

# RECLAMATION

*Managing Water in the West*

## Investigating the Need for a Center for Programmatic Data Management

Research and Development Office  
Science and Technology Program  
Final Report ST-2015-7671



U.S. Department of the Interior  
Bureau of Reclamation  
Research and Development Office

## **Mission Statements**

The U.S. Department of the Interior protects America's natural resources and heritage, honors our cultures and tribal communities, and supplies the energy to power our future.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188		
<b>T1. REPORT DATE</b> 3/16/2016		<b>T2. REPORT TYPE</b> Research		<b>T3. DATES COVERED</b> 2015	
<b>T4. TITLE AND SUBTITLE</b> Investigating of the Need for a Center for Programmatic Data Management			<b>5a. CONTRACT NUMBER</b> 15XR0680A1-RY1541RD201527671		
			<b>5b. GRANT NUMBER</b>		
			<b>5c. PROGRAM ELEMENT NUMBER</b> 1541 (S&T)		
<b>6. AUTHOR(S)</b> Thomas Heinzer, Diane Williams			<b>5d. PROJECT NUMBER</b> 7671		
			<b>5e. TASK NUMBER</b>		
			<b>5f. WORK UNIT NUMBER</b> MPGIS		
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> U.S. Bureau of Reclamation			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>		
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Research and Development Office U.S. Department of the Interior, Bureau of Reclamation, PO Box 25007, Denver CO 80225-0007			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b> <b>R&amp;D: Research and Development Office</b> <b>BOR/USBR: Bureau of Reclamation</b> <b>DOI: Department of the Interior</b>		
			<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>		
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Final report can be downloaded from Reclamation's website: <a href="https://www.usbr.gov/research/">https://www.usbr.gov/research/</a>					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT (Maximum 200 words)</b> This paper summarizes the advantages Reclamation programs could achieve by the establishment of a centralized information system architecture. Advantages include facilitation of interactive data assimilation, application development, structured data management, stakeholder interaction, data dissemination and public transparency. For the purposes of this paper, this centralized architecture will be referred to as a "Center for Programmatic Data Management", or CPDM for short.					
<b>15. SUBJECT TERMS</b> CPDM, Data, Management, Program					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  U	<b>18. NUMBER OF PAGES</b>  16	<b>19a. NAME OF RESPONSIBLE PERSON</b> Thomas Heinzer
<b>a. REPORT</b> U	<b>b. ABSTRACT</b> U	<b>c. THIS PAGE</b> U			<b>19b. TELEPHONE NUMBER</b> 916-978-5273

# PEER REVIEW DOCUMENTATION

## Project and Document Information

Project Name: Investigating the need for a Center for Programmatic Data Management

WOID\_X7671

Document Paper

---

Document Author(s) Thomas Heinzer, Diane Williams

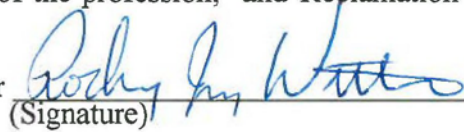
Document date 8 April 16

Peer Reviewer Rodney Whittler

## Review Certification

**Peer Reviewer:** I have reviewed the assigned items/sections(s) noted for the above document and believe them to be in accordance with the project requirements, standards of the profession, and Reclamation policy.

Reviewer

  
(Signature)

Date reviewed

8 Apr 16

# Acknowledgements

The author would like to thank the many people that contributed their thoughts and experiences, which were the centerpiece of this paper. The data allow report readers to really get a feel for the consensus levels of various issues.

# Executive Summary

There are a number of policies requiring transparency, inclusiveness and efficiency; however, policies to facilitate public transparency do not exist.

Reclamation programmatic needs relating to data management, data dissemination and stakeholder interaction are not being met.

Impacts of restricted systems on programs are significant (lack of stakeholder involvement, data entry errors, etc.).

A solution to the issues is paramount as CVPIA, SJRRP, TRRP and CVO constitute 75% of MP Region's budget.

