Data Management System/User Interface Project

DMS Solution Options Analysis

Prepared for:
OERA/FORCE

Submitted by:
SEG Consulting Inc.
www.segconsulting.ca
Chad Amirault

Version 2.0
June 6, 2017

Version Log

The following table is used to track each version of this document and the changes that occur at each version.

Version	Release Date	Description	Author
0.1	May 12, 2017	Initial incomplete draft	Chad Amirault
0.2	May 17, 2017	Updated version of incomplete draft	Charlotte Wyatt
1.0	May 19, 2017	First complete draft	Chad Amirault and Charlotte Wyatt
1.1	May 19, 2017	Updated draft with revisions to the conceptual architecture figures.	Chad Amirault
2.0	June 6, 2017	Updated to include additional weightings and scoring around the options	Chad Amirault

Version: An incremental number to identify every version of this document beginning with version 1.0 to distributed versions (1.x – x.x). Note that any document version below 1.0 indicates that the first final version has not been released (0.1 – 0.x).

Release Date: The date the new version of the document was released.

Description: Describes the modifications that occurred for the new version.

Author: Individual responsible for initiating and completing the version change.

Table of Contents

List of Tables

List of Figures

1.0 INTRODUCTION

The Offshore Energy Research Association (OERA) and the Fundy Ocean Research Center for Energy (FORCE) have partnered with SEG Consulting Inc. (SEG) to define a conceptual Data Management System/User Interface (DMS). This project defines a go-forward plan to acquire a best-fit DMS in line with the current and future business needs of FORCE.

1.1 Purpose

The overall purpose of this document is to provide several valid options for designing and building a DMS solution for FORCE. These options are described and compared based on FORCE's needs. These options are provided without recommendation as their purpose is to elicit FORCE priorities.

1.2 Scope

After completing the current and future state analysis based on data gathered during interviews with FORCE staff, the project team developed several options for a future DMS solution for FORCE. These options are described at a conceptual level to guide the next step in developing a conceptual solution design.

In addition to the options, this document includes a comparison between each DMS solution design. Each criterion within the comparison is assigned a level of weight. These criteria are defined based on the current and future state business objectives and solution requirements.

The following is considered out of scope for this document:

- Financial data and documentation;
- Legal data and documentation;
- Software.

1.3 Assumptions

The following is a list of assumptions that were made when developing the requirements in this document:

- Due to the nature of our data analysis, it is assumed that the data and documentation received from FORCE are representative of the data that is in scope for a future DMS.

  NOTE: Data types and formats dictate how a data relationship is established in the DMS.

2.0 FUTURE STATE CRITERIA

This section summarizes the findings previously submitted to FORCE in the *Current and Future State Business Needs and Solution Requirements* document (see document for details). Within this document, the solution requirements (Section 5.1) were used as a reference to build a set of concise comparison criteria (see table 1). These concise criteria will be used to define the DMS solution options.

Table 1: DMS criteria matrix

ID	SOLUTION REQUIREMENT	COMPARISON CRITERIA
SR01	Handle variety of data types	Flexibility
SR02	Easy to use for internal stakeholders	Usability
SR03	Levels or data security	Security
SR04	Read access through a variety of media	Accessibility
SR05	Read/write access (internal)	Accessibility
SR06	Sharing of data between stakeholders	Accessibility
SR07	Hosting outsourcing	Hosting
SR08	Regular back-up cycle	Storage
SR09	Single source (restricted)	Flexibility
SR10	"Live" data connectivity	Flexibility
SR11	Regular data archiving cycle	Storage
SR12	Centralized data storage	Storage
SR13	Scalable capacity	Scalability
SR14	Read/write access (external)	Accessibility
SR15	Consistent organized file structure	Storage
SR16	Support data visualization/presentation	Flexibility
SR17	Direct access to open data	Flexibility
SR18	Cost effective	Cost Effectiveness
SR19	Low maintenance	Usability
SR20	Solution performance that meets FORCE needs	Performance
SR21	Easy for partners to upload data	Accessibility

