

Policy and Implementation Plan for Public Access to
Scientific Publications and Digital Data from Research
Funded by the Department of Veterans Affairs

COMMENTS ON THIS PLAN MAY BE FORWARDED TO
TOM.PUGLISI@VA.GOV
IN THE VHA OFFICE OF RESEARCH OVERSIGHT

1. Purpose

This document describes policy and implementation plans for the Department of Veterans Affairs (VA) to implement the February 22, 2013, memorandum from the White House Office of Science and Technology Policy (OSTP) directing Federal agencies to support increased public access to the results of Federally-funded research.

2. Background

On February 22, 2013, OSTP directed each Federal agency with over \$100 million in annual expenditures for the conduct of research and development to develop a plan to support increased public access to the results of research funded by the Federal government, including any results published in peer-reviewed scholarly publications that are based on research that directly arises from Federal funds (*OSTP, Increasing Access to the Results of Federally Funded Scientific Research, February 22, 2013*).

On May 9, 2013, the President signed Executive Order 13642, *Making Open and Machine Readable the New Default for Government Information*. To implement the Executive Order, Office of Management and Budget (OMB) Memorandum M-13-13 instructed agencies on establishing a framework for information management at each stage of the information's life cycle to promote interoperability and openness (*OMB, Open Data Policy – Managing Information as an Asset, May 9, 2013*).

The VA Office of Policy and Planning (OPP) Data Governance and Analysis Office is charged with developing overall policy and implementation planning to adhere to the broad principles and specific requirements of Executive Order 13642 and OMB Memorandum M-13-13. Consistent with this undertaking, the VA Office of Information and Technology (OI&T) is developing an enterprise framework for information management to promote interoperability and openness. As part of this framework, VA OI&T is currently developing comprehensive inventories of VA datasets. VA's policy and implementation plan for increased public access to the results of its research must unfold within the larger context of the OPP Data Governance planning process and the OI&T enterprise information management framework.

NOTE: Effective dates and timelines for VA requirements on public access to publications and digital data from VA-funded research are provided in Section 8 of this document.

3. Mission and Values

a. Foundation. VA's mission is to fulfill President Lincoln's promise "To care for him who shall have borne the battle, and for his widow, and his orphan" by serving and honoring the men and women who are America's veterans. In fulfilling this mission, VA has adopted the Core Values of Integrity, Commitment, Advocacy, Respect and Excellence ("I CARE") to define a culture of caring for Veterans, their families and other beneficiaries. The mission of the Veterans Health Administration (VHA) is to honor America's Veterans by providing exceptional health care that improves their well-being, educating the Nation's health care providers, developing new treatments through research, and contributing to National emergency response.

Within this context, VA and the Veterans Health Administration (VHA) recognize that Veterans and the public at large have a substantive interest in accessing the results of the research that VA funds. However, VA's primary mission is to provide health care to the Nation's Veterans, and the efforts of VA OI&T are focused first on that essential mission. VHA research efforts must operate within this context, and accordingly, the provisioning of VHA research data to the public at large will necessarily need to integrate into enterprise-wide VA OI&T patient-support data storage and access initiatives. Nevertheless, VA is committed to implementing the following principle articulated by OSTP:

To the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; U.S. national, homeland, and economic security; . . . digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze (OSTP, Increasing Access to the Results of Federally Funded Scientific Research, February 22, 2013, Section 4).

b. Primary Commitment to Veterans. In implementing its policies on public access to research publications and digital research data, VA must first remain cognizant of its ethical and legal obligations to safeguard the privacy of Veterans (and VA's other research subjects) and the confidentiality of their private information, while promoting the highest quality science. This responsibility precludes unlimited public access to private information about individual research subjects. As a result, VA has carefully weighed the public benefits versus the risks of harm to Veterans in establishing the requirements for public access to data resulting from research funded by VA (see Section 5 of this plan).

VA is committed to ensuring that the final study results, including publications and digital data, of VA-funded scientific research are made available to the scientific community, industry, and the general public in machine-readable formats with the fewest constraints possible, while (i) protecting the privacy of the Veterans (and others) from whom research data are obtained and (ii) safeguarding the confidentiality of their data.

c. Stakeholder Involvement. To ensure that VA's policies on public access to research publications and digital data meet the needs of Veterans, the research community, and the

public, VA will publish a *Federal Register* announcement seeking public comment on its policy and implementation plan. VA will also seek advice and comment from the VA National Research Advisory Council (NRAC), and other stakeholders, including researchers, clinical subject matter experts in conditions pertinent to Veterans, Veterans' Service Organizations, VA's many university affiliates, libraries, publishers, users of Federally-funded research results, and civil society groups.

4. Scope and Requirements

a. Scope. All VA research must be conducted by VA employees operating within a research program authorized by VHA. Other VA Administrations (i.e., the Veterans Benefits Administration and the National Cemetery Administration) are not authorized to fund research or to conduct research outside a VHA-authorized research program.

VA has no extramural research programs and is not authorized to award grants to support research external to VA. Most VA-funded research relies upon funds specifically appropriated by Congress for mission-related research conducted by VA employees for the benefit of Veterans. These funds are administered by the VHA Office of Research and Development (ORD), which provides peer-awards for VA employee-investigators to conduct peer-reviewed biomedical laboratory research, clinical sciences research, health services research, and rehabilitation research beneficial to Veterans. In addition, certain VHA Program Offices may fund research in support of their specific program missions without utilizing VA's appropriated research funds (see Section 7.ee.).

The requirements of this *Plan for Public Access to Scientific Publications and Digital Data from Research Funded by the Department of Veterans Affairs* will apply to all VA-funded research, including all research funded by ORD and all research funded by all VHA Program Offices, as well as to all research funded by any VA entity that may acquire research funding authority in the future.

VA's policies for public access to the results of VA-funded research are, and will continue to be, subject to applicable law; agency mission; resource constraints; U.S. national, homeland, and economic security requirements; and the objectives listed the OSTP memorandum dated February 22, 2014. VA is not authorized to fund or conduct classified research. Thus, VA requirements for public access to the results of VA-funded research will apply only to the results of non-classified research.

b. Partnerships for innovation and long-term stewardship. VA research targets biomedical and behavioral investigations to benefit the health of the Nation's Veterans and shares certain commonalities with the types of research supported by the National Institutes of Health (NIH) or regulated by the Food and Drug Administration (FDA), both located within the Department of Health and Human Services (HHS), or supported by the Department of Defense (DoD).

VA's data access requirements will maximize the potential for creative and scientifically appropriate reuse to enhance value to all stakeholders; avoids unnecessary duplication of

existing mechanisms; and maximizes the impact of the Federal research investment. VA policies will strongly encourage cooperation with the private sector to improve data access and compatibility; foster the development of public-private partnerships with foundations and other organizations that sponsor research; and facilitate search and analysis of peer-reviewed scholarly publications directly arising from research funded by the Federal Government. Examples of specific actions include the following.

(1) VA has established an agreement with HHS to make the results of ORD-funded research accessible to the public through the National Library of Medicine (NLM) PubMed Central (PMC) archive, a firmly established public-private partnership that currently provides access to publications of research supported by NIH. Use of PMC ensures that members of the public can read, download, and analyze final manuscripts or final published documents in digital form. Use of PMC also ensures that texts and their associated content will be stored in nonproprietary and/or widely-distributed archival, machine readable formats; provide access to persons with disabilities in accordance with Section 508 of the Rehabilitation Act of 1973; enable interoperability with other Federal public access archival solutions and other appropriate archives; and ensure that attribution to authors, journals, and original publishers will be maintained. VA is currently negotiating with NIH PMC to provide the same public access to the results of research funded by VHA Program Offices (or by any other VA entity with funding authority) and expects to have a completed agreement in place prior to the effective dates provided in Section 8 of this plan.

(2) VA also currently requires, and will continue to require, that the results of applicable VA-funded Clinical Trials must be provided to the public through the *ClinicalTrials.gov* archive, which provides access to the results of clinical trials involving products regulated by FDA.