The objective of this effort was to interview program and data managers in an attempt to summarize the advantages Reclamation programs can gain through the centralized, managed use of technology and information systems for interactive data assimilation, management, stakeholder interaction, dissemination and public transparency.

The research summarized the views of many program and data managers within Reclamation. After establishing a consensus that some form of Center for Programmatic Data Management (CPDM) is strongly needed, a general discussion investigates what a CPDM might look like, and the functions that it could support.

The paper is intended to be a spring board and reference document for further discussions on establishing a CPDM.

# Contents

- Introduction..... 1
- Issue ..... 1
- Method ..... 1
- Interviews..... 2
  - David Mooney, CVPIA Administrator ..... 2
  - Jim Nagode, Reclamation Data Architect ..... 2
  - Ali Forsythe – Manager, San Joaquin Restoration Program ..... 3
  - Rod Wittler – Fish Resource Area Coordinator – CVPIA ..... 3
  - Katrina Harrison – Program Engineer (Civil), San Joaquin Restoration Program ..... 4
  - Sonja Kokos – Adaptive Management Specialist, LCR MSCP..... 4
  - Jerry Johnston – Director, Information and Technical Services, OCIO ..... 5
  - Greg Gault - BORGIS Systems Manager ..... 5
  - Kurt Wille, Manager, Emergency Management and GIS Group, TSC..... 6
  - Bryan Iversen – IT Specialist, Infrastructure Operations Services Group..... 6
  - Syd Poulton, IT Applications Chief, Mid Pacific Region..... 6
  - Doug Clark, PhD, Reclamation Data Policy Analyst ..... 7
  - Eric Petersen – Data Manager, Trinity River Restoration Program..... 8
  - Tom Heinzer – Manager MPGIS Service Center..... 9
- Proposal..... 10
- What a CPDM Might Look Like ..... 10
  - Physical Infrastructure..... 10
    - Databases ..... 11
    - Application Software..... 11
    - GIS applications: ArcGIS Server, SDE, Access to Organizational ArcGIS Online ..... 11
    - Other third party visualization and analysis products. ..... 11
  - The Need for Geospatial ..... 11
  - Technical Staffing ..... 12
  - Types of Data the CPDM Would Handle..... 13
  - Data Management ..... 14
  - Document Management Systems..... 14
  - Funding..... 15
- Conclusion ..... 15

# Introduction

This paper summarizes the advantages Reclamation programs could achieve by the establishment of a centralized information system architecture. Advantages include facilitation of interactive data assimilation, application development, structured data management, stakeholder interaction, data dissemination and public transparency. For the purposes of this paper, this centralized architecture will be referred to as a “Center for Programmatic Data Management”, or CPDM for short. This paper is intended to be a foundational exploration to facilitate further work.

## Issue

We often hear that large programs in Reclamation need technical assistance to transparently provide collaboration tools and scientific data to fulfill their missions. As mitigation and restoration projects are on the increase, data acquisition and analysis requirements have skyrocketed. On one hand, the inability of the programs to efficiently share data with stakeholders is hampering required interaction with external scientists and the general public, directly impacted our mission. On the other hand, security concerns exacerbated by the recent OPM breach are well-founded and should be taken seriously.

Historically, many IT departments had ‘applications groups’ that could at least partially assist the programs. Due to a variety of reasons, those groups no longer exist. As a result, no mechanism for assistance currently exists so the programs are forced to independently forage for help, often from external consultants and delivery systems. While results may be realized in the short term, the potential benefits that could be gained by agency-level oversight, adherence to data management standards, security support, software development standards and lifecycle management are largely absent. This nonexistent supporting environment severely reduces Reclamation’s ability to meet long term programmatic and stakeholder needs.

Initiatives like Open Data are providing a promising general direction for data transparency and collaboration. Reclamation programs however, typically have very specific legislated and/or regulatory requirements which call for specific solutions. These solutions require teams with focused skillsets to implement data management and collaboration strategies, as well as support and maintain these implementations.

