Health and Medical Program of the Department of Veterans Affairs (CHAMPVA)	029-066
Foreign Medical Program (FMP)	029-067
Spina Bifida Program	029-068
Children of Women Vietnam Veterans (CWVV)	029-069
Department of Housing and Urban Development-VA Supportive Housing (HUD-VASH)	029-070
Grant and Per Diem (GPD) Program	029-071
Health Care for Homeless Veterans (HCHV)	029-072
Supportive Services for Low Income Veterans and Families (SSVF)	029-073
National Call Center for Homeless Veterans (NCCHV)	029-074
Veterans Justice Outreach (VJO) Program	029-075
Homeless Veterans Supported Employment Program (HVSEP)	029-076
National Homeless Registry	029-077

Veterans Benefits Administration (20)

OMB Nomenclature	Program Code
Veterans Benefits – Administration of Benefits	029-003
Burial Benefits	029-004
Veterans Disability Compensation	029-005
Survivors Compensation	029-006
Dependency Indemnity Compensation (DIC)	029-007
Filipino Veterans Compensation	029-008
Veterans Pension	029-009
Survivors Pension	029-010
Fiduciary Services	029-011
Veterans Mortgage Life Insurance	029-012
United State Government Life Insurance	029-013
National Service Life Insurance	029-014
Veterans' Special Life Insurance	029-015
Veterans Reopened Insurance	029-016
Service-Disabled Veterans' Insurance	029-017
Servicemembers' Group Life Insurance (SGLI)	029-018
Family Servicemembers' Group Life Insurance (FSGLI)	029-019
Servicemembers' Group Life Insurance Traumatic Injury Protection (TSGLI)	029-020
Veterans' Group Life Insurance (VGLI)	029-021
Housing Guaranteed Loan	029-022
Acquired Direct Loan	029-023
Vendee Direct Loan	029-024
Guaranteed Loan Sale Securities	029-025
Native American Direct Loan	029-026
Specially Adapted Housing (SAH) & Special Housing Adaptation (SHA)	029-027
Vocational Rehabilitation Loan	029-028
Transitional Housing Loan	029-029
Post-9/11 GI Bill	029-030
Montgomery GI Bill – Active Duty	029-031
Montgomery GI Bill – Selected Reserve	029-032
Survivors' and Dependents' Educational Assistance	029-033
Reserve Educational Assistance Program	029-034
Veterans Retraining Assistance Program	029-035
Post-Vietnam Era Veteran's Educational Assistance Program	029-036
Vocational Rehabilitation and Employment Program	029-037
Vocational Rehabilitation and Employment –	029-038

VetSuccess On Campus (VSOC)	
Vocational Rehabilitation and Employment – Independent Living (IL)	029-039

National Cemetery Administration (40)

OMB Nomenclature	Program Code
Memorial Services	029-001
Veterans' Cemetery Grants	029-002

Office of the Inspector General (50)

OMB Nomenclature	Program Code
Office of the Inspector General	029-091

Office of the Deputy Secretary (001)

OMB Nomenclature	Program Code
Office of the Secretary	029-081

Office of the Chief of Staff (00A)

OMB Nomenclature	Program Code
Office of the Secretary	029-081

Office of Acquisitions and Logistics (001AL)

OMB Nomenclature	Program Code
Office of Acquisitions, Logistics and Construction	029-090
Construction – Major Projects	029-092
Construction – Minor Projects	029-093

Office of Public and Intergovernmental Affairs (002)

OMB Nomenclature	Program Code
Office of Public and Intergovernmental Affairs	029-088

Office of Management (004)

OMB Nomenclature	Program Code
-------------------------	---------------------

Office of Management	029-084
----------------------	---------

Office of Information and Technology (005)

OMB Nomenclature	Program Code
Operations and Maintenance	029-078
Development	029-079
Staffing and Administration	029-080

Office of Human Resources and Administration (006)

OMB Nomenclature	Program Code
Office of Human Resources and Administration	029-085

Office of Security and Preparedness (007)

OMB Nomenclature	Program Code
Office of Security and Preparedness	029-087

Office of Policy and Planning (008)

OMB Nomenclature	Program Code
Office of Policy and Planning	029-086

Office of Congressional and Legislative Affairs (009)

OMB Nomenclature	Program Code
Office of Congressional and Legislative Affairs	029-089

ATTACHMENT 6

Frequently Asked Questions

For the purposes of assisting Administrations and Staff Officers with VA's Open Data efforts, the following frequently asked questions are outlined.

