

---

# DRAFT FREQUENTLY ASKED QUESTIONS:

## RESEARCH DATA INVENTORY AND OIG REPORT 11-01823-294

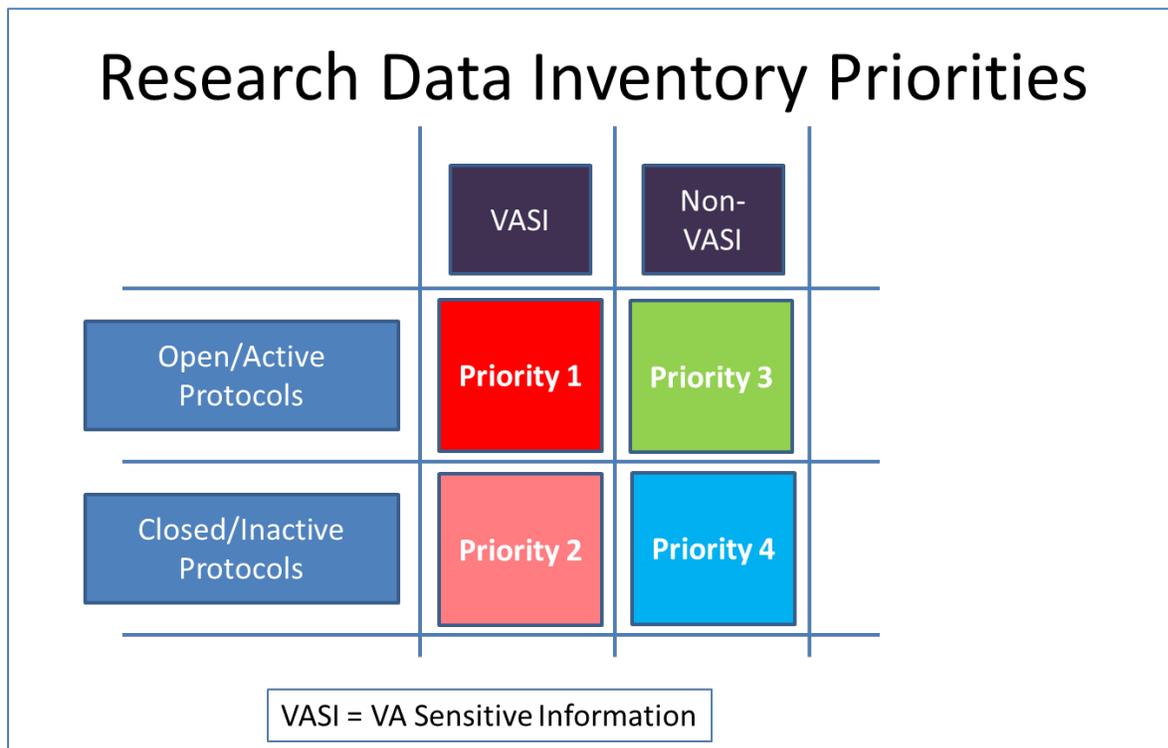
February 12, 2015

**1. WHAT DID THE OIG FIND IN THEIR REPORT?**

They found that VA could not provide an accurate inventory of research data, including where the data were hosted. They also found, that even if the data was inventoried, no sensitivity level of the data (VA Sensitive Information (VASI) or non-VASI label) was associated with the inventory. They also found unsecured and unencrypted electronic (backup tape, CD/DVD, thumb drive, external hard drive, etc.) and hard copy (paper) research data at VAMC and co-located research facilities.