## Method

In order to formalize the problem and the need for a better approach, a number of program managers, data managers, and IT managers were interviewed. Interestingly, 100% of those interviewed felt that a CPDM in some form to assist Reclamation programs was a good idea. All of those interviewed felt that the programmatic needs, IT Transformation, and cloud platform availability have evolved to a point where a CPDM is achievable, and that historic obstacles can be overcome with the ‘right people in the room’, proper oversight and investment.



The results of the interviews will be used to formulate comments regarding what a CPDM might look like. It should be understood that these early stage comments will necessarily evolve.

## **Interviews**

The following is a summary of the comments from a number of program managers and data managers within Reclamation. A focused attempt was made to accurately capture the comments, however any questions for clarification should be addressed to the individuals. Inclusion of complete interview transcripts would be too unwieldy, so the remarks have been condensed and paraphrased for brevity. Some views that became consensus are summarized in other areas.

### **David Mooney, CVPIA Administrator**

The main issue is that we have a collection of subject matter experts (SMEs) in a program, and we have a computing infrastructure, but we don't have a group to interface between the two. For example, if a program needs to have external web applications, they could contact the CPDM, and the CPDM personnel would then coordinate on the infrastructure/application side to make the outside web and technology happen.

A lot of weight is being given to security in risk discussions, but insufficient weight is being given to the benefits of a more open IT policy and data approach.

The programs are letting contracts with different components containing a lot of seemingly disparate stuff. However, in many cases these acquisition efforts are similar and there are opportunities where a solution architect could design single databases that could handle these data flows, which would greatly enhance accessibility, querying and management.

### **Jim Nagode, Reclamation Data Architect**

The idea of merging programmatic needs with IT services has been floated before, and was sporadically implemented in the regional offices. However, centralizing the effort has not been discussed seriously. There is support for the notion of a CPDM and the feeling that at least initially, it could be supported in the form of directed S&T funds. We already spend millions on data acquisition, let's look at doing it better.

A cloud-based solution is appropriate, and the only real legal requirement pertaining to data protection is around PII. FOUO designations are self-imposed and self-regulated. One area of exploration is the possibility of creating a new service center that is a division in IT. It may be prudent to consider vendor solutions for various areas (e.g., Socrata).

## **Ali Forsythe – Manager, San Joaquin Restoration Program**

A CPDM is a great idea as the SJRRP has had significant problems with data handling and dissemination. Since the public paid for the information, they should have access to it. The SJRRP staff receive a large number of time consuming queries and requests for data from students, citizens, and teachers (e.g., to use SJRRP as an example in a science lesson). Established applications and data delivery systems would be a big time saver and would free up staff for direct programmatic functions.

SJRRP would be willing to consider funding at some level.

## **Rod Wittler – Fish Resource Area Coordinator – CVPIA**

As lead scientist 2004-2008 for the Trinity River Restoration Program, I have worked in a multi-state-county-federal and stakeholder group where the pledge was made to be transparent and also be the hub for all science supporting the adaptive management process.

We struggled in 2004 to determine what hoops we had to jump through to develop this transparent ‘Integrated Information Management System’. There was no guidance. As the main repository, the idea was there would be open access to all the parties involved (upload and download). It was understood that the various data that were uploaded would have different levels of QA/QC. As a whole, this hub in theory supported the “Adaptive Management Program”. In the end, with support from the R & D office, we utilized an external contractor, and an online data portal was established, meeting a subset of our goals.

As a manager, I need a cookbook, a recipe of how we can set such a system up within the security and data management requirements of the agency. The goal is to have our agency, other agencies, tribes, NGO’s, etc. swim in the same data pool.

A CDPM would be standardized and develop a level of trust and a sense of collaboration, cooperation and participation. One approach would be to have different levels of QA/QC: raw data, provisional data, and finalized data. Data in the earlier stages would be labeled “Use At Your Own Risk”, while finalized data would have more formal metadata assignments.