Table 2 provides descriptions of the comparison criteria identified in Table 1. Table also included a weighting that is meant to represent priority of the business. For example, is quality more important than costs, and is security more important than quality. *NOTE: These weightings have not yet been validated by the business.*

Table 2: Future state comparison criteria

CRITERIA	COMPARISON CRITERIA DESCRIPTION	WEIGHT
Storage	All FORCE data will be stored within a centralized locality with a consistent organization of data. The data will be structured in an intuitive manner with pre-defined data categories. Redundancy of data should be minimized. Archiving and back-ups will occur on a regular cycle.	High
Hosting	The hosting of the DMS must be outsourced. The DMS will not be hosted internally by FORCE. *NOTE: It was confirmed with FORCE staff that they will not host a future DMS. Therefore, this option has been omitted from analysis.*	High

Accessibility	All FORCE data will be easily accessible to all necessary stakeholders. Data will be accessible depending on user status: read only access, read/write access (internal), or read/write access (external).	High
Flexibility	The DMS will support the ability to store a wide variety of data types and formats and will be the single authoritative source for all restricted data. The DMS will support future development of data visualization and will support direct access to open data.	High
Usability	The DMS will be easy to use and will not require a database manager. The solution will require a minimal level of effort to maintain.	Moderate
Security	The DMS will include user-defined levels of security (from open to restricted). Security will be dependent on dataset/data theme.	Moderate
Cost Effectiveness	The DMS will be reasonably priced within the FORCE budget.	Moderate
Scalability	The DMS will be capable of scaling in storage capacity.	Moderate
Performance	The DMS will meet FORCE needs with both direct access and indirect access via web applications and interfaces.	Low
Mapping	The solution should be compatible with Web-based mapping applications for data visualization	Low

3.0 DMS SOLUTION OPTIONS

This section identifies and describes the DMS solution options at a non-technical, conceptual level. Each solution option includes a high-level architecture design diagram. The DMS **solution** options are highlighted:

- Option 1 – Fully Centralized DMS
- Option 2 – Semi Centralized DMS
- Option 3 – Decentralized DMS
- Option 4 – Partial DMS

NOTE that the figures in this section are not technically accurate, they are meant to present the different options from a non-technical perspective. These diagrams are meant to illustrate data the main parts of a future DMS.

3.1 Platform Options

This section describes five DMS **platform** options. Each of these options includes an element of outsourced data storage and hosting. To meet FORCE's DMS solution requirements, FORCE will need one or two data storage and hosting services.

While conducting this platform options analysis, the project team identified a selection of possible data hosting providers for FORCE to consider in reviewing this document. Table 3 includes a list of sample hosting providers along with descriptions of their service and general cost ranges.

NOTE: Both ACENET and DeepSense have verbally confirmed (by phone) that FORCE would be eligible to participate in these solutions. These are cost effective solutions; however, the University does control access to the repository.

NOTE: Information was limited related to the ORE Catapult, however based on our research, FORCE may be eligible for this service as well.

Table 3: Data Storage and Hosting Platforms and Service Provider Options

Type	Solution / Platform	Description	Cost Range
Private	Various such as Amazon and Azure	Base server platform where a customer can develop their own purpose-built solution	$10-20k/year
Academic	ACENET (Compute Canada)	A network (regional clusters) of pre-defined (setup) hosting environments/platforms that exist to support academic research	Minimal to Free
Academic	DeepSense (Dalhousie and IBM)	A proposed future (Fall 2017) hosting environment/platform that is intended to be customized for each industry partner but also support academic research opportunities	Minimal to Free
Government	CNSOPB Data Management Centre	This is an existing environment that houses CNSOPB data intended to be shared to external stakeholders including the public	Unknown if FORCE is eligible

Type	Solution / Platform	Description	Cost Range
Government	ORE Catapult (UK)	A UK government funded organization that provides various support services to private organizations. This service is government subsidized to support economic growth and industry commercialization	>$10k/year (subsidized)

NOTE: The Tidal Energy Atlas, developed in partnership with Acadia University and TekMap is currently hosted by Compute Canada (ACENET). Acadia University has a three-year contract with Compute Canada to provide a hosting environment for free.