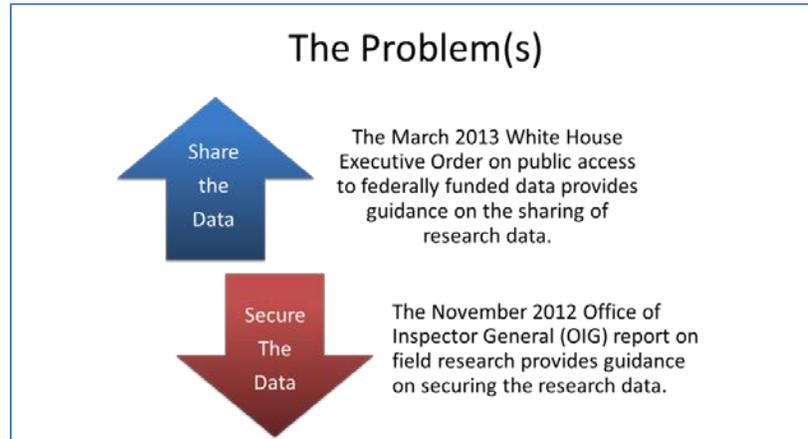
**2. WHAT TYPES OF DATA ARE THEY MOST CONCERNED ABOUT?**

They are concerned about all active and inactive data – insisting that the VA manage all research data over the entire research lifecycle. They are more concerned about VASI than non-VASI. But because there is so much research data in the ORD field, they agreed we could focus on the Research Data Inventory in stages. The highest priority would be given to creating an inventory of Active research that included VASI. They expect other types of VA Research Data to be included over time (management over the entire data lifecycle). We can identify the following 4 quadrants for research data:



**3. BESIDES THE OIG, WHAT OTHER DRIVERS ARE REQUIRING US TO INVENTORY OUR RESEARCH DATA?**

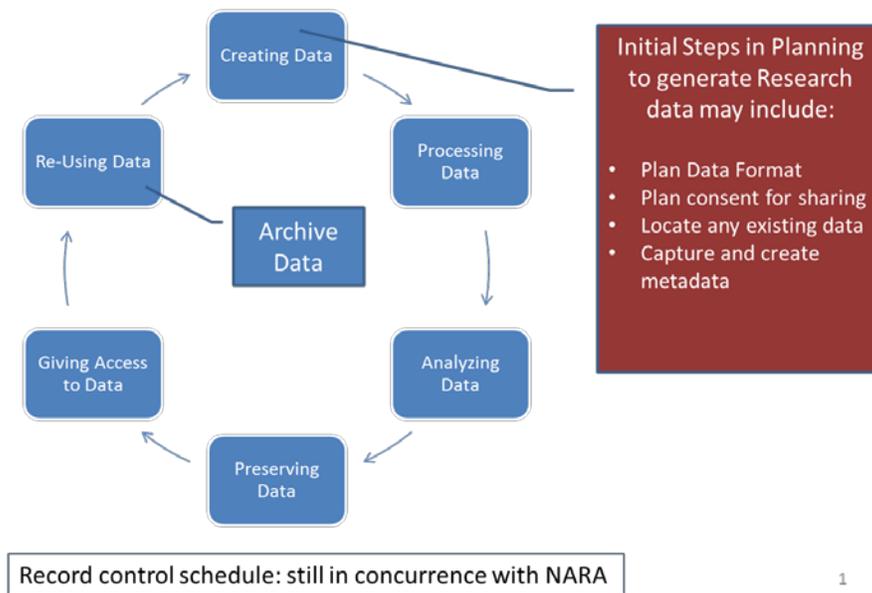
The White House Office of Science and Technology Policy (OSTP) and OMB have mandates for “open data” and to make federally funded research publicly available. The VA/VHA has been developing a draft policy to comply with these mandates (as have all federal agencies). The draft VA policy will require ORD to inventory all research data (VASI/non-VASI, open/closed). In some ways the requirements of OIG and OSTP contradict each other (see graphic), but the end result is that ORD will need better process, policy and IT systems to support both drivers.