(3) VA will leverage its continuing public-private partnership with the Association of American Medical Colleges (AAMC) to utilize the existing *Working Group on Information Technology Security and Privacy in VA and NIH sponsored research* in the ongoing evolution of its research data access requirements and implementation mechanisms.

(4) VA will seek partnerships with HHS, NIH, FDA, and DoD to identify and share effective mechanisms (such as the NIH Commons) to make digital research data accessible to the public in a manner that optimizes search, archival, and dissemination features that encourage innovation in accessibility and interoperability, while ensuring long-term stewardship of research data results. VA will seek to establish productive partnerships with HHS and DoD to achieve economies of scale that will make it possible to fulfill its research data access objectives.

(5) Evaluation of whether or not to preserve data will take into account the relative values of long-term preservation and access versus the associated cost and administrative burden. VA notes that current costs of long term storage per se are relatively modest.

(6) VA will ensure that research data resources are fully integrated with VA OPP enterprise initiatives to comply with Executive Order 13642 on open and machine readable information

and OMB Memorandum M-13-13 on establishing a framework for information management at each stage of the information's life cycle to promote interoperability and openness, including data inventory requirements. VA will ensure full public access to relevant metadata without charge, availability in a data format that ensures interoperability with current and future search technology, and links to full text after any embargo period through its partnership with NIH for using PMC.

c. Requirement for Data Management Plans. All VA investigators conducting VA-funded research will be required to develop prospective data management plans for that research. (Effective date: December 31, 2015) *Note: If located in VHA Central Office, investigators should consult the Privacy Officer (PO), Records Officer (RO), and Information Security Officer (ISO) assigned to their Program Office for assistance, as needed, in developing acceptable data management plans. If located at a VHA research facility, investigators should consult their facility PO, RO, and ISO for assistance as needed.*

(1) VA will require that all VA-funded research include a written data management plan to be evaluated for merit as part of the proposal review and approval process. The data management plan for VA-funded research must describe how and where data resulting from the research will be made available to the public and must specifically include how any data underlying scientific publications will be made available for discovery, retrieval, and analysis, including which materials will be available in machine readable formats. Investigators will be held accountable for sharing publications and data in accordance with the approved data management plan. Failure to implement the approved plan may result in loss of current or future funding or other restrictions on the investigator's research activities.

(2) Many VA research studies are supported by funding from non-VA entities such as NIH and industry sponsors. Regardless of funding source, all VA investigators will be required to include in their research proposals a data management plan that describes how, where, and the extent to which they will make the data and results of their research available to the public, including which materials will be available in machine readable formats.

d. Requirements for Public Access to Publications Generated by VA-Funded Research.

Requirements for public access to scientific publications will apply to all publications reporting the results of VA-funded research. Such publications must be stored for long-term preservation and must remain publicly accessible for search, retrieval, and analysis. Proposals for the conduct of such research must include a data management plan that addresses these requirements. (Effective date: December 31, 2015)

e. Requirements for Public Access to Digital Data Generated by VA-Funded Research. All proposals for VA-funded research must include a data management plan describing the mechanisms for providing public access to the digital data resulting from the research. The plan must specifically include how the *final research datasets* underlying all publications reporting results of VA-funded research will be made available for discovery, retrieval, and analysis, including which materials will be available in machine readable formats. (Effective date: December 31, 2015)

(1) VA will begin sharing digital data from VA-funded research through **controlled public access mechanisms** (e.g., through data use agreements (DUAs) and other written agreements, as appropriate) and move as expeditiously as possible toward open public access mechanisms that ensure the protection of Veterans' identifiable private information in a manner consistent with VA's ethical and statutory responsibilities to the Nation's Veterans.

(2) All VA-funded researchers will be required to share **all digital data** underlying the published results from all VA funded research at least under **controlled public access mechanisms** where privacy, intellectual property, or other concerns preclude open public access. (Effective date: December 31, 2015)

(3) VA will implement requirements for digital research data sharing in phases, beginning with studies funded by the ORD Cooperative Studies Program (CSP), followed by other research funded through ORD, and finally by research funded by VHA program Offices. Based on the standards developed by NLM PMC and, where applicable, *ClinicalTrials.gov*, VA's research data sharing mechanisms will identify and ensure accurate attribution relative to all digital datasets made available under this plan. In accordance with the 2013 OSTP memorandum, VA will require that access to digital data from VA-funded research be without charge. (Completion date: December 31, 2015)

f. Limits of Applicability. In keeping with broader VA-wide open data policies being developed by VA OPP in response to OMB Memorandum M-13-13 (which requires that all existing and newly developed datasets be added to VA's enterprise inventory), final research data resulting from VA-funded research will be made publicly available under mechanisms that ensure that (i) the release of such data is compliant with all Federal statutes and regulations, and (ii) the privacy of individual Veterans participating in research is assured (see Section 5 of this plan for additional discussion of related privacy, confidentiality, and data security concerns).¹ These mechanisms will ensure compliance with OMB Memorandum M-13-13 §4, which requires "strengthened measures to ensure that privacy and confidentiality are fully protected and that data are properly secured", including "reviewing information for valid restrictions to release."

(1) Informed consent agreements, HIPAA authorizations, and other documents (such as DUAs, contracts, and memoranda of understanding) associated with currently existing research data have typically included restrictions on use of the data outside VA and/or beyond the study for which the data were collected. For this reason, VA's public access requirement relative to research data will not be applied retrospectively to previously initiated research (see Section 5

¹ From the *VA Open Data Policy Narrative, Paragraph 3, Data Publication Process (draft October 30, 2013)*, "Any dataset that contains information protected by applicable confidentiality law and regulation will not be released as open data without effective de-identification. Characteristics that would lead to a determination to not release include a risk of re-identification of data that contains record level data that has been de-identified. This would include medical records that have been de-identified pursuant to the HIPAA safe harbor, but which in combination with publicly available data and modern data mining methods poses a risk of loss of confidentiality."

of this plan for additional discussion of related privacy, confidentiality, and data security concerns).

(2) VA's requirements for public access to scientific publications and digital research data apply only to research that is funded (i) by ORD from VA's research appropriation, or (ii) by VHA Program Offices (see Section 4.a. above).

Nevertheless, all Federally-funded research conducted by VA must comply with the publication and data access requirements of the Federal funding agency, which in turn must comply with all applicable statutes, regulations, and government-wide OSTP and OMB requirements. In the case of research that is funded by industry and/or private foundations, VA must ensure that the data sharing plan for the study is consistent with VA's ethical and statutory responsibilities to the Nation's Veterans.

(3) As a matter of policy, VA's requirements for public access to digital research data will not apply to release of certain individually identifiable data generated by studies that are protected by Federal Confidentiality Certificates. However, VA will make every effort to provide limited, restricted access to such data under written agreements protecting subjects' privacy and the confidentiality of their data. *Note: Such limitations are consistent with the requirements of OMB Memorandum M-13-13 §4 referenced above. Certificates of Confidentiality protect identifying information about research subjects from compelled disclosure relative to any civil, criminal, administrative, legislative, or other proceeding, whether at the federal, state, or local level.*

(4) Internal operational activities that are not designed to develop or contribute to generalizable knowledge do not constitute research and are not covered under this plan (see VHA Directive 1200, VHA R&D Program and VHA Handbook 1058.05, VHA Operations Activities That May Constitute Research). Datasets generated by such activity will be addressed under the broader VA OPP initiatives to comply with Executive Order 13642 on open and machine readable information and OMB Memorandum M-13-13 on establishing a framework for information management at each stage of the information's life cycle to promote interoperability and openness, including data inventory requirements.

5. Privacy, Confidentiality, and Data Security

a. Limited Utility of De-Identified Data. Wholly de-identified datasets are often not sufficient to meet the needs of investigators who request VA research data. As a result, in order to share data requested by secondary users in a manner consistent with their needs, VA investigators must typically develop (i) detailed documentation to enable data stewards to meet all legislative and regulatory governance requirements for release (e.g., privacy, security, data quality requirements); (ii) prepare the dataset according to required safeguards for sharing; and (iii) execute specific DUAs or memoranda of understanding to ensure that the requestor/data custodian understands and can fulfill VA's requirements regarding data ownership and secure data transfer and storage.