3.2 Option 1 – Fully Centralized DMS

This option proposes the centralization all FORCE data to a single outsourced (e.g. cloud-based) hosting service. The hosting service can either be an academic and research hosted environment (e.g. ACENET) or a private environment (e.g. Amazon Web Services (AWS) or Azure).

This option will allow the authoritative source of all FORCE data to be in a single location. Option 1 will support the back-up and archival of all FORCE data. Note that advantages and disadvantages of each option are presented in section 4.2 below.

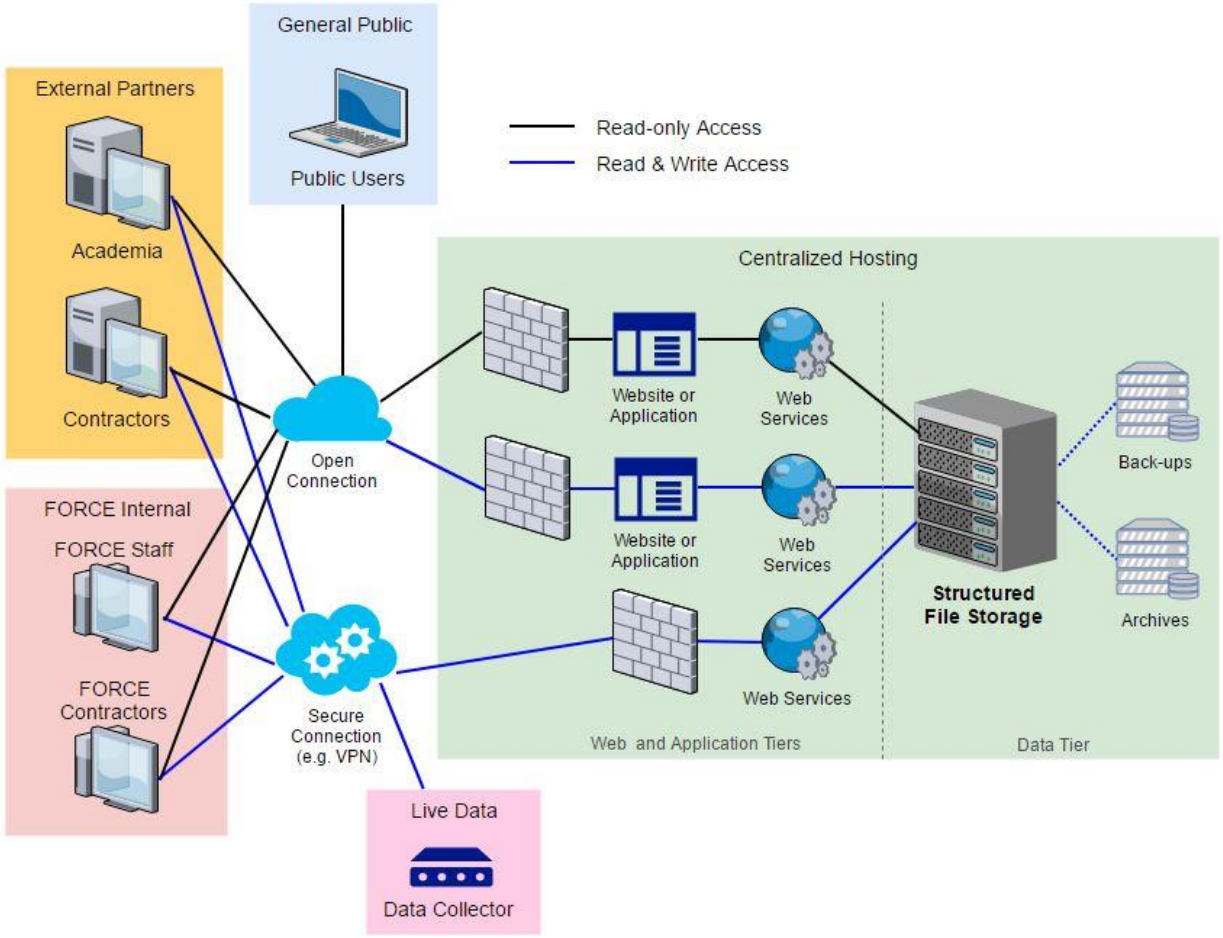

Figure 1 – Conceptual Architecture for Option 1

3.3 Option 2 – Semi Centralized DMS

This option includes a partial centralization of FORCE data through the distribution of data between two (no more than two) hosting environments.

All data is hosted between two physical locations (localities); an academic and research hosted environment (e.g. ACENET) and a private cloud environment (e.g. Amazon Web Services (AWS) or Azure). This data model is meant to minimize (or remove) redundancy of data between multiple hosting locations. Data redundancy means that staff are managing and maintaining multiple copies of the same data. Backups and archives are not considered redundancy as they are not maintained, just stored.

The assumptions around this option is that all non-sensitive (open) data would be hosed in the Academia locality, whereas the sensitive (restricted access) data would be stored in a primary, secure cloud locality.

Option 2 will support the centralization (no redundancy), back-up and archival of all FORCE data.