General

What is Open Data? An actionable government-wide initiative aimed at having agencies manage information as an asset across all executive departments and agencies. By managing government information as an asset, the desired outcome is to increase operational efficiencies, reduce costs, improve services, support mission needs, safeguard personal information, and increase public access to government information. Open Data guidance was officially disseminated on 9 May 2013 with the release of [Executive Order 13642](#), *Making Open and Machine Readable the New Default for Government Information* and [OMB Memorandum M-13-13](#), *Open Data Policy-Managing Information as an Asset*. Subsequent guidance is routinely released via a combined OMB and Office of Science and Technology Policy (OSTP) resource known as Project Open Data (<http://project-open-data.github.io/>)

What will Open Data do for VA? Open Data will collect or create information in a way that supports downstream information processing and dissemination. This will be accomplished in part because VA data assets will be in machine-readable and open formats, use enterprise-wide data standards, and have common core metadata for all newly created information. Additionally, Open Data will ensure information stewardship through the use of open licenses and the review of information for privacy, confidentiality, security, or other restrictions on release. Finally, Open Data will build or modernize its information systems so that interoperability and information accessibility is maximized, an inventory for all internal and external data assets can be maintained, information safeguards are enhanced and information management responsibilities are clearly defined.

What is VA's Data Asset Collection effort? An internal VA effort to identify, collect, record, and review all of the Department's data assets and associated metadata using a standardized OMB reporting format by November 30, 2014. All collected data assets will be stored in VA's enterprise-wide data asset inventory and uploaded to OMB's Government-wide inventory. A small portion of all VA data assets will be available to the public – to the extent permissible by law - via the inventory's public data asset listing.

What is a data asset? The term "data asset" can mean different things to different people. As defined by OMB's Project Open Data, a data asset is essentially "a collection of data elements or datasets that make sense to group together." For VA purposes, a data asset is a work product generated by a VA employee or VA-affiliated entity which possesses one or more of the following attributes: 1) contains a collection of record-level data (data about an individual) and/or aggregated data (a form of information) generated in support of the Department's mission; 2) structured in composition and done so in a purposeful and meaningful way; 3) encased in one of many file formats to include but not limited to .doc, .xls, .csv, or .xml; 4) presented in one of many mediums to include but not limited to electronic, digital, or paper; 5) composed primarily of numerics (record-level data or aggregated information) and/or numerics with descriptive narratives such as in a report; 6) physically housed either within VA's domain (Intranet, data warehouse, bookshelf etc.) or in the public domain (Internet); 7) may or may not have a data dictionary or system of records notice; 8) may or may not be releasable in part or in whole under the Freedom of Information Act (FOIA) and Privacy Act; 9) may require permanent information

ATTACHMENT 6 *continued*

Frequently Asked Questions

security safeguards and dissemination restrictions; 10) can consist exclusively of internal VA data/information or partial VA data/information merged with another agency's records; and, 11) was not generated in its entirety by another agency.

Enterprise- Wide Data Asset Inventory

What is the Enterprise-Wide Data Asset Inventory? Within VA, the major output of the Open Data activity will be the generation of a comprehensive listing of all data assets called the Enterprise-Wide Data Asset Inventory. By using a form hosted on a SharePoint platform, this inventory will serve as the repository for all identified VA data assets. Select metadata for each Administration and Staff Office's data asset will populate the fields contained there within.

Where will the Enterprise-Wide Data Asset inventory be stored? Both VA and OMB will have copies of the Inventory. VA will maintain a copy of all information that will be submitted to OMB via quarterly data collection uploads. The VA uploads will be appended with other Federal agency submissions. The General Services Administration (GSA) will be hosting and managing OMB's Government-wide inventory known in some quarters as the data catalog or CKAN.

Will the public have access to the entire Enterprise-Wide Data Asset Inventory? The public will only have access to a subset of the inventory known as the public data asset listing. These are data assets which have an access level of "public."

Will OMB have access to the entire Enterprise-Wide Data Asset Inventory? Regardless of the access level designation, both OMB and other Federal employees will have access to the entire inventory.

Will databases or data sets be included in the Enterprise-Wide Data Asset inventory? Under no circumstances will the metadata contain URLs to record-level data (information about a veteran). That said, characteristics (metadata) about a database or data set that contains record-level data will be provided. The composition and structure of the inventory prevents document and file uploads.