## **Katrina Harrison – Program Engineer (Civil), San Joaquin Restoration Program**

The idea of creating a CPDM is definitively a good idea.

The CPDM should assist with bureaucracy, servers, software development, and have standardized templates anyone can use if they don't want a custom application. For example, getting approval for cloud-based databases, database support, developers to assist with interfaces

and building database tables, set schemas and interfaces that anyone in the agency can grab to add to their own websites, have a server bank to hold all the data so no one has to worry about storage space, etc.

As for SJRRP activities, if a CPDM would have been in place, MPGIS efforts could have been focused on other SJRRP activities instead of wrestling with bureaucracy.

## **Sonja Kokos – Adaptive Management Specialist, LCR MSCP**

LCR MSCP is ten years in, and is required to monitor system wide (e.g., birds, fish, etc.). Historically, no standards were in place for data collection and storage. When attempting to implement adaptive management, the lack of standard data collection and storage methodologies precluded the program level analysis needed to implement plans.

Because of the lack of standardized data and volume of the data, it is impossible to programmatically analyze LCR MSCP. Proper emphasis on data verification and QA/QC is lacking. The staff doesn't have the background necessary to perform appropriate QA/QC. Additionally, LCR MSCP has lacked understanding and support of technologies (e.g., data dictionaries, enterprise schemas) to adequately support programmatic operations. Nor is there training or exposure to data management principles.

Reclamation restoration programs need to be able to solicit help with contract language and content review for various efforts which relate standards and business practices. A CPDM could connect program personnel with people who know what is current and commercially available, which is rarely visible from the programmatic vantage point.

Reclamation needs to start being leaders and get up to speed with some of these new technologies. LCR MSCP has made some progress with data dictionaries for collection devices, and is trying to standardize the data sets across the program.

The LCR MSCP is required to store large amounts of data for the life of the program. Not having Reclamation wide support in terms of technical expertise in the areas is an issue that will continue to be exasperated as data collection continues annually. Also, MSCP is required to provide or disseminate much of this information/data, and if they can't disseminate, they can't meet their objectives.

There is a need for centralized Database Administrators (DBAs) and application developers to do business intelligence, as well as a need to serve data on the web both internal and external. Hiring technically qualified personnel historically have fallen short. Reclamation doesn't pay enough to attract the required skill levels. Internal personnel use Excel and Access only. Need enterprise databases to meet the big picture goals.

A need for professional spatial analyst/geodatabase skills beyond creating maps has been identified.

LCR MSCP would be willing to give funding support.

## **Jerry Johnston – Director, Information and Technical Services, OCIO**

This initiative should be at the agency level, and noted that staffing and funding levels would have to be permanently in place to be successful. Personnel need not be co-located. The best place for a CPDM is in the DOI Cloud infrastructure, which would alleviate some of the problems with external interaction. The cloud systems would likely be designated as FISMA moderate. Moving in the direction of a CPDM is certainly the way of the future.

## **Greg Gault - BORGIS Systems Manager**

The programs are charged with a mission and have money, but no good solutions exist to date to handle the programmatic data component. In place of database solutions, a ‘spreadsheet culture’ has proliferated within Reclamation. There are different data collection protocols between programs, people, and sensors, which compound the problem. It would be beneficial to establish a central service group or office to provide expertise and technical resources to support programs and projects in meeting mission data management needs that also meet the Open Data Policy mandate and reporting requirements.

A CPDM would have to be staffed with the ‘right’ people (the challenging/missing piece). Initially, a system architect would need to be consulted. The infrastructure would have to be a high availability, robust system. Certain clients of the system would have to assume a data responsibility role.

One suggestion is to have the programs put together rough figures on what they spend now and what they estimate they currently lose by not having a CPDM in place. Another problem is granularity and breadth of data, which will require enterprise data managers and enterprise application developers. Many of the clients want statistical systems (i.e., statistics on the data). We should perhaps look at what other people are doing.

You can’t just build based on requirements. General enterprise storage and management techniques need to be considered.