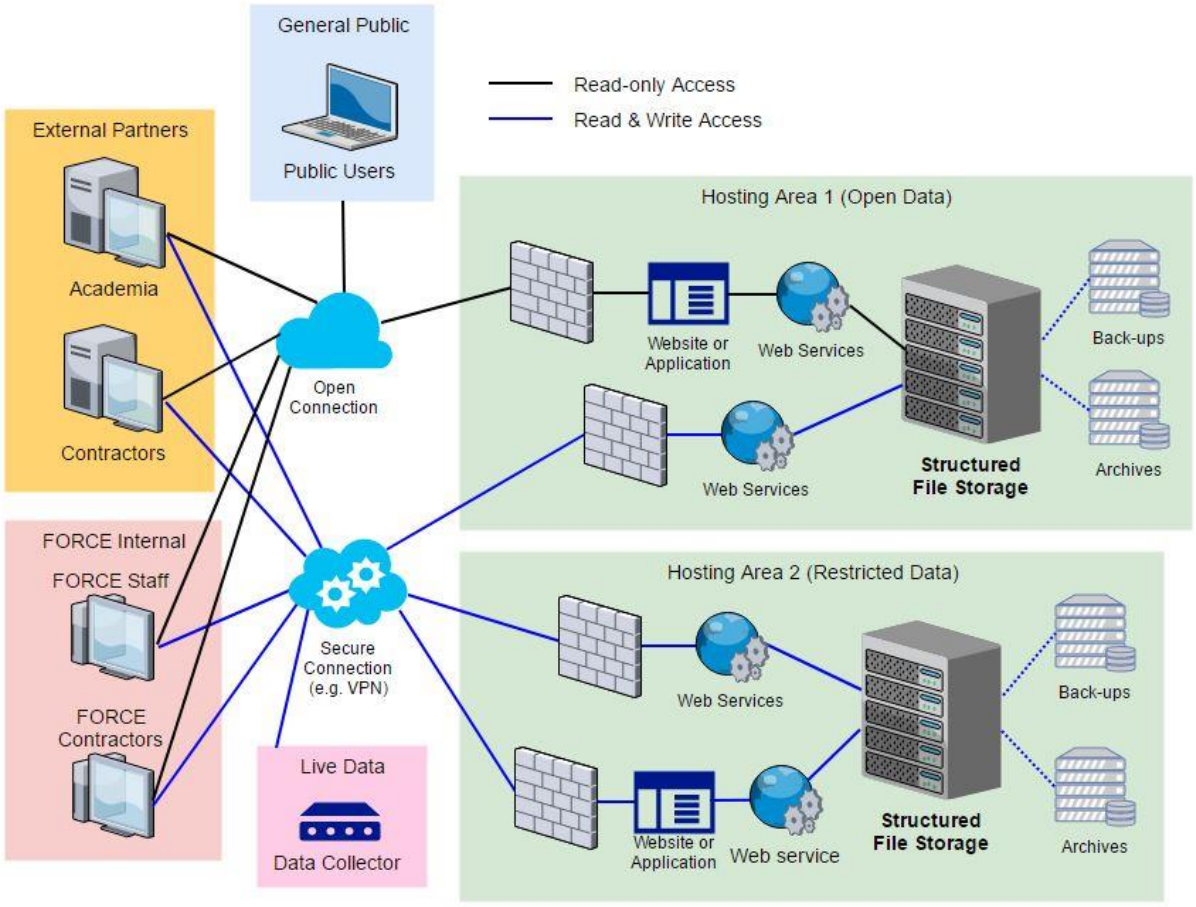

Figure 2 – Conceptual Architecture for Option 2

3.4 Option 3 – Decentralized DMS

This option proposes a decentralized model where FORCE data is distributed between more than two hosting environments. FORCE's current state is closest to this solution as data is hosted by several external partners and stakeholders both locally and through hosting services.

FORCE data is hosted from multiple locations (e.g. personal hard drives, local servers, remote servers owned by partners, and cloud-based servers).

There will be a significant amount of documentation and workflow agreements to ensure data redundancy is minimized and data backup and archive are regularly scheduled.