Metadata

What is metadata? Metadata represents the characteristics of the data asset. The maximum number of metadata fields that a designated point of contact will need to provide for each data asset is 29. Some types of data assets to include statistical or analytical reports will require far less. Typical metadata fields include items such as title, description, and data steward/point of contact information. For those organizations that inputted metadata into Data.gov, the metadata for Open Data is virtually the same. Reference the full definition in Attachment 7.

ATTACHMENT 6 *continued*

Frequently Asked Questions

Does the metadata contain a field to discriminate the availability of the data asset to the public? Yes, the Inventory calls for the designated point of contact to complete a field known as “access level” which represents the degree to which a data asset could be made publicly-available, regardless of whether it has been made available. Options are as “Public” (data asset is or could be made available to all without restricts); “Restricted Public” (data asset is available under certain use restrictions); and “Non-Public” (data asset is not available to members of the public). Reference Attachment 7 for detailed definitions.

Does the actual file for a data asset described in the public data asset listing need to be made publicly available? Yes, the goal of the public data asset listing is to make these data assets available via a URL and thus eliminate the need for a formal request such as a Freedom of Information Act (FOIA) request.

Will the Database Architecture Repository (DAR) be involved with this effort? The DAR will be utilized to capture physical data schemas identified during the data asset collection efforts.

Data

Will data assets containing record-level data or data about an individual be uploaded and/or stored in the Enterprise-Wide Data Asset Inventory? Under no circumstances will this type of sensitive data be accessible to anyone regardless of designated access level. The design of the Inventory prevents such uploads. What is permissible is for select documents – after only having been reviewed and approved via multiple layers of concurrence – to be available via URLs made available to the public. These documents will be of the sort that describes the data asset to include a data dictionary, SORN, and/or a technical document.

Documents

Will documents or files be uploaded into the Inventory? No. That said, documentation such as data dictionaries and salient technical documents (if necessary) will be made available via URLs.

Where will these documents be stored? Work products - represented by a URL - will reside on a forthcoming Open Data website (URL to be distributed under separate cover) and centralized VA server location accessible by the responsible Administration or Staff Office. In doing so, the public will be able to access all documents with “Public” access levels as well as other documents associated with a “Restricted Public” or “Non-Public” data asset such as a SORN or data dictionary.

Who creates the necessary URL links? Each Administration and Staff office will be responsible for generating its own URLs.

ATTACHMENT 6 *continued*

Frequently Asked Questions

What is machine-readable and open format? Information or data that is in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost. Source: <http://project-open-data.github.io/glossary/>

What are the best formats to use? Most generic formats provide an acceptable level of openness to include but not limited to .csv; .xml; and .json. While ISO 32000 is an open standard, the Portable Document Format (PDF) does not achieve the same level of openness.

Information Release

Can the entire inventory be released under FOIA? Yes.

Who will be responsible for responding to the received FOIA? The responsible program as designated by the data asset's program code.

Open Data Initiative Coordination

What organization is responsible for coordinating the Department's Open Data efforts? On an interim basis, the OPP is the responsible element. The effort is being facilitated by the Open Data Integrated Project Team which is operating under the auspices of the Department's Data Governance Council. The executive lead for the effort is the Chief Technology Officer.

Who is the lead for the Open Data IPT and the Data Governance Council (DGC)? The Open Data IPT lead is Joe Salvatore and the DGC Chair is Mr. Dat Tran, the Deputy Assistant Secretary for Data Governance and Analysis.

Do OMB and OSTP provide direction? In addition to the released Executive Order and OMB Memorandum, VA receives updated initiative information via the Project Open Data website and participation in bi-weekly Open Data meetings. Additionally, the effort is occasionally discussed at a strategic level during the respective sessions of the Chief Financial Officer Council and Chief Information Officer Council.

How can I get involved with the Open Data effort? After clearing involvement with your supervisor first, please contact your organizational lead. If you do not know your lead, please e-mail Mr. Joe Salvatore at joe.salvatore@va.gov.

ATTACHMENT 6 *continued* **Frequently Asked Questions**

Reporting

When an individual data asset is identified, where do I store it? Each Staff Office and Administration must individually input the data asset and associated metadata in the VHA-hosted SharePoint site. This will require that each Staff Office and Administration secure “edit” access.