Part of problem is that the IT world is all about infrastructure, i.e., they don’t have anybody that does enterprise database management and application development. They build the container, but that is it.

## **Kurt Wille, Manager, Emergency Management and GIS Group, TSC**

The proposal of a CPDM is a great idea. The CPDM would have to be staffed at the appropriate level, most notably application programmers. Instead of a pilot, we just need to ‘do it’. It is likely that a CPDM would increase data stewardship in the agency.

## **Bryan Iversen – IT Specialist, Infrastructure Operations Services Group**

IT is at a crossroads. Currently, the technology is at a point where people (program personnel and stakeholders) should be able to get what they want. There are too many silos, and not much will change until this issue is addressed.

There is an initiative coming down related to water and open data. There is a need to at least adopt a data structure for water related activities. There is not a lot of guidance. We need a team – akin to a board of directors – that has ownership of water data initiatives.

The problem is systemic. Enacting change in the government is like steering the Titanic. We could learn a lot from the private sector. The position and engagement of leadership needs to evolve and embrace current technologies to meet these emerging issues.

## **Syd Poulton, IT Applications Chief, Mid Pacific Region**

Yes, I agree it is a good idea. We need a standard user interface and data structure across the Bureau for the same types of program data. We need a system that sits on a modern technology stack.

CPDM would be a single system running on a public or private Cloud limited to Reclamation staff. The system would reduce security vulnerabilities, licensing costs, and points of failure. It could be the feeder system for Public Facing sites.

Staffing considerations:

Each region would be given data manager access in order to QC the data.  
Each region would be given system access in order to manage their regions interfaces.  
Each region would be accountable for the quality and availability of their data.

## **Doug Clark, PhD, Reclamation Data Policy Analyst**

The center is a good idea. Reclamation needs a pool of data stewardship/management experts. Based on a survey the Data Stewardship Core Team (DSCT) did last year, these folks need to have *expertise* in the following areas:

- *Data Management Technology.* Understanding of the technologies that pertain to data management and how and where to obtain the relevant expertise. Examples: data planning protocols, web services, data collection protocols, data logging, data modeling, data standard development, database development, application development, and cloud computing.

- *Collaboration.* Managing collaborative data collection efforts. Obtaining collaboration, buy-in, and support for data stewardship practices from management, partners, researchers, principal investigators, advisors, reviewers, and/or stakeholders.
- *Comprehensive Data Management Planning.* Data collections must be viewed and managed in an integrated, holistically planned manner through the entire data lifecycle: design, acquisition, evaluation, maintenance, access, analysis, reporting, archival. Comprehensive data management planning means looking at the whole data management effort, along with each interdependent process and component, when undertaking a data collection. This includes, but is not limited to management of the following:
  - 1. personnel and turnover,
    2. relevant evolving technology,
    3. acquisition and maintaining relevant expertise
    4. budgets,
    5. multiple databases (often in a variety of locations) over many years,
    6. timely, transparent, and successful data sharing and reporting,
    7. data standards, protocols, naming conventions, and metadata,
    8. data quality and integrity,
    9. locating, managing, and/or applying standards to legacy/historical data sets and metadata,
    10. contracting (i.e. making sure contractors get the right data to the agency in a timely manner),
    11. project scope creep, and change management,
    12. voluminous data sets and related demands on IT systems,
    13. training of personnel, contractors, etc. in data management best practice,
    14. back-up plans in the event of data loss,
    15. data interpretation: objectivity
    16. access, permitting, and other challenges such as remoteness in collecting field data,
    17. data requests: who is eligible to obtain the data and under what requests,
    18. working with decision makers to ensure that the data collected are the data they require to make sound assessments, rulings, and judgments, etc.
    19. sampling design and statistical power analysis,
    20. analytic processes,
    21. data screening and release processes,
    22. data sharing agreements.

Every DSCT survey respondent indicated that there is a requirement for more rational processes for data management within the agency.