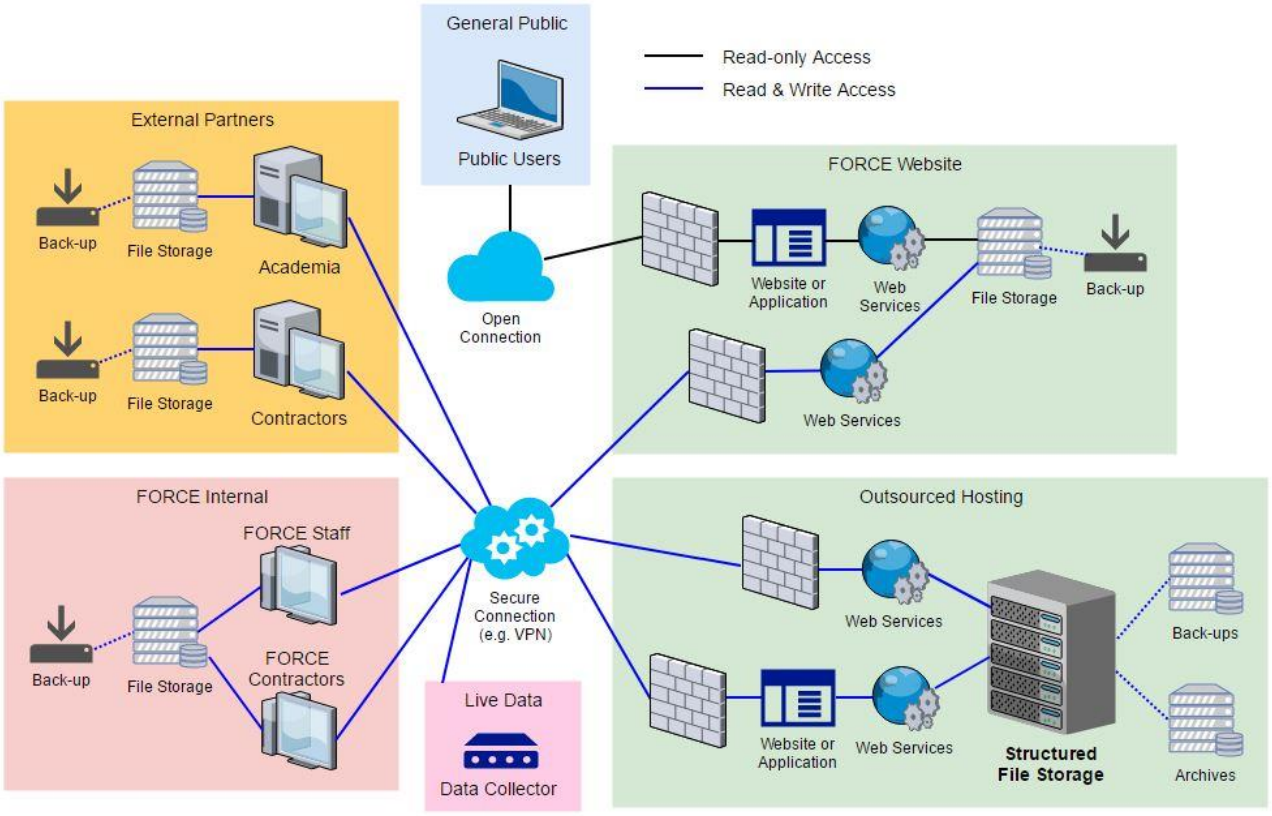

Figure 3 – Conceptual Architecture for Option 3

3.5 Option 4 – Partial DMS

The final DMS solution is similar in architecture as options 1 and 2, however, it would only house a subset of FORCE data. This differs from the other options as this DMS solution focusses on a subset (to be defined by FORCE) of FORCE data. This subset of data will be controlled, backed-up and archived regularly. Whereas, all other FORCE data will be managed as is today, with less process and control.

The hosted data can either be hosted in a single location or two locations, similarly to Option 1 or Option 2. Option 4 will only support the back-up and archival for the subset of FORCE data hosted within the DMS.

4.0 OPTION COMPARISON

To assist FORCE in reviewing and deciding on a preferred DMS option, a comparative analysis based on the predefined, weighted criteria from Table 2 above is presented in Table 4 below

4.1 Comparison Analysis

Table 4 compares each proposed DMS option to the future state criteria. Rather than scoring each option based on the weighting factors in Table 2, the matrix defines who well each option fulfils each criterion. Requirements fulfillment is depicted using the following colouring:

- Green represents a strong fulfillment of the criteria;
- Blue represented a moderate fulfillment of the criteria;
- Orange represents a low or partial fulfillment of the criteria.

Table 4: Future state use by objective

SOLUTION CRITERIA	STORAGE	ACCESS	USEABILITY	SECURITY	FLEXIBILITY	SCALABILITY	PERFORMANCE	COST
Option 1 – Fully Centralized DMS								
Option 2 – Semi Centralized DMS								
Option 3 – Decentralized DMS								
Option 4 – Partial DMS								

NOTE: The colouring provided in Table 4 are based on general industry conventions.

4.2 Comparison Analysis Description