**4. WHAT IS THE RESEARCH DATA LIFECYCLE AND HOW MUCH VA RESEARCH DATA IS THERE?**

A project team of ORD and OIT staff have been working to identify the scope of the mandates from OSTP and OIG. The research data lifecycle may be thought of as all steps including those below and the current estimate (probably an underestimate) is shown below:

## Research Data Life Cycle



The project team estimates ~4-6 petabytes of “on-line” data currently exists in VHA ORD and that the rate of new data accumulation is ~500-1,000 terabytes annually. The amount of closed/inactive protocol data that is now “off-line” is unknown.

**5. WHAT IS THE DEFINITION OF VA SENSITIVE INFORMATION?**

The definitions of VASI, Protected Health Information, and Personally Identifiable Information are found in VA Handbook 6500 and Appendix D. VA sensitive information is all VA data on any storage media or in any form or format, which requires protection due to the risk of harm that could result from inadvertent or deliberate disclosure, alteration, or destruction of the information. The term includes information whose improper use or disclosure could adversely affect the ability of an agency to accomplish its mission, proprietary information, records about individuals requiring protection under various confidentiality provisions such as the Privacy Act and the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, and information that can be withheld under the Freedom of Information Act (FOIA). Examples of VA sensitive information may include:

- Individually-identifiable medical, benefits, and personnel information; VHA HANDBOOK 1200.01
- Intellectual property related to research
- Non-human primate research data (a risk assessment should be done to determine sensitivity level)
- Other information which, if released, could result in violation of law or harm or unfairness to any individual or group, or could adversely affect the national interest or the conduct of Federal programs.

**6. WHAT IS THE SIGNIFICANCE OF THE JANUARY 15, 2015 DATE OF THE MEMO?**

VHA and ORD are required to report to the OIG quarterly until their last recommendation in the Report 11-01823-294 is closed. Their inspectors are looking for assurance that the controls on Research Data are working, evidence of ongoing monitoring by the local Research Office and that the VHA can provide documentation of the activities. This means that by March 2015, we will report to the OIG the number of VAMC Research Offices that certify they have complied with the data inventory. For those sites that are not able to comply, they must ask for an extension and that extension will also be tracked and reported to the OIG. The documentation of compliance, monitoring and reporting to the OIG required to close out the recommended action is being developed and will be shared with the ORD Field when available.

**7. IS THE INVENTORY TO BE KEPT BY THE RESEARCH OFFICE?**

Yes. We expect this responsibility of compiling the inventory to be carried out by the ACOS/Research and Research Office. Individual Principle Investigators will have the responsibility for individual protocol data inventories. A Research Data Inventory can be on paper, simple Excel sheet, SharePoint custom list or other tool that meets the intent of the memo. Some centers have electronic tools for this purpose, but many do not. ORD is planning an enterprise solution to this problem but that will take time. Any Research Office that decides to automate their Research Data Inventory should proceed with the understanding that at some future point in time, the Inventories may be consolidated or migrated to an enterprise tool. We will use the interim period to determine business processes, policies and requirements to make the best tool selection we can. Our ORD IT Portfolio will include requests for future IT resources to accomplish this goal.

**8. DOES THE RESEARCH DATA INVENTORY NEED TO CONTAIN THE PRIMARY RESEARCH DATA TOO?**

No. Only the characteristics of the protocol as outlined in the form provided. Storage of research data is the subject of another joint ORD/OIT project team planning an acquisition of local, regional and central storage for all research data to comply with the OSTP/OMB mandates for Open data. The RSORDD team (Regional Storage of ORD Data) is planning this 3-tiered storage system to include tools to crawl, index,

search and retrieve metadata associated with research data and would also permit Research Office staff to manage multiple research folders and permissions for those folders to comply with all mandates.

**9. IF DATA IS COLLECTED AND RESIDES IN VINCI OR SOME OTHER CENTRAL DATA REPOSITORY, IS IT STILL DECLARED IN THE INVENTORY?**

If the Principle Investigator is at your facility and the protocol was approved by your IRB, then your Research Data Inventory should note that VINCI is the location of the data. If the primary data resides in VINCI but you have downloaded a copy or just a de-identified data set, your inventory needs to mention both locations. The OIG is concerned about management over the entire data lifecycle and part of that lifecycle includes the steps that are related to process and analysis of data.

**10. WHAT HAPPENS TO DATA THAT ARE ARCHIVED TO AN APPROVED FEDERAL SITE WHEN CLOSED?**

Since data archive is a step in the data lifecycle, the Research Data Inventory must be updated with the new/final location of the research data. The approved Federal site may be the last notation in the inventory until such time as the Record Control Schedule (RCS) would allow destruction of the data.

**11. IF ALL THE RESEARCH DATA IS KEPT ON A SPECIFIC SERVER, DO WE HAVE TO SPECIFY DOWN TO THE FOLDER LEVEL OR WILL THE SERVER NAME AND PHYSICAL LOCATION SUFFICE?**

A very specific question from a site with relatively advanced practices of Research Data Inventory and lucky enough to have reasonable IT resources available! The Inventory should specify the location of the data with enough detail to allow an OIG auditor to independently verify the accuracy of the inventory. The server name and location (station, building, floor, room, or some other similar locator) should be sufficient, but if you have created research folders and manage those folders, then you should be more specific (citing the root/directory/folder path/folder name if available).

**12. INVESTIGATORS THAT ONLY DO BENCH WORK AND HAVE DATA BOOKS, WHAT ARE THE RESPONSIBILITIES FOR THOSE STUDIES OR DO WE FOCUS ONLY ON HUMAN RESEARCH?**

Data books and other paper records that contain information related to human subject's research are especially noted by the OIG. They fall into the category of paper records and removable electronic media (backup tape, thumb drives, external hard drives, CD/DVD, etc.) that the OIG is especially concerned about. All such paper or removable electronic data (whether encrypted or not) should be locked in secure cabinets designed for data storage and the location (station, building, floor, room, etc.) noted in the Research Data Inventory. We suggest that you determine which quadrant of the matrix (page 1) you are dealing with and proceed accordingly; recognizing the requirement to inventory the data will eventually extend to all 4 quadrants.

**13. WILL A NEW RECORD CONTROL SCHEDULE CHANGE THIS RESEARCH DATA INVENTORY REQUIREMENT?**

Yes. Right now the new RCS for VA Office of Research and Development is in concurrence with the National Archive and Records Administration (NARA). It is being reviewed after a long concurrence process within VA/VHA. In many cases, a new RCS may allow destruction of closed/inactive material that is now being kept in perpetuity. The sheer quantity of paper and electronic storage space required may be less because of a new RCS. However, note that the rate at which new electronic research data is acquired is increasing exponentially, so the size of the problem will continue to grow even with an approved RCS.

**14. SHOULD NON-VA LOCATIONS OF NON-SENSITIVE VA DATA (E.G., BASIC SCIENCE DATA STORED AT AN AFFILIATE LOCATION) ALSO BE RECORDED?**

Consult the quadrant matrix on page 1. We will eventually want to inventory all VA data regardless of where it is stored conducted VA research data. If there is a question regarding data ownership, consult with the local Research Service. We hope to release an example of a Work Breakdown Structure or project plan that outlines the milestones and dates for full compliance.

**15. IS VASI THAT HAS BEEN DISCLOSED TO A NON-VA RECIPIENT (UNDER A VALID HIPAA AUTHORIZATION) BELONGING TO THE VA OR TO THE RECIPIENT? WOULD THIS HAVE TO BE DECLARED IN THE INVENTORY?**

It depends and you should consult with your ISO/Privacy Officer.

**16. ARE PHI AND PII THE ONLY TYPES OF VA SENSITIVE DATA?**

No. Examples of other research data that may be sensitive include non-human primate data, intellectual property related research data (CRADA, patent-related, etc.). Consult your ISO or Privacy Officer if uncertain.

**17. INVESTIGATORS AT OUR FACILITY ALREADY IDENTIFY AND INVENTORY THEIR DATA BUT THAT INFORMATION IS IN THE DOCUMENTATION OF THE PROTOCOL ONLY. WHAT SHOULD WE DO?**

That information in the protocol that relates to the required elements of the form should be extracted from the protocol, elevated to the Research Data Inventory and managed Inventory. The rationale is that the ability to manage the research data over the entire data lifecycle is better accomplished in the Inventory than by having to investigate the forms on file. Oversight and monitoring of compliance with the requirement is also made easier.

**18. WHAT DO YOU ANTICIPATE FOR ORO'S ROLE?**

We are working with ORO and the part it might play in determining compliance with the Research Data Inventory requirement of the OIG. Without a policy, directive or handbook section related to this requirement, ORO is limited in its authority to enforce compliance. ORO is working with ORD on the VA/VHA Public Access to Federally Funded Research (OSTP) policy and it might be that part of the implementation of that policy would be a new ORD Handbook (1200.xx) dealing with Public Access to data. Alternatively, VA Handbook 1200.12 dealing with Research Data Repositories might contain a section related to Research Data Inventories. Input will be sought and any policy changes communicated prior to implementation.

**19. A COUPLE OF YEARS AGO A 9-PAGE PRIVACY AND SECURITY PLAN WAS CIRCULATED. DOES THAT REPLACE THE NEED FOR THE RESEARCH DATA INVENTORY?**

The "Checklist for Reviewing Privacy, Confidentiality and Information Security in Research" was issued by OIT, ORD and the Privacy Office in 2011. That tool was distributed to aid in facilities in meeting the requirements related to the ISO and PO review of human subjects' protocols as outlined in the 2012 version of VHA Handbook 1200.05. This inventory initiative is based on a separate requirement to address the recommendations in the OIG Report 11-01823-294. However, facilities may find that creating a Research Data Inventory is a useful step that could be incorporated into the local processes for the review of protocols for information security and privacy requirements. A more modular approach to all the privacy and security regulations is being considered with the Research Data Inventory as the key step in establishing control over the data. The issue is being discussed in the Security of Research Data Working Group.

**20. WHAT IS THE SECURITY OF RESEARCH DATA WORKING GROUP (SRD-WG)?**

The SRD-WG is a joint VHA/ORD/OIT working group chartered to discuss and recommend issues of policy related to the security of research data. It was chartered by the VHA Chief of Staff, the VHA Chief Research and Development Officer (CRADO) and the VA CIO in January 2014. The SRD-WG meets monthly and is working on many policy issues related to data at rest, data in flight, data sharing, scientific computing and isolation architecture, problems in collaboration with non-VA colleagues, and more. The Chair is Jim Breeling ([james.breeling@va.gov](mailto:james.breeling@va.gov)) and local problems and issues should be referred to the Working Group.

**21. IS VA FORM 3203 STILL REQUIRED TO BE COMPLETED FOR RESEARCH PHOTOS AND RECORDINGS?**

The requirement for use of VA Form 3202 was removed from the recent update of VHA Handbook 1200.05. Therefore, the reference to its use in the Research Data Inventory tool should be removed if used at the local facility.

**22. SECTION 7 OF THE RESEARCH DATA INVENTORY TOOL ASKS “WILL A COPY OF VASI BE SHARED OR DISCLOSED WITHOUT HIPAA AUTHORIZATION? (THIS IS RARELY APPROVED)”**

This will be clarified. Preliminary inquiry suggests VASI can be shared or disclosed under some specific conditions:

- If it does not contain PHI; in this case HIPAA does not apply (such as proprietary research information that contains no PHI).
- Pursuant to a valid waiver of HIPAA authorization and as long as all other applicable requirements are met (e.g., it is releasable under the Privacy Act).
- As a limited data set as defined by HIPAA and with the required data use agreement in place.
- As a release of information in accordance with those processes outlined in VHA Handbook 1605.1.