These activities consume considerable time and effort beyond that required for conventional

data documentation. Consequently, VA will implement digital research data sharing requirements within a staged approach that begins with controlled public access to digital research data and moves towards open public access where the protection of Veterans' Personally Identifiable Information, Individually Identifiable Health Information, and Protected Health Information (PII/IIHI/PHI) can be ensured.

Prudence dictates that access to VA research datasets containing PII/IIHI/PHI should occur only when allowed by Federal law or regulation and within the stipulations of written agreements that define the permitted uses of the data, prohibit actions to use or combine de-identified data to "re-identify" patient-subjects, specify information security requirements, clarify data ownership, and provide for the disposition of the data upon project completion.

As indicated in Section 4.e. above, this approach will ensure compliance with OMB Memorandum M-13-13 §4, which requires "strengthened measures to ensure that privacy and confidentiality are fully protected and that data are properly secured", including "reviewing information for valid restrictions to release."

b. Risks Associated with Research Involving PII/IIHI/PHI. VA's research mission is to conduct biomedical and behavioral studies that benefit the health of the Nation's Veterans. The vast majority of VA's research datasets contain the PII/IIHI/PHI of VA patients and other research subjects. VA research relies heavily not only on health information collected specifically for research purposes, but also on Veterans' medical records containing health information originally obtained for clinical purposes.

From the earliest statements on medical ethics, confidentiality has been considered an obligation owed by the health care provider to the patient. The basis for the obligation is two-fold: respect for persons, and the establishment of trust between patients and health care providers as the basis for delivery of safe, high-quality health care. As a civilian health care system, VHA is committed to upholding this standard of professional ethics.

Based on this obligation, Veterans rightly expect that their medical record information will not be released outside VA unless authorized by law, and consistent with professional ethics standards. Veterans trust VA to honor its commitment to preserve the confidentiality of their information (PII/IIHI/PHI). Any real or apparent breach of Veterans' trust in VA would justifiably be met with outrage, would result in irreparable damage to VA, and would severely compromise VA's ability to care for the Nation's Veterans, VA's primary responsibility under law [38 USC 7301].

VA takes extremely seriously its ethical and legal responsibility (i) to safeguard the privacy of Veterans who receive their medical care from VA, as well as the privacy of all persons who freely choose to participate in VA research, and (ii) to ensure the confidentiality of their PII/IIHI/PHI in accordance with requirements related to individual privacy, informed consent, authorization for use and disclosure of PHI, Institutional Review Board (IRB) oversight, and information security. [See applicable Federal privacy laws and regulations, such as the Privacy Act of 1974 at 5 USC 552a, and 38 USC 5701 & 7332; the Federal Policy (Common Rule) for the

Protection of Human Subjects at 38 CFR 16; FDA regulations at 21 CFR 50 & 56; and the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule at 45 CFR 160 & 164].

Research examining the mosaic effect has demonstrated that the wide availability on the internet of large-scale databases and increasingly sophisticated internet search capabilities have made possible the “re-identification” of individual patients/subjects from health-related datasets that were “de-identified” in accordance with the HIPAA Privacy Rule Safe Harbor standards.

Given this potential, accessibility of Veterans’ information for prolonged periods of time in public databases and advances in information technology capabilities will likely put Veterans’ identities at some level of risk. Before VA can grant **unrestricted**, open public access to its large-scale health-related research datasets, VA must ensure that the privacy of Veterans and the confidentiality of their data are protected adequately. VA cannot release data that can be identified in combination with public datasets and the knowledge that the information came from a Veteran. However, **controlled** public access will be provided to the greatest extent possible (see Sections 5.g and 5.h and Figure 1 of this plan).

VA will need to enlist Federal and non-Federal experts who can assist in the assessment of privacy risks associated with matching of “de-identified” health information with both publicly available and readily available private databases. Assessment of these risks and possible solutions **for future expansion of unrestricted, open public** access to VA’s health-related research datasets will be completed by December 31, 2016. (Effective dates and timelines are provided in Section 8 and Figure 2 of this plan.)

Until such assessments can be completed, VA will not be able to provide **unrestricted**, open public access to large-scale health-related datasets because of re-identification concerns and the obligation to protect Veterans’ private information. However, as indicated above, **controlled** public access will be provided to the greatest extent possible under specific DUAs or other appropriate written agreements, and open access will be provided to the final datasets underlying peer-reviewed publications (e.g., aggregated data that can be released without privacy and confidentiality risks). See Sections 5.g and 5.h and Figure 1 of this plan.

c. Risks Associated with Genomic Information of Subjects’ Relatives. A related concern is that although informed consent and HIPAA authorization for research use and disclosure of genomic information can be requested during the informed consent process from the source subject, it is virtually impossible to obtain comparable informed consent and HIPAA authorization from all of the source subject’s 1st degree relatives (i.e., parents, siblings, and children), who share the source subject’s genomic information. The ramifications of the potential violation of privacy of the source subject’s 1st degree relatives and disregard for their right to decline research participation have yet to be explicated satisfactorily or incorporated into broadly accepted ethical and legal standards.

In connection with its genomic research initiatives, VA is consulting widely with experts in

relevant disciplines such as ethics, science, and law; Veterans' advocacy and service organizations; and relevant public stakeholders to identify acceptable mechanisms for access to genetic information. One mechanism that is currently under consideration is permitting some form of public access to genomic and other large VA datasets through the use of protected VA hardware and software that would permit data manipulation and analysis without the actual release of potentially identifiable data outside VA.

Preliminary assessment of the risks associated with genomic and other large VA datasets and possible mechanisms for expanded access will be completed by July 31, 2016. Viable mechanisms will be incorporated into VA's public access requirements and continue to be revised as additional insights become available.

d. Risks Associated with Proprietary Interests and Intellectual Property. The February 22, 2013, OSTP Memorandum specifies that agency plans must avoid "... significant negative impact on intellectual property rights ...". VA recognizes the importance of proprietary interests and confidential business information and will comply with the OSTP Memorandum in this regard. VA's research data access requirements will be designed to enhance innovation and competitiveness and foster public-private collaborations.

On the other hand, VA-funded research may result in intellectual property that may warrant exercising the Department's authority to request patent protection (35 USC 207). Accordingly, mechanisms must be in place to prevent premature disclosure to the public domain of data that might constitute patentable intellectual property prior to application for a provisional patent. VA's partnership with NIH in the use of PubMed Central (PMC) will prevent the unauthorized mass redistribution of scholarly publications.

e. Constraints of Prior Conditions of Informed Consent. Informed consent agreements, HIPAA authorizations, and other documents (such as DUAs, contracts, and memoranda of understanding) associated with currently existing research data have typically included restrictions on use of the data outside VA and/or beyond the study for which the data were collected.

Although constraints such as these could be alleviated going forward by including appropriate language in future informed consent agreements and HIPAA authorizations voluntarily signed by prospective research subjects, achieving the objective to provide maximum feasible electronic public access may not be feasible for VA research data whose use has been limited by the already existing conditions of research subjects' informed consents and authorizations.

For these reasons, and as indicated previously, VA's public access requirements will not include results of research initiated prior to the effective dates provided in Section 8 of this plan.

f. Information Technology (IT) Considerations. VA fully recognizes the value of making digital data available to researchers and the public and will leverage, to the extent possible, current ongoing initiatives in the development of the VA Informatics and Computing Infrastructure

(VINCI), electronic research management systems, and VA's enterprise-level open access data initiatives to foster this objective.

(1) VA's primary missions are to provide health care, benefits, and cemetery services to the Nation's Veterans as mandated by law. VHA research and data access efforts support its health care mission, and the provisioning of VHA research data to the public will necessarily need to be integrated with the various VA OI&T enterprise-wide patient-support data storage and access initiatives, which include VA's enterprise-level programs to implement the requirements of OMB Memorandum M-13-13 .

The development of archiving, data platforms, and sharing resources for VA research data will necessarily unfold within the larger context of the VA OPP Data Governance process and the VA OI&T enterprise framework for information management, which are designed to promote interoperability and openness in accordance with OMB Memorandum M-13-13 and related requirements. This OI&T framework employs a "federated" approach to data storage that leverages existing resources such as the websites of offices that generate and/or maintain data assets that are made publicly available.