4.2.1 Option 1 – Fully Centralized DMS

The following are the advantages and disadvantages for Option 1.

Advantages:

- Data integrity is high, data redundancy is low
- Single authoritative source means high security and control
- Easiest and cheapest option for maintaining back-ups and archives of data
- Makes it easy to change or re-organize the data structure (e.g. file structure and naming conventions)
- Single database design allowing for streamlined data sharing dissemination processes
- Easy to maintain and administer single data warehouse (locality)

Disadvantages:

- Highly dependent on network connection; internet speeds and drop outs, high traffic times, service provider issues (e.g. regular maintenance windows). This is a low to very low risk.
- May be slower performance speeds since there is a large amount of data in one location

4.2.2 Option 2 – Semi Centralized DMS

The following are the advantages and disadvantages for Option 2.

Advantages:

- Higher level of data security for sensitive data (dedicated hosting environment)
- Having two dedicated environments provides strong flexibility and scalability
- Most users will only have access to the open data, which allows for brief learning curve
- Higher performance speeds since there is less data bogging down the server(s)
- Lower risk of failure since data is spread in two locations

Disadvantages:

- Highly dependent on network connection; internet speeds and drop outs, high traffic times, service provider issues
- Double the cost since there will be two server subscriptions (if free academic solution is not available)
- Requires data design documentation to ensure there is no redundancy in the data and that the source of truth is known to all users

4.2.3 Option 3 – Decentralized DMS

The following are the highlighted advantages and dis-advantages for Option 3.