In addition to planning for single project data collections, plans should be laid for the management of certain selected data sets across the agency: e.g. anadromous fish, invasive species, endangered species, etc. These collections should be overseen with a business data steward.

Lastly, the center ought to build partnerships with other ongoing data management efforts such as what PNAMP (Jen Bayer), BLM (Tom Chatfield), and USGS (Viv Hutchison) are doing. This could save a lot of time and effort, because these folks are pretty far down the road.

## **Eric Petersen – Data Manager, Trinity River Restoration Program**

In a way, TRRP already has a 'center for programmatic data management' (CPDM) in the form of a library for data sets and reports completed among the partner agencies that form TRRP. I would consider our activities to be a good model for similar centers.

My vision is based on the assumption that multiple offices are participating in a 'program'. The center would fulfill two roles: (1) maintaining a reference "library" of data collected across the program, and (2) providing that data as needed for future use. It should be the responsibility of the individuals collecting the data to organize and document it, as they are closest and most knowledgeable about collection protocols, sampling design, etc.

The CPDM should require that data be documented to some degree for acceptance into the library. The library itself should be a server of sufficient capacity to store the volume of data, with regular backups, however additional thought needs to be given to an off-site mirror or some other assurance that data cannot be accidentally deleted.

Making data available online either via intranet, or the public web, will be dependent on data sensitivity and end-user needs. In the case of TRRP, much is made available publicly, but it is only a small portion of our total data library. I also recommend maintaining reporting products as well, as they often inform data documentation and future uses, plus additional information in reports can sometimes be used as data in meta-analyses.

## **Tom Heinzer – Manager MPGIS Service Center**

The programs are in need of professional assistance to meet their regulatory requirements. They do not have the internal resources to even begin to architect scientific and geospatial analysis and data delivery systems that promote transparency and collaboration. Even if they did, it would be a waste of taxpayer money to have these high level groups replicated in each program when there could be one central resource.

A properly equipped data management center would be able to say to the programs, “we can assist you with application development, data storage and management, cloud technologies, collaboration tools, data collection methods, and contract procurements.” The centralized methodologies of the CPDM would *naturally* lead to data standards.

A CPDM would be positioned to evaluate the potential consolidation of similar data collection efforts across the agency. This consolidation could take the following forms:

1. Assist in standardized data collection schema to enable consolidation for enterprise efforts.
2. Develop translators for necessarily different data collection formats to facilitate enterprise efforts (e.g. different sensors arrays have different data formats, but collect the same data type).

The CPDM would need a manager that understands scientific, engineering monitoring programs, databases, GIS, software development processes, etc. Staff in the various disciplines would need to be focused subject matter professionals (e.g., qualified DBAs). This may require contract staff for some disciplines. A good place to house the CPDM might be IT, however it is unclear at this point if our internal IT personnel are going to be working in the DOI cloud infrastructure or just internal to the agency.

In situations where contractors would be needed, assistance needs to be available for the spectrum of the procurement process: RFP formulations, evaluation of offers, and access for external workers.

A certain level of base funding is necessary to sustain the CPDM, it is unlikely to succeed if it relies totally on programmatic funding. Reclamation needs leadership in this area going into the future. We are going to be collecting, housing and making decisions with data for a long time, and these data are going to be accessed by future generations to solve important societal problems.

## Proposal

The proposed recommendation is that steps be taken to further examine the feasibility of establishing a CPDM within Reclamation. The interview results suggest a CPDM would assist with the following:

1. Data collaboration tools
2. Security issues
3. Application development
4. Contracting procedures
5. Liaison between program staff and IT
6. Liaison with Public Affairs
7. Data management strategies
8. Data storage
9. Reporting services
10. Interaction with non-Reclamation program partners



The interviews also suggest that if groups representing regulatory requirements, security objectives, procurement requirements, infrastructure resources were brought to the table at a high level, a solution could be architected that meets the requirements of all concerned. This collaboration would provide a sound foundation to our programs, our open data initiatives and the quest for better transparency.