(2) Under the "federated" approach, each VA office will be responsible for identifying the data assets they create or manage as data stewards. Metadata describing those assets will be submitted to a centralized VA data inventory that is made available to the public. Per Section 4.b(6), VA will ensure full public access to relevant metadata without charge, availability in machine readable data formats that ensure interoperability with current and future search technology, and links to full text after any embargo period through its partnership with NIH for using PMC. Research publications and the underlying final research data generated by VA-funded research will be made available through NIH PMC and existing public websites, according to the project's approved data management plan.

(3) VHA tasked an Integrated Project Team (IPT) for Regionalized Storage for ORD Data (RSORDD) to perform the analysis and planning necessary to estimate the costs and features of an IT investment into systems that would leverage *existing* nationwide VA IT infrastructure, providing critical integration into current clinical systems and databases to support VA research needs and for the purposes of compliance with the 2013 OSTP memorandum.

The IPT completed its analysis in December 2014, reporting that it discovered approximately 1.5 petabytes of research data currently on-line and estimating that the actual total amount of data (on-line and stored off-line in electronic format such as backup tape, CD/DVD, external hard drives, thumb drives, other storage) to be in the realm of 4-6 petabytes. The IPT also estimated that between 500-700 terabytes of NEW DATA are generated PER YEAR by ORD-funded research projects, and that the rate of new data acquisition is increasing over time.

The IPT concluded that creating the systems needed to support access to all ORD-funded research datasets will require a significant investment in IT infrastructure over a number of years, as well as long-term Federal agency and private sector collaboration to achieve economies of scale. It recommended developing a solution architecture comprised of (a) more

than **30 local research systems** to service the application, database, and storage needs of **active research studies**, and (b) **centralized repository resources** to provide retention and cataloging of all **completed studies**.

(4) The IPT recommended adopting a 4-year plan (FY2015 through FY2018) for development of the local storage systems for active studies, to include phases for (a) data analysis and classification with respect to sensitivity, categorization, required metadata, and retention; (b) engineering, purchasing, and implementation of the local research systems; and (c) preparation of data from completed studies for migration to one or more centralized repositories..

(5) VA will explore the development of a research data commons, a shared space for research output including data, software and narrative associated with biomedical research, both basic and clinical. In collaboration with other Departments and Agencies, such a commons will adhere to the FAIR principles of Find, Access, Interoperate, and Reuse. The goal will be to make digital research data available to the public in a manner that optimizes search, archival, and dissemination features that encourage innovation in accessibility and interoperability, while ensuring long-term stewardship of research data results. A particular focus of the effort will be on making the data underlying the conclusions of peer-reviewed scientific publications resulting from federally funded scientific research available for free at the time of publication.

In the interim, investigators for VA-funded research will be required to complete a data inventory and submit this inventory at the time of application to indicate exactly where their data will be stored. Until the infrastructure recommended by the IPT can be implemented, that requirement will be handled locally and thus be very decentralized.

g. Achieving Fundamental Research Data Access Goals. The large volume of data generated by VA research studies, which enroll over 100,000 individual research subjects in over 15,000 studies annually (not including the Million Veteran Program which enrolls an additional 100,000 or more subjects annually); the obligation to safeguard the privacy of Veterans and the confidentiality and security of their PII/IIHI/PHI; and the mandate for all Federal agencies to develop public access data systems, together require that the standards for public access to the results of VA research be implemented in stages over a period of several years.

Consequently, VA will pursue a staged approach to expanded access to research results that simultaneously advances three fundamental goals:

- Making VA research data available to the public with the fewest constraints possible.
- Protecting the privacy of Veterans and meeting VA's obligation to maintain the confidentiality and security of their PII/IIHI/PHI.
- Utilizing data sharing mechanisms that can be implemented at reasonable cost.

VA will develop long-term objectives for digital research data sharing in a series of stages to spread budget impact across multiple years. Sharing mechanisms for studies supported by appropriated research funds (administered by VHA ORD) will rely on those appropriated funds. Sharing mechanisms applicable to the broader VA open data initiative will rely upon VA OI&T

funds.

Requests for VA research funding will be required to include appropriate costs for publication and data access and management. Requests by VA personnel for funding from other Federal agencies will be required to include costs appropriate to the publication and data access and management requirements of the funding agency.

h. Open Versus Controlled Public Access. VA's expanded research data access mechanisms will ultimately include both *open public access* (i.e., unrestricted access without prior permission or approval) and *controlled public access* (i.e., access under restricted conditions) to digital research data as warranted to protect the privacy of VA's research subjects and the confidentiality of their PII/IIHI/PHI.

In order to achieve fully the objective to provide maximum feasible *open digital public access* to VA research data, VA must develop clear standards to ensure privacy, confidentiality, and identity protections consistent with industry and scientific research community standards and best practices. However, until such protections can be developed and implemented, VA will encourage, and publicize the availability of, the following six *controlled public access* mechanisms (see Figure 1) under which access to *final research results data* can currently be granted. VA is committed to the responsible use of these six mechanisms in implementing its plan for expanded public access to the results of VA- funded research.

(1) *Individually Identifiable Data/Information* may be shared pursuant to valid *HIPAA authorization*. Research informed consent documents, applicable written agreements, protocol documents, etc. must be consistent with such authorization.

(2) In the absence of valid HIPAA authorization,

(a) *Individually Identifiable Data/Information, excluding Veterans' names and 38 USC 7332-protected information*, may be shared, pursuant to a written request and an IRB-approved waiver of HIPAA authorization, with the approval of the Under Secretary for Health, in accordance with VHA Handbook 1605.1 §13.b(1)(b) or §13.b(1)(c) or superseding versions of that Handbook (see 38 USC 7332).

(b) *Individually Identifiable Data/Information, including 38 USC 7332-protected information*, may be shared, pursuant to a written request, IRB approved waiver of HIPAA authorization, approval of the Under Secretary for Health, and a written assurance from the recipient that the information will be maintained in accordance with the security requirements of 38 CFR 1.466, or more stringent requirements, the information will not be re-disclosed except back to VA, and the information will not identify any individual patient in any report of the research or otherwise disclose patient identities.

(c) *A Limited Dataset (LDS)* may be created and shared pursuant to an appropriately formulated DUA prohibiting the recipient from identifying or re-identifying (or taking steps to identify or re-identify) any individual whose data are included in the dataset.

(d) ***A De-identified and/or Apparently Anonymized Dataset*** may be created and shared pursuant to a written assurance from the recipient prohibiting the recipient from identifying or re-identifying (or taking steps to identify or re-identify) any individual whose data are included in the dataset.

(3) All non-exempt, reasonably segregable information must be disclosed pursuant to a valid ***Freedom of Information Act (FOIA) request***. All relevant data, including information that may ultimately be withheld, must be provided to the FOIA officer processing such a request for proper disclosure as determined by the FOIA Officer.

6. Responsible Offices

a. Policy. The VHA Office of Research and Development (ORD) will be responsible for establishing and maintaining policies for public access to scientific publications and digital research data for VA-funded research, including standards for evaluating the merits of the proposed data management plan for VA-funded research and policies for the use of existing resources (e.g., PubMed) for data and publication inventory and archiving.

(1) Procedures for evaluating the merits of proposed data management plans will be incorporated into the established peer review and administrative review processes. VA will require that periodic progress reports on such research include documentation that approved data management plans have been implemented. ORD will be responsible for implementing the requirement to provide data management plans and progress reports for ORD-funded research. VHA Program Offices will be responsible for implementing the requirement to provide data management plans and progress reports for research that they fund.

(2) ORD, the VHA Office of Information and Analytics (OIA), and VA OI&T will be jointly responsible for assessing long-term needs for the preservation of scientific data in fields that VA supports, including consideration of VA data repository resources and the efforts of other public and private sector entities.

(3) ORD (in consultation with the VHA Privacy Office, VA OPP, VA OI&T, the Office of General Counsel, and other relevant offices) will ensure that VA's public access policies related to VA-funded research are consistent with all relevant Federal requirements, including requirements related to privacy, confidentiality, and data security.