Advantages:

- Not much change required from current state
- Very easy connectivity for staff with data stored locally and on external drives
- May have lower costs since there is no server (hosting) subscription
- Data can be partitioned so that each smaller database represents a specific data theme

Disadvantages:

- Poor data accessibility (beyond those that have external drives)
- High effort required to ensure everyone is using the best copies of data
- Very difficult to manage version control on data
- Poor data control and security (back-up of data in manual)
- High risk of failure and data loss, due to hardware quality
- High probability of data redundancy between staff and various storage locations

4.2.4 Option 4 – Partial DMS

This option closely resembles the advantages and disadvantages of options 1 and 2 however, they only pertain to a subset of FORCE data. The following are the highlighted advantages and disadvantages that differentiate option 4 from options 1 & 2.

Advantages:

- Costs are lower than options 1 and 2 due to less data being stored in dedicated environment(s)
- Performance in higher than options 1 and 2 due to less data

Disadvantages:

- Less control, security and accessibility to data the data not centralized in dedicated environment(s)

5.0 SUMMARY AND CONCLUSION

This summary and conclusion section provides our expert opinion of what direction should be the most practical and effective for FORCE. However, to create the following, we had to make the following assumptions about the data to be stored and managed (that need to be validated by FORCE).

1. **FORCE wants to have secure and controlled storage of all RAW and final versions of data.** This means that interim versions of data will not be managed by FORCE as they can be re-created from the RAW.

2. **FORCE wants to make a large portion of their data publicly available, however keep some of their research to restricted access.** This means that the majority of FORCE data can be stored in any private or academic hosting environment. However, depending on the data accessibility and sharing agreement with the hosting provider - especially academia; this data may need to reside in a private hosting area. *NOTE: DeepSense may also be any option, as they said that they do not require full access to the data in their environment.*

3. **FORCE wants to make data accessible to a wide variety of users (internally and externally) and applications.** This means that data should be located/accessed in one spot (centralized where possible) and be well structured (e.g. folder structure). Data redundancy and duplication should be minimized, as much as possible. Trust in the data comes from ensuring data integrity (e.g. currency).

4. **FORCE will continue to work with external partners.** This means that FORCE partners will continue to collect, store and manage research data that is either partially or fully owned by FORCE. In this case, partnership agreements should be established with partners regarding how FORCE data is stored and managed through out a research project. FORCE data can be held by a partner during project execution, however must be provided to FORCE as deliverables of the project.

5. **FORCE want to improve the presentation and visualization of its data.** This means that FORCE data should be organized in a way to be best compatible with websites and web applications. This is best done by publishing web services (e.g. web map services) from your environment.

Based on the above assumptions, SEG considers options one or two to be the best options for FORCE. FORCE should centralize all their RAW and final version of their data to one or too main locations. This does not include interim version of data.

The following are the primary differences between Option 1 and Option 2:

- If FORCE is fine with an Academia partner controlling access to all their data – including their proprietary research data, then FORCE should go for Option 1 under the existing ACENET program or wait for the future DeepSense program. These programs include heavily subsidized hosting services. These providers would be less expensive as a private hosting provider like Amazon or Azure.
    - FORCE should still understand the data admiration rules under these programs before concluding with these providers.
- If the DeepSense project is approved and FORCE is eligible to enter into this subsidized environment program, then FORCE should consider Option 1 using the future DeepSense environment.
- If FORCE wishes to keep and control the restricted access to some of their proprietary data and wants to eliminate the risk associated with an academic partner hosting their data, then FORCE should go

to Option 1 with a private sector hosting provider. There are many providers out there that offer various levels of service related to both quality and performance. This option is the most expensive.