## What a CPDM Might Look Like

A detailed picture of what a CPDM might look like will require input put from many disciplines. However, what follows is a general outline sketched out by talking to program managers and personnel, solution architects, developers, and system managers within Reclamation.

### Physical Infrastructure

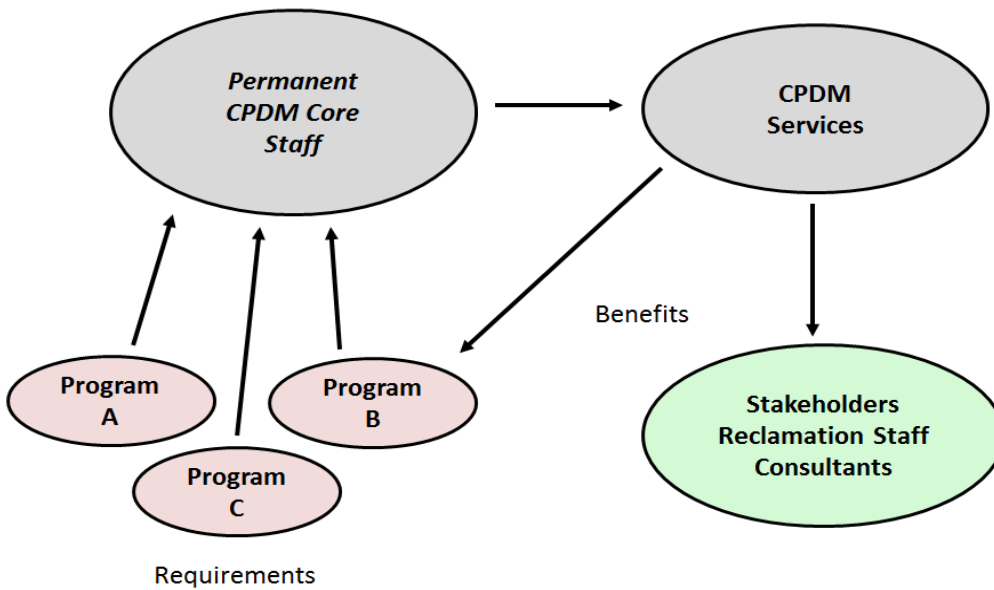
The following presents options to consider when formulating the physical architecture of the CPDM.

The core system infrastructure of the CPDM will consist of:

1. Servers internal to the DOI domain
2. One or more large cloud servers sitting in the DOI Gov cloud platform. This platform will enable flexible scaling of virtual machines both vertically (databases) and horizontally (web applications, mapping applications).

Some of these servers will be web servers running some or all of the following: IIS, Apache Tomcat, NodeJS, and PHP.

The CPDM servers outside of the DOI domain will have a FISMA specification likely set to 'FISMA Low or Moderate'. Application and data access will be anonymous or optionally controlled using appropriate security protocol (e.g., SAML, etc.). Security controls will be set to allow consultants and contractors to perform work directly in the CPDM. Virtual Desktop Infrastructure (VDI) technologies should be considered as a viable external access option.



**Figure 1. Schematic illustrating the relationships between requirements and benefits**

## Databases

The CPDM servers should be running enterprise database systems, which would likely be Oracle and SQL Server.

## Application Software

Collaboration Tools: SharePoint, Google Tools (Docs, Drive, and Hangout), Live Meeting, FTP services

## GIS applications: ArcGIS Server, SDE, Access to Organizational ArcGIS Online

Social Media tools where it makes sense to facilitate public interaction and issue tacking (e.g., Twitter)

## Other third party visualization and analysis products.

## The Need for Geospatial

Mapping applications are a critical component of many programmatic functions. As early as 1992, it was estimated that 70%+ of all government data had a geo-locational component (*Franklin, Carl and Paula Hane, "An introduction to GIS: linking maps to databases," Database. 15 (2) April, 1992, 17-22*). With the proliferation of GPS-enabled devices, locational analytics is playing an ever more increasing role in programmatic efforts and is certainly

transforming the way we look at the world. GIS integrates the scientific disciplines by providing an underlying spatial framework.