(4) ORD, OIA, and VA OI&T will be jointly responsible for coordinating the requirements for public access to scientific publications and digital research data from VA-funded research with the rules developed by the VA Data Governance Council for release of VA data at the record level.

(5) ORD is responsible for notifying VA-funded researchers of the requirements for public access to scientific publications and digital research data by revising relevant policy Handbooks, operational procedures, and training programs to implement these requirements.

(6) ORD, OIA, and VA OI&T will be jointly responsible for providing training, education, and workforce development in data management, analysis, storage, preservation, and stewardship (in coordination with other agencies and the private sector) as they relate to VA's policy for public access to the results of VA-funded research.

(7) ORD, OIA, and VA OI&T will be jointly responsible for developing and maintaining comprehensive inventories of VA research datasets within the broader VA OPP initiatives to comply with Executive Order 13642.

(8) ORD, OIA, and VA OI&T will be jointly responsible for designing and implementing mechanisms to publish descriptions of VA-funded research, and all publications resulting from such research, in "plain language" that improves access for Veterans and the lay public.

(9) ORD, OIA, and VA OI&T will be jointly responsible for designing and implementing an ongoing evaluation process to assess the efficiency and effectiveness of VA's research data access requirements.

(10) ORD, OIA, and VA OI&T will be jointly responsible for developing mechanisms for the assessment of long-term needs for the preservation of scientific data in fields that VA supports, including options for developing and sustaining repositories for scientific data in digital formats that take into account the efforts of public and private sector entities. ORD, OIA, and VA OI&T will ensure that these efforts are fully integrated with VA OPP's broader initiatives for enterprise-wide data storage and access.

b. Compliance Oversight and Enforcement. The VHA Office of Research Oversight (ORO) will be responsible for conducting compliance oversight and enforcement of VA's policies for public access to scientific publications and digital research data resulting from VA-funded research.

(1) Compliance oversight and enforcement of requirements for data access plans, access to publications from VA-funded research, and access to digital data from VA-funded research will be exercised under ORO's independent statutory authority at 38 USC 7303 to monitor, review, and investigate matters of regulatory compliance in VA research, halt VA research where warranted, and require corrective actions to ensure compliance with all applicable requirements related to the conduct of VA research.

(2) ORO will be responsible for incorporating VA's research public access requirements into its existing compliance oversight inspection and enforcement programs. Every VA research facility receives proactive onsite compliance reviews at 3-4 year intervals, for-cause onsite compliance reviews as needed, remote compliance reviews of reported noncompliance, mandatory audits by facility-based research compliance officers, annual facility director certifications of research oversight, and annual collection and monitoring of research compliance performance metrics.

(3) As soon as they become effective (see Section 8 and Figure 2 of this plan), the requirements for data management plans (regardless of funding source) and access to

publications and digital data from VA-funded research will be incorporated into ORO's proactive onsite compliance reviews. These intensive onsite inspections by ORO staff are scheduled at regular 3-4 year intervals at every VA research facility. ORO's annual Facility Director Certification of Research Oversight process, as well as ORO's remote compliance reviews, compliance oversight tools, and technical assistance materials, will also be modified to ensure compliance with these requirements by VA investigators. In addition, facility Research Compliance Officer (RCO) annual and triennial audit requirements (whose findings must be reported to ORO) will be modified to include annual audits to confirm existence of the required data management plans and triennial audits to ensure investigator compliance with those plans.

(4) ORO monitoring and compliance oversight programs will ensure that all identified noncompliance is documented, appropriate remedial action plans are developed, and remedial actions are implemented in a timely fashion to ensure compliance with VA data management and data access requirements. ORO will promptly notify ORD of any failure to comply with the approved data management plan in any project funded by ORD. ORO will promptly notify the Principal Deputy Under Secretary for Health (PDUSH), who serves as the Institutional Official for the VHA Program Office Human Research Protection Program (HRPP), of any failure to comply with the approved data management plan in any project funded by any VHA Program Office. In addition to timely remediation of any identified noncompliance, noncompliance with VA's data management and data access requirements may result in loss of the investigator's current and/or future research funding.

(5) ORO will develop, annually compile, and track performance metrics for compliance with VA's data management and data access requirements. ORO will monitor these metrics at the investigator, facility, network, and systems levels to identify trends and, where warranted, the need for interventions to ensure compliance. Findings will be made available to the public annually.

7. Definitions

Note: Where applicable, the definitions in this Section are meant to conform substantively to the corresponding definitions from the relevant OMB circulars, OSTP memoranda, and VA/VHA Directives and Handbooks, as currently in effect and as may subsequently be revised. In no case shall the definitions in this Section limit the applicability of the corresponding OMB or OSTP definitions.

a. Certificates of Confidentiality. Certificates of Confidentiality are issued by HHS to protect identifiable research information from forced disclosure pursuant to civil, criminal, administrative, legislative, or other such proceedings.

b. Controlled-Access Data. Data that can be obtained only if a prospective user has met defined conditions and obtained the approval of the data owner or designated data manager.

c. Cooperative Research and Development Agreement (CRADA). An agreement between VA

and one or more non-Federal parties under which VA may accept, retain, and use funds, personnel, services, facilities, equipment, or other resources from collaborating parties in order to conduct research and development in a particular project. CRADAs establish the terms of sponsored collaborative research and are designed to protect the parties' prior inventions while allowing the government and its private sector partner(s) to negotiate management of any new discovery or intellectual property that may result from the collaboration. VA authority to enter CRADAs derives from the Federal Technology Transfer Act of 1986, 15 USC 3710a and 38 USC 7303.

d. Data. For purposes of this policy, data refers to all information obtained in research.

e. Data Asset. A collection of data from one or more sources that describes a topic of interest. A data asset may be decomposed into multiple datasets.

f. Data Archive. A place where machine-readable data are acquired, manipulated, documented, and finally distributed to the scientific community for further analysis.

g. Data Enclave. A controlled, secure environment in which eligible researchers can perform analyses using restricted data sources.

h. Data Use Agreement (DUA). A written agreement that (i) governs the sharing of data between a data owner and a data requester (recipient), (ii) defines ownership as related to the data exchange, (iii) establishes the specific terms of use and disclosure for the requester, (iv) provides a means to transfer liability for the protection of data to the requester; (v) serves as a means to establish criteria for using, disclosing, storing, processing, and disposing of data; and (vi) satisfies HIPAA requirements when providing information within a Limited Dataset.

i. Dataset. An organized collection of data. The most basic representation of a dataset consists of data elements presented in tabular form.

j. Data Sharing. Providing data to individuals not involved in the research under which it was acquired. Data can be shared through publication of research results, making data available on open-access or controlled-access web sites, providing copies of data upon request, through the use of data archives and data enclaves, or combinations of these mechanisms.

k. De-Identified Data/Information. Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual, as provided by and in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule at 45 CFR Part 164.514; VHA Handbook 1605.01; and the Federal Policy (Common Rule) for the Protection of Human Subjects at Title 38 Code of Federal Regulations Part 16. De-identified information is no longer covered by the Privacy Act, 5 USC §552a, 38 USC §7332, or the HIPAA Privacy Rule.