<https://vawww.dwh.cdw.portal.va.gov/Regions/NDS/Open%20Data%20Initiative/Forms/AllItems.aspx>

Rather than individually inputting the data asset, can my organization upload a quarterly spreadsheet to the SharePoint site? Each identified data asset must be individually inputted into the SharePoint site.

Who do I contact if I have a technical problem with Open Data SharePoint site input?

Please send an e-mail to Mr. Robert Street of VHA’s National Data Systems at robert.street2@va.gov.

Who do I contact if I have a non-technical question? Please contact your organizational lead for information. Otherwise, send an e-mail to Mr. Joe Salvatore (joe.salvatore@va.gov).

How does VA provide the information to OMB? On a quarterly basis, VA will electronically upload new data asset entries in a JSON file (JavaScript Object Notation or JSON is a lightweight data-exchange format that is easy to read, parse and generate) to OMB’s password-protected OMB MAX system.

Will there be multiple JSON files or one JSON file? VA will only submit one master JSON file to the OMB MAX system. Another JSON composing only of the public data asset listing will be posted to www.va.gov/data.json.

Does my organization need to generate a quarterly JSON file? No, individual Administrations and Staff Offices will not be generating a JSON.

Does my organization need a CKAN account? A CKAN account is no longer required.

Codes

What is a bureau code? Where can I find my bureau code? Definitions for a bureau code are located in Attachments 3 and 7. VA’s bureau codes are located in Attachment 4.

What is a program code? Where can I find my program code? Definitions for a program code are located in Attachments 3 and 7. VA’s program codes are located in Attachment 5.

Can I input my own code? Since all codes are OMB-prescribed, no other inputs are permissible.

ATTACHMENT 6 *continued*

Frequently Asked Questions

What if I can't find a precise program code? While Administrations for the most part have detailed program codes, there may not be a 1:1 match between Administration program and OMB program code. If the outlined codes are not suitable for your program, please default to the program codes for your parent organization.

CKAN

What is CKAN? An OMB and GSA-provided inventory tool suitable for customization which is available to agencies for their Open Data inventory efforts. Usage of this is optional, not mandatory.

Is VA using CKAN? While VA may utilize this catalog in the future, it will not be doing so right now. VA's data asset entries will populate VHA's SharePoint site, not CKAN unless done so by GSA.

Data.gov

What happened with the data assets that my organization previously inputted into Data.gov? Data assets submitted by the National Cemetery Administration, OPP, VBA, and the VHA were used to generate VA's initial Enterprise-Wide Data Asset Inventory. However, the metadata will need to be checked against the fields that were carried over by Data.gov metadata fields. For example, the required input of bureau and program codes were not Data.gov requirements.

Can I still input information into Data.gov? The oldest Dataset Management System that agency Data.gov points of contact utilized to enter data asset metadata has been decommissioned. That said, the Data.gov catalog (similar to an inventory) will remain in existence and continue to grow by harvesting the available public data asset listing (located at www.va.gov/data.json) on each agency's websites. Since the public data asset listing will grow each quarter, the activity will be repeated on a recurring basis.

Will there be two concurrent data asset repositories: Open Data and Data.gov? Because the Data.gov catalog will be harvesting each agency's data.json file, there should not be any concern about duplication. While the information is available both at Data.gov and the agency's website, there is only one definitive listing at any time of an agency's data holdings--the agency's data.json file made available on the agency's website.

ATTACHMENT 7

Key Definitions

Open Data Policy utilizes terminology that is of a technical nature and is not commonly utilized in mainstream conversations. Outlined below are some key definitions that your team will encounter as part of this quarterly data asset collection effort.

Access Levels: A metadata field assigned by the data steward which signifies the level of availability as well as release restrictions for an identified data asset in the enterprise-wide data asset inventory. The three access levels are as follows:

- a. **Public:** Data asset is or could be made publicly available to all without restrictions.
- b. **Restricted Public:** Data asset is available under certain use restrictions.

Example 1: Data asset that can only be made available to select researchers under certain conditions, because the data asset contains sufficient granularity or linkages that make it possible to re-identify individuals, even though the data asset is stripped of Personally Identifiable Information (PII) and Protected Health Information (PHI).

Example 2: Data asset that contains PII and is made available to select researchers under strong legal protections. This category includes some but not all data assets designated as Controlled Unclassified Information (CUI), consistent with Executive Order 13556.

- c. **Non-Public:** Data asset is not available to members of the public. This category includes data assets that are only available for internal use by the Federal Government, such as by a single program, single agency, or across multiple agencies.