Despite this, GIS as a transformational architecture is not being implemented in Reclamation at the pace and degree it should be. Empowering employees, consultants, and stakeholders, with geospatial maps and applications is essential to the process of becoming collaborative and transparent.

## Technical Staffing

A critical consideration is to ensure the success of the CPDM is the acquisition and *retention* of highly qualified, dedicated staff. A strong argument can be made that Reclamation has *underestimated the on-board staff required to successfully operate in the information management and dissemination era.*

Federal skill levels and contractor salaries need to be at a level that attracts and retains the required talent. The industry mantra is that the need for colocation of staff is far less important than team composition and formalization of sound processes. In the case of Reclamation, the CPDM could be established as a service center within IT.

The CPDM core staff will contain IT professionals, data management professionals, DBAs, and application developers. Additionally, liaisons will exist to bridge the gap between CPDM core staff and program subject matter experts (biologists, geologists, modelers, engineers, meteorologists, hydrologists, soil scientists, etc.).

A potential CPDM staffing list might be:

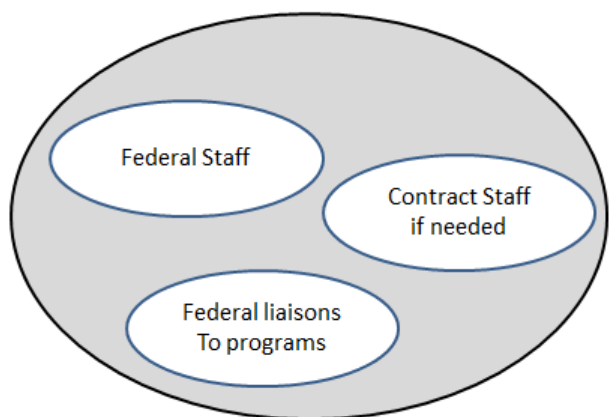
Liaisons – these people establish the link between the core CPDM personnel and the programs.  
DBA(s) – Oracle, SQL Server, or other (may be more than one)

Security Analysts – these people would assist in implementing access control, and are especially needed when implementing external cloud asset access control.

GIS Professional(s) – analysts and web mapping developers

Application Developer(s) - These developers should be versed in the following: modern design patterns, bootstrap style responsive design patterns, UI design patterns in mobile platforms.

Systems Integrator – this is a person that can construct interoperability or interface components between systems. For example, many times programs need applications that link GIS features with document management systems.



**Figure 2. Staffing of the CPDM (the permanent core staff would interface with and receive direction from various Reclamation activities).**

## **Types of Data the CPDM Would Handle**

The programs are typically involved in environmental monitoring functions, which require the management of large amounts of collected data. The problem is the sheer volume and diversity of the data makes it very difficult to manage without proper investment.

Business data stewards should be identified and oversee mission critical data sets where consultation with the center is required. The Reclamation system BORGIS should be brought in to assist with the geospatial component of these data.

These data would likely include the following:

- (1) monitoring data (time series)
- (2) imagery
- (3) real property data describing the location and disposition of Reclamation lands, buildings, structures, and facilities
- (4) data related to the safety and security of Reclamation dams, power plants, and other critical infrastructure
- (5) threatened, endangered, and invasive species;
- (6) water storage or delivery
- (7) power generation and delivery
- (8) facility operations and maintenance
- (9) cultural resources

The centralized management and accessibility of these base data would establish a foundation for peripheral programmatic data to be built.

The CPDM could develop procedures for working seamlessly with the Open Data Team and with the various data communities of interest.

## **Data Management**

There were two interesting lines of thought that came out of the interviews, which will facilitate further discussion. In general, people who had a background in data management felt that formal data management principles should be implemented top down in a CPDM. On the other hand, some programs (while conceding the importance of data management) felt the overly restