BRD for Supporting Sensitive Data

Document Approval	
Requirements Gathering Deadline	**Feb 10, 2016**
Approved by Executive Committee & Stakeholders	[*Initial and date here…*]

Business Requirement Description
Dataverse needs to be able to support sensitive data. The first implementation will be for the Dataverse instance hosted by LTS (Harvard Dataverse). In this case, this means to provide a hosting infrastructure and policies, as well changes in the Dataverse software equivalent to Harvard security level 3. The specific requirements are listed here: https://drive.google.com/open?id=1t5sZiqTacunAXhfjcOnj4l81jNiCpswE8DlbVNqyKgU

At a high level these requirements are:

- encrypting data during transmission and backups
- allowing access to users with a verifiable business need
- requiring stricter password standards, including periodic password change
- actively scanning for security threats
- a policy for maintaining current system patches
- monitoring for security breaches
- scanning for unidentified sensitive data
- maintaining data depositor contact information in case of security issues.

This work is tied to an NSF grant on Privacy Tools for Sharing Research Data. In addition to be a requirement for Harvard Dataverse, this is also of high interest for various Dataverse installations partners (including ODUM, DANS, HMS).

Assumptions, Dependencies, Constraints & Risks
Assumptions include:

- The LTS, Harvard Security, HUIT, and Dataverse teams will work together to identify requirements and plan a solution.
- Work is divided between hosting (LTS and HUIT SOC) and software (Dataverse).
- Support for security level 3 can be accomplished with the existing Dataverse architecture that will serve both sensitive and non-sensitive data. Backups will need to occur through HUIT Data Domain rather than the current tape library.
- Test environments can be assembled from currently existing hardware, no additional hardware is needed.
- There is separate related work to integrate DataTags asssignment to datasets, but this is covered by a separate BRD.

Constraints include:

- There is a time frame for deliverables for the NSF grant (fall 2016 to support up to orange DataTag, which maps to level 3)

Dependencies include:
Working with LTS requires coordinating with them and their availability.

There is intersection and dependency with three other projects (which have - or will have - their own BRDs). For Harvard Dataverse to host sensitive data, the three projects below need to be completed:

1. ***Remote Authentication/Shibboleth:*** Users need to login through an approved federation, such as InCommon to upload or download sensitive data. See the BRD: https://docs.google.com/document/d/1vcAmo2nkFYavAr7OwwXzxM0IFQbkRZYZrrX43q-wqGE/edit?usp=sharing (Phase 1)
2. ***DataTags:*** Workflow to upload sensitive data, and assign a datatag to each file at the time of file upload, and then set the access requirements based on that datatag.
3. ***Modular Upload/Ingest and Download:*** Data upload and download components need to be separated and secured from the main Dataverse application hosting. Sensitive data files never go through the main application. The Upload Component sends the metadata and file location to the Dataverse. This is to support Level 4 data in the future. However, the feature will be implemented sooner for other reasons, such as scalability and extensibility.

Risks include:

- Not finalizing the solution soon enough to work on and meet deliverables.
- Changing requirements due to integration with other projects such as single sign on and DataTags.
- Undiscovered requirements from outside partners such as Odum and partners from the Netherlands.
- We need to review the mapping between DataTags and Harvard Security Levels. This may result in additional requirements on Level 3 to support DataTag orange.

Stakeholders: Dataverse Advisory Board; LTS and HUIT Security

Resources:

- Kevin Condon and Merce Crosas work on FRD
- (1 or 2) Developers (Philip Durbin and TBD) to specify and implement software changes.
- (1) System Administrator, Benson Smith, to specify and implement hosting changes.
- (1) QA engineer, Kevin Condon, to represent Dataverse at meetings and test solution.

High-Level Effort Estimates

2 weeks to produce FRD for software changes
1 month to produce FRD for hosting changes

3-5 months, depending on availability of external resources, for hosting and software changes, including testing (TBD)

External deliverable deadlines:
This work is tied to an NSF grant.

Strategic Goal this aligns with: SUPPORT SENSITIVE DATA