Example 1: Data assets that are only available for internal use by the Federal Government, such as by a single program, single agency, or across multiple agencies.

Example 2: Some but not all data assets designated as Controlled Unclassified Information (CUI), consistent with Executive Order 13556. Some non-public data assets may still potentially be available to other intra-agency operating units and/or other government agencies, as discussed in OMB Memorandum M-11-02: Sharing Data While Protecting Privacy.

ATTACHMENT 7 *continued*

Key Definitions

API: An application programming interface, which is a set of definitions of the ways one piece of computer software communicates with another. It is a method of achieving abstraction, usually (but not necessarily) between higher-level and lower-level software.

Catalog: A catalog is a collection of datasets or web services.

CKAN: CKAN is a popular, open-source data catalog that can be installed independently or run as a service. CKAN provides a solution with which to begin maintaining an internal or external catalog. In addition to providing an effective means of organizing the agency's data catalog, CKAN provides several opportunities for generating the needed data.json file. CKAN provides a convenient CSV export that can then be imported into the Catalog Generator for conversion into an appropriately formatted JSON file. Alternatively, CKAN provides a JSON API of the data catalog which can then be altered to provide an automatic update for the appropriately formatted JSON file.

Data: A value or set of values representing a specific concept or concepts. Data become "information" when analyzed and possibly combined with other data in order to extract meaning, and to provide context. The meaning of data can vary depending on its context. Data includes *all* data. It includes, but is not limited to, 1) geospatial data 2) unstructured data, 3) structured data, etc.

Data Asset: For VA purposes, a data asset is an Administration or Staff Office work product which contains a collection of aggregated data and was generated in support of the their respective business function and the Department's overall mission. Additionally, the asset's aggregated data makes sense when grouped together such as in an analytical report. A data asset's formatting is not limited to one type, but rather to multiple types to include but not limited to.csv and .xml. Common work products are statistical or analytical reports, databases, or APIs. Source: VA Open Data IPT.

Dataset: A dataset is an organized collection of data. The most basic representation of a dataset is data elements presented in tabular form. Each column represents a particular variable. Each row corresponds to a given value of that column's variable. A dataset may also present information in a variety of non-tabular formats, such as an extensible mark-up language (XML) file, a geospatial data file, or an image file, etc.

Database: A collection of data stored according to a schema and manipulated according to the rules set out in one Data Modelling Facility.

Data Publication Process: A process, developed in consultation with VA's Office of General Counsel, to determine whether data assets have a valid restriction to release. The objectives of this process are to: review information for valid restrictions to public release in order to ensure proper safeguarding of privacy, security, and confidentiality of government information; document reasons why a data asset or certain components of a data asset should not be made public at this time; consult with VA's Office of General Counsel and Office of Information

ATTACHMENT 7 *continued*

Key Definitions

Security regarding the barriers identified. A general overview of this process is published on the www.va.gov/digitalstrategy page. The overview includes information on the actual process by which data is determined to have a valid restriction to release and examples of what kinds of characteristics a data asset has that leads to a determination to not release. If data cannot be released, Senior leadership must ensure a complete analysis of issues that could preclude public disclosure of information collected or created. If it is determined that the data should not be made publicly available because of law, regulation, or policy or because the data are subject to privacy, confidentiality, security, trade secret, contractual, or other valid restrictions to release, reasons must be documented in consultation with Office of General Counsel. For each data asset listed in the inventory, one of three “access levels” will be designated: public, restricted public, and non-public. Source: VA Open Data IPT.

Enterprise-wide Data Asset Inventory: Within VA, the major output of Open Data will be the generation of a comprehensive listing of all data assets called the Enterprise-Wide Data Asset Inventory. By using a form hosted on a SharePoint platform, this inventory will serve as the repository for all identified VA data assets. Select metadata for each Administration and Staff Office’s data asset will populate the fields contained there within. Source: VA Open Data IPT.

JSON: JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language.

Machine-Readable File: Refers to information or data that is in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost.