Note: Removal of the specific elements listed in the safe harbor provision at 45 CFR 164.514(b) of the HIPAA Privacy Rule does not necessarily ensure continued de-identification or anonymity of

record subjects. See Section 5 of this plan for details. Under appropriate circumstances, de-identified research data and aggregate data may be shared with non-VA investigators who are not involved in a VA- approved research study upon written request from those investigators and documentation confirming the de-identified status of the data. It is strongly recommended that all such requests and related agreements be reviewed by the facility Privacy Officer and maintained by the facility Research Service together with any additional documentation that may be relevant.

l. Final Research Data. The recorded factual material accepted by the scientific community as necessary to document, support, and validate research findings, including the data on which published summary statistics and tables are based. Final research data refers to data analyzed and formatted to serve as the basis for results reported in peer-reviewed publications. As circumstances warrant, final research data may be shared as an identifiable dataset or may be converted to a de-identified dataset or a limited dataset for sharing purposes.

m. Health Information. Any information, including genetic information, whether oral or recorded in any form or medium, created or received by a health care provider or health plan that relates to the past, present, or future physical or mental health or condition of an individual or the provision of health care to an individual; or the payment for the provision of health care to an individual. Health Information includes information pertaining to examination, medical history, diagnosis, findings, or treatment, and includes such information as laboratory examinations, x-rays, microscopic slides, photographs, prescriptions, and other similar data.

n. Individually Identifiable Health Information (IIHI). A subset of health information, including demographic information collected from an individual, that is: (1) created or received by a health care provider, health plan, or health care clearinghouse; (2) relates to the past, present, or future condition of an individual and provision of or payment for health care; and (3) identifies the individual or for which a reasonable basis exists to believe the information can be used to identify the individual. *Note: VHA uses the term individually-identifiable health information to define information covered by the Privacy Act and the Title 38 confidentiality statutes in addition to HIPAA. Individually-identifiable health information does not have to be retrieved by name or other unique identifier to be covered by this term.*

o. Individually-Identifiable Data/Information. Individually-identifiable data/information is any data/information, including health information maintained by VHA, pertaining to an individual that also identifies the individual and, except for individually-identifiable health information, is retrieved by the individual's name or other unique identifier. Individually-identifiable health information is covered regardless of whether or not the information is retrieved by name.

p. Limited Dataset. Protected health information from which certain specified direct identifiers of the individuals and their relatives, household members, and employers have been removed. These identifiers include name, address (other than town or city, state, or zip code), phone number, fax number, e-mail address, Social Security Number (SSN), medical record

number, health plan number, account number, certificate and/or license numbers, vehicle identification, device identifiers, web universal resource locators (URL), internet protocol (IP) address numbers, biometric identifiers, and full-face photographic images. A limited dataset is not de-identified information or data. A limited dataset may be used for research, health care operations, and public health purposes. VHA may disclose a limited dataset for research, health care operations, and public health purposes pursuant to a DUA. *Note: Limited datasets may typically be shared with non-VA investigators who are not involved in a VA-approved research study upon written request from those investigators and documentation from the VA facility Privacy Officer confirming that status of the research data. Additionally, a DUA required by the HIPAA Privacy Rule must also be completed by the VA and non-VA parties per VHA Handbook 1605.1 Appendix F and VHA Handbook 1080.01, Data Use Agreements. It is strongly recommended that all such agreements be reviewed by the facility Privacy Officer and maintained by the facility Research Service together with any additional documentation that may be relevant.*

q. Machine Readable. Machine readable refers to data represented in a format intended for automated consumption by software. Examples of such formats include, but are not limited to, JSON, XML, and CSV.

r. Metadata. Metadata refers to characteristics of data that facilitate common understanding. The metadata for scientific data generated by VA-funded research will include, at a minimum, the common core metadata schema in use by the Federal government, found at <https://project-open-data.cio.gov/>. *Note: Per Section 4.b(6) above, VA will ensure full public access to relevant metadata without charge, availability in a data format that ensures interoperability with current and future search technology, and links to full text after any embargo period through its partnership with NIH for using PMC.*

s. Mosaic Effect. The mosaic effect occurs when information from a dataset is combined with other available information in a way that could result in the release of sensitive information (e.g., enabling the identification of an individual, posing a threat to national security, etc.).

t. Non-Public Dataset. A non-public dataset refers to a data asset that is not available to the public. A data asset marked as non-public may still potentially be available to other intra-agency operating units and/or other government agencies.

u. Open-Access Data. Data that can be browsed or downloaded from a web site without prior permission or approval. Open Data refers to publicly available data structured in a way that enables the data to be fully discoverable and usable by end users.

v. Open Standard. A standard developed or adopted by voluntary consensus standards bodies, both domestic and international. These standards include provisions requiring that owners of relevant intellectual property have agreed to make that intellectual property available on a non-discriminatory, royalty-free or reasonable royalty basis to all interested parties.

w. Primary Outcomes Publication. A peer-reviewed publication that reports results addressing the primary aim(s) specified in the protocol and/or funding application.

x. Program Office. Any office within the VHA Office of the Under Secretary for Health. A Program Office includes all of its component offices and subdivisions, regardless of physical location (i.e., including Program Office components hosted by VA Medical Centers). *Note: Components of the VHA Office of the Under Secretary for Health may be found on the VHA Web site at <http://www.va.gov/health/orgs.asp>.*

y. Personally Identifiable Data/Information (PII). Any data/information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc. *Note: Information does not have to be retrieved by any specific individual or unique identifier (i.e., covered by the Privacy Act) to be personally identifiable information. The term "Personally Identifiable Information" is synonymous and interchangeable with "Sensitive Personal Information."*

z. Protected Health Information (PHI). Individually-identifiable health information subject to HIPAA requirements and transmitted or maintained in any form or medium by a covered entity, such as VHA. *Note: VHA uses the term protected health information to define information that is covered by HIPAA but, unlike individually-identifiable health information, may or may not be covered by the Privacy Act or Title 38 confidentiality statutes. In addition, PHI excludes employment records held by VHA in its role as an employer.*

aa. Public Dataset. Public refers to a data asset that *is currently, or could possibly be*, made publicly available. This is the default for all newly collected or created data assets maintained by the Federal government. A data asset whose only barrier to access is a written or electronic request from any person, with no criteria other than that a request be made, should be marked public, not restricted, because those data *could* be made publicly available by removing the intermediate step of the request.

bb. Public Access. Availability of data by the general public through open-access databases or defined controlled-access procedures that do not require a Freedom of Information Act (FOIA) request.

cc. Research. Research is the testing of concepts by the scientific method of formulating a hypothesis or research question, systematically collecting and recording relevant data, and interpreting the results in terms of the hypothesis or question. Research is a systematic investigation including research development, testing and evaluation designed to develop or contribute to generalized knowledge. *Note: Internal operational activities that are not designed to develop or contribute to generalizable knowledge do not constitute research under this plan. See VHA Directive 1200, VHA R&D Program and VHA Handbook 1058.05, VHA Operations Activities That May Constitute Research.*

dd. Restricted Data. Datasets that cannot be distributed to the general public because of participant confidentiality concerns, third-party licensing or usage agreements, etc.

ee. Restricted Public Dataset. A restricted public dataset refers to a data asset that is available under certain use restrictions to members of the public, determined by an agency on a per data asset basis. For example, a data asset containing PII/IIHI/PHI might be made available to select researchers under a contract or DUA.

ff. VA-Funded Research. VA-funded research means (i) research funded by VHA ORD using Congressionally-appropriated research funds, and (ii) research funded by VHA Program Offices.

ff. VA Research. Research that is approved by a VA facility's Research & Development Committee and conducted by VA Investigators including PIs, co-PIs, and site investigators on VA time (serving on compensated, WOC, or IPA appointments), utilizing VA resources (e.g., equipment), or on VA property including space leased to, and used by VA. The research may be funded by VA, by other sponsors, or be unfunded.

8. Effective Dates and Timelines for Implementation

a. Preliminary Actions. No later than October 1, 2015, VA will:

(1) Initiate outreach to the VA research community, the VA National Research Advisory Council (NRAC), Veterans' Service Organizations, and other stakeholders, including a *Federal Register* announcement, seeking public comment on this public access plan.

(2) Finalize procedures to implement the agreement with NLM PMC to host publications resulting from VA-funded research studies.

(3) Finalize procedures for implementing public access to the results of clinical trials funded by the VA Cooperative Studies Program (CSP), including the process for entering information into the ClinicalTrials.gov archive. *Note: In general, Section 801 of the FDA Amendments Act (FDAAA 801) requires that results of trials involving products that are approved, licensed, or cleared by FDA must be submitted to ClinicalTrials.gov no later than 12 months after completion. Where the trial is completed before the product is initially approved, licensed, or cleared by FDA, results must be submitted within 30 days after the product is approved.*

(4) Submit, for formal VA approval, all revisions to existing VHA policy Handbooks needed to implement the requirements of this policy and implementation plan.

b. Full Implementation of Public Access Requirements. No later than December 31, 2015, VA will:

(1) Require that all VA investigators include in their research proposals, a data management plan that describes how, where, and the extent to which they will make the data and results of

their research available to the public and specifying what data will be available in machine readable formats.

(2) Require that all publications based on VA-funded research be made available to the public through the NLM PMC open public access web site within one year of the date of publication.

(3) Through its partnership with NLM PMC, require that publications and related metadata resulting from VA-funded research be stored in archival solutions that provide for long-term preservation and access to the content without charge.

(4) Require public access to the results of clinical trials funded by the VA Cooperative Studies Program (CSP) through submission of such results to the *ClinicalTrials.gov* archive.

(5) Require that data management plans for **all VA-funded research**:

(a) Specifically include how the *final research datasets* underlying scientific publications will be made available for discovery, retrieval, and analysis and what data will be available in machine readable formats.

(b) Include provisions for long-term preservation of, and access in digital machine-readable format to, the scientific data resulting from the research (including the final research datasets underlying any publications), or explaining why long-term preservation and access cannot be provided.

(c) Include an explanation of how data sharing and preservation will enable validation of results, or how results could be validated if data are not to be shared or preserved.

(d) Describe the mechanisms to ensure the personal privacy of research subjects, the confidentiality of individually identifiable private information, and the secure maintenance of proprietary data and information.

(6) Require the inclusion of appropriate costs for data management and access in proposals for ORD-funded research.

(7) Require that an evaluation of the merit of data management plans for VHA-funded research be incorporated into the existing peer review or administrative review processes.

(8) Require that investigators of all VA-funded research be held accountable for sharing publications and data in accordance with the approved data management plan and specify that failure to implement the approved plan may result in loss of current or future funding or other restrictions on the investigator's research activities.

(9) Modify RCO audit requirements and ORO protocols for onsite compliance reviews to identify and document noncompliance with approved data management plans and monitor remedial actions to ensure compliance.

(10) Implement a mechanism for stakeholders to petition to change the embargo period for a specific field by demonstrating that the established period would, by a preponderance of evidence, be inconsistent with the public interest and with the objectives articulated by OMB and OSTP. Such petitions must be presented in writing to the VHA Chief Research and Development Officer (CRADO), who shall render a decision on petitions within 90 days of receipt. Decisions may be appealed in writing to the Under Secretary for Health, who shall render a final agency decision within 90 days of the receipt of the appeal.

c. Planning for Future Expansion of Datasets Available for Unrestricted Public Access. No later than December 31, 2016, VA will:

- (1) Complete a detailed analysis of mosaic effects, whereby the release of ostensibly “de-identified” information from datasets such as those obtained for VA research, can be reasonably expected to be combined with information from large, readily available datasets to identify the individuals from whom the data originated.
- (2) Develop and implement specific standards related to ensuring identity protections for VA research subjects, including mechanisms to prevent “re-identification” of subjects from “de-identified” datasets.
- (3) Develop and implement specific standards to protect patentable intellectual property, where warranted.
- (4) Complete a detailed analysis of, and determine a path forward regarding, the most viable long-term mechanisms for effecting public access to VA research data, including one or more of the following:
 - (a) Creating a dedicated, centralized research data archive hosted by VA (perhaps expanding upon VA’s current VINCI initiative);
 - (b) Relying on a group of decentralized VA research data archives;
 - (c) Making use of components of VA’s emerging enterprise framework for information management;
 - (d) Using an archive hosted by another government entity but available to VA;
 - (e) Permitting some form of public access to genomic and other large VA datasets through the use of protected VA hardware and software that would permit data manipulation and analysis without the actual release outside VA of potentially identifiable information;
 - (f) Combinations and/or other alternatives to the above.
- (5) Develop and implement mechanisms for continued assessment of the long-term needs for

the preservation of scientific data in fields that VA funds, including options for developing and sustaining data warehouse and repository resources for scientific data in digital formats, including consideration of other public and private efforts.

(6) Develop, and make available to the VA research community, standardized template language for voluntary informed consent, voluntary HIPAA authorization, and DUAs to support public access to VA research results under specified protections for maintaining the privacy of subjects and the confidentiality of their data.

9. References

1. Association of American Medical Colleges (2010), *Working Group on Information Technology Security and Privacy in VA and NIH-Sponsored Research*, https://www.aamc.org/download/138118/data/va_report.pdf.pdf.
2. Borgman, C.L. (2011), *The Conundrum of Data Sharing*, *Journal of the American Society for Information Science and Technology*. Available from Social Science Research Network, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155.
3. Data.Gov, Data Policy Statements, <http://www.data.gov/data-policy>.
4. Department of Veteran Affairs, *VA Information and Computing Infrastructure (VINCI)*, http://www.hsrp.research.va.gov/for_researchers/vinci/default.cfm.
5. Executive Office of the President, *Increasing Access to the Results of Federally Funded Scientific Research*, http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
6. Executive Office of the President, *Making Open and Machine Readable the New Default for Government Information*, <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>.
7. Executive Office of the President, Office of Management and Budget, Memorandum M 13-13, May 9, 2013, Open Data Policy-Managing Information as an Asset.
8. Inter-University Consortium for Political and Social Research, *Guide to Social Science Data Preparation and Archiving*, 5th Edition, <http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>.
9. National Aeronautics and Space Administration, Open Government Initiative, <http://www.nasa.gov/open/plan/science-data-access.html>.

10. National Institutes of Health, Guidelines for the Conduct of Research in the Intramural Research Program at NIH, <http://sourcebook.od.nih.gov/ethic-conduct/Conduct%20Research%206-11-07.pdf>.
11. National Institutes of Health, NIH Data Sharing Policy, Publication No. 03-5399, http://grants.nih.gov/grants/policy/data_sharing/data_sharing_brochure.pdf.
12. National Institutes of Health, NIH Data Sharing Policy Implementation and Guidance, http://grants2.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
13. National Institutes of Health, NIH Grants Policy Statement, October 1, 2011, http://grants.nih.gov/grants/policy/nihqps_2011/index.htm.
14. National Institutes of Health, National Center for Biotechnology Information, dbGaP Overview, <http://www.ncbi.nlm.nih.gov/projects/qap/cgi-bin/about.html>.
15. National Institutes of Health, NIH Public Access & PMC, <http://www.ncbi.nlm.nih.gov/pmc/about/public-access-info/>.
16. National Institutes of Health, PubMed Central (PMC) Overview, <http://www.ncbi.nlm.nih.gov/pmc/about/intro>.
17. Office of Management and Budget Memorandum M-13-13, *Open Data Policy – Managing Information as an Asset*, <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.
18. Office of Management and Budget, *Project Open Data*, <http://project-open-data.github.io/>.
19. Title 5 United States Code Section 552a, *The Privacy Act*.
20. Title 15 United States Code Section 3710a, *Cooperative Research and Development Agreements*.
21. Title 35 United States Code Section 207, *Domestic and foreign protection of Federally owned inventions*.
22. Title 38 Code of Federal Regulations Part 16, *Protection of Human Subjects*.
23. Title 38 United States Code Section 5701, *Records and Investigations*.
24. Title 38 United States Code Section 7301, *Functions of Veterans Health Administration: In general*.
25. Title 38 United States Code Section 7303, *Functions of Veterans Health Administration: Research programs*

26. Title 38 United States Code Section 7307, *Office of Research Oversight*.
27. Title 38 United States Code Section 7332, *Confidentiality of Certain Medical Records*.
28. Title 45 Code of Federal Regulations Part 160, *Standards for Privacy of Individually Identifiable Health Information: General Administrative Requirements*.
29. Title 45 Code of Federal Regulations Parts 164, *Standards for Privacy of Individually Identifiable Health Information: Security and Privacy*.
30. VA Information Resources Management, *Strategic Plan*, Available upon request and at http://www.ea.OI&T.va.gov/EAOI&T/docs/Sept2013Docs/VA_IRM_Strategic_Plan_5_15_13_for_Publication.pdf.
31. VA Handbook 6500, *Information Security Program*.
32. VHA Directive 1058, *The Office of Research Oversight*.
28. VHA Handbook 1058.05, *VHA Operations Activities That May Constitute Research*.
29. VHA Handbook 1080.01, *Data Use Agreements*.
30. VHA Directive 1200, *VHA Research and Development Program*.
31. VHA Handbook 1605.01, *Privacy and Release of Information*.
32. VHA Integrated Project Team, December 2014, *Regionalized Storage for Office of Research and Development Data (RSORDD), Project Analysis*.

Figure 1.
Current Mechanisms for Access to the Data Resulting from
Research Funded by the Department of Veterans Affairs

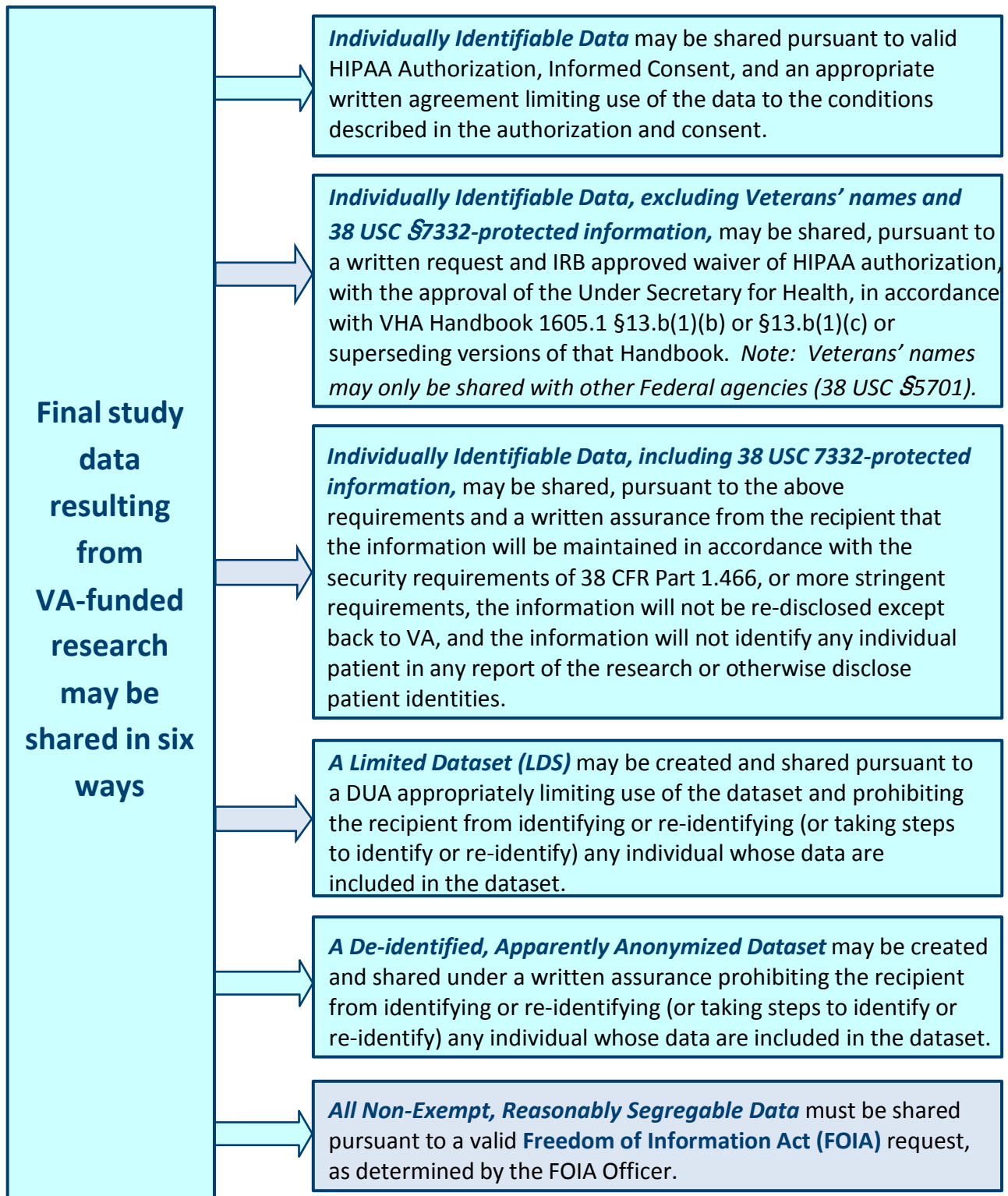


Figure 2.
Timeline for Implementation of
Policy for Public Access to Scientific Publications and Digital Data from
Research Funded by the Department of Veterans Affairs

Preliminary Actions: October 1, 2015
(1) Seek public comment on the VA plan for public access to the results of VA-funded research. (2) Finalize procedures to implement the agreement with NLM PMC to host publications. (3) Finalize procedures for implementing public access to the results of clinical trials funded by the VA Cooperative Studies Program (CSP). (4) Submit, for formal VA approval, all necessary revisions to existing VHA policy Handbooks.
Full Implementation of Public Access Requirements for all VA-Funded Research: December 31, 2015
(1) Require that all VA investigators provide a data management plan that describes how, where, and the extent to which they will make the data and results of their research available to the public and what data will be available in machine readable formats. (2) Require that all publications based on VA-funded research be made available through NLM PMC within one year of the date of publication. (3) Require that publications and related metadata from VA-funded research be stored through NLM PCM in archival solutions providing long-term preservation and access without charge. (4) Require access to the results of CSP-funded clinical trials through the <i>ClinicalTrials.gov</i> archive. (5) Require that data management plans for all VA-funded research: (a) Specifically include how the <i>final research datasets</i> underlying scientific publications will be made available for discovery, retrieval, and analysis and what data will be available in machine readable formats. (b) Include provisions for long-term preservation of, and access in digital machine-readable format to, the scientific data resulting from the research or explaining why such cannot be provided. (c) Include an explanation of how data sharing and preservation will enable validation of results, or how results could be validated if data are not to be shared or preserved. (d) Describe the mechanisms to ensure the personal privacy of research subjects, the confidentiality of individually identifiable private information, and the secure maintenance of proprietary data and information. (6) Allow the inclusion of appropriate costs for data management and access in ORD-funded research. (7) Require that an evaluation of the merit of data management plans for VHA-funded research be incorporated into the existing peer review or administrative review processes. (8) Require that investigators of all VA-funded research be held accountable for sharing publications and data in accordance with the approved data management plan. (9) Modify RCO audit requirements and ORO protocols for onsite compliance reviews to identify and document noncompliance with approved data management plans and monitor remedial actions to ensure compliance. (10) Implement a mechanism for stakeholders to petition to change the embargo period for a specific field.

**Planning for Future Expansion of Datasets Available for Unrestricted Public Access:
December 31, 2016**

- (1) Complete a detailed analysis of mosaic effects permitting identification of the individuals from whom data originated.
- (2) Develop and implement standards to ensure fail-safe identity protections for VA research subjects, including mechanisms to prevent “re-identification” of subjects from “de-identified” datasets.
- (3) Develop and implement standards to protect patentable intellectual property, as warranted.
- (4) Complete a detailed analysis of, and determine a path forward regarding, the most viable long-term mechanisms for effecting public access to VA research data, including one or more of the following:
 - (a) Creating a dedicated, centralized research data archive hosted by VA (perhaps expanding upon VA’s current VINCI initiative);
 - (b) Relying on a group of decentralized VA research data archives;
 - (c) Making use of components of VA’s emerging enterprise framework for information management;
 - (d) Using an archive hosted by another government entity but available to VA;
 - (e) Permitting some form of public access to genomic and other large VA datasets through the use of protected VA hardware and software that would permit data manipulation and analysis without the actual release outside VA of potentially identifiable information;
 - (f) Combinations and/or other alternatives to the above.
- (5) Develop and implement mechanisms for continued assessment of the long-term needs for the preservation of scientific data in fields that VA funds.
- (6) Develop, and make available to the VA research community, standardized template language for voluntary informed consent, voluntary HIPAA authorization, and DUAs to support public access to VA research results.