Metadata: To facilitate common understanding, a number of characteristics, or attributes, of data are defined. These characteristics of data are known as “metadata”, that is, “data that describes data.” For any particular datum, the metadata may describe how the datum is represented, ranges of acceptable values, its relationship to other data, and how it should be labeled. Metadata also may provide other relevant information, such as the responsible steward, associated laws and regulations, and access management policy. Each of the types of data described above has a corresponding set of metadata. Two of the many metadata standards are the Dublin Core Metadata Initiative (DCMI) and Department of Defense Discovery Metadata Standard (DDMS). The metadata for structured data objects describes the structure, data elements, interrelationships, and other characteristics of information, including its creation, disposition, access and handling controls, formats, content, and context, as well as related audit trails. Metadata includes data element names (such as organization name, address, etc.), their definition, and their format (numeric, date, text, etc.). In contrast, data is the actual data values such as the “US Patent and Trade Office” or the “Social Security Administration” for the

ATTACHMENT 7 *continued*

Key Definitions

metadata called “Organization Name”. Metadata may include metrics about an organization’s data including its data quality (accuracy, completeness, etc.).

Mosaic Effect: The mosaic effect occurs when the information in an individual dataset, in isolation, may not pose a risk of identifying an individual (or threatening some other important interest such as security), but when combined with other available information, could pose such risk. Before disclosing potential PII or other potentially sensitive information, agencies must consider other publicly available data -in any medium and from any source-to determine whether some combination of existing data and the data intended to be publicly released could allow for the identification of an individual or pose another security concern. Source: OMB Memorandum 13-13, <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

Open Data: For the purposes of OMB’s Open Data Policy, Open Data refers to publicly available data structured in a way that enables the data to be fully discoverable and usable by end users. In general, open data will be consistent with the following principles:¹

- *Public:* Consistent with OMB’s *Open Government Directive*, agencies must adopt a presumption in favor of openness to the extent permitted by law and subject to privacy, confidentiality, security, or other valid restrictions.
- *Accessible:* Open data are made available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched. Formats should be machine-readable (i.e., data are reasonably structured to allow automated processing). Open data structures do not discriminate against any person or group of persons and should be made available to the widest range of users for the widest range of purposes, often by providing the data in multiple formats for consumption. To the extent permitted by law, these formats should be non-proprietary, publicly available, and no restrictions should be placed upon their use.
- *Described:* Open data are described fully so that consumers of the data have sufficient information to understand their strengths, weaknesses, analytical limitations, security requirements, as well as how to process them. This involves the use of robust, granular metadata (i.e., fields or elements that describe data), thorough documentation of data elements, data dictionaries, and, if applicable, additional descriptions of the purpose of the collection, the population of interest, the characteristics of the sample, and the method of data collection.
- *Reusable:* Open data are made available under an open license that places no restrictions on their use.
- *Complete:* Open data are published in primary forms (i.e., as collected at the source), with the finest possible level of granularity that is practicable and permitted by law and other requirements. Derived or aggregate open data should also be published but must reference the primary data.
- *Timely:* Open data are made available as quickly as necessary to preserve the value of the data. Frequency of release should account for key audiences and downstream needs.

- *Managed Post-Release*: A point of contact must be designated to assist with data use and to respond to complaints about adherence to these open data requirements

Source: OMB Memorandum 13-13, Open Data Policy – Managing Information as an Asset, March 9, 2013.

Project Open Data: "Project Open Data," a new OMB and OSTP resource, is an online repository of tools, best practices, and schema to help agencies adopt the framework presented in this guidance. Project Open Data can be accessed at <http://project-open-data.github.io>. Project Open Data will evolve over time as a community resource to facilitate adoption of open data practices. The repository includes definitions, code, checklists, case studies, and more, and enables collaboration across the Federal Government, in partnership with public developers, as applicable. Agencies can visit Project Open Data for a more comprehensive glossary of terms related to open data.

Source: OMB Memorandum 13-13,
<http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

Source: Unless specified otherwise, the source for definition is Project Open Data
<http://project-open-data.github.io>.

ATTACHMENT 8 Key References

Since May 2013, the Executive Office of the President has released a series of guidance regarding its Open Data Policy. Note: Components of OMB Memorandum 13-13 have changed since it is release in May 2013. Contents of these documents have in part supported the construction of this Memorandum.

Document	Title	Release Date
Executive Order 13642	Making Open and Machine Readable the New Default for Government Information	May 9, 2013
OMB Memorandum 13-13	Open Data Policy- Managing Information as an Asset	May 9, 2013
Open Data Implementation Guide	Implementation Guide Supplemental Guidance on the Implementation of M13-13 “Open Data Policy - Managing Information as an Asset”	Continually undergoing revisions
Project Open Data Website http://project-open-data.github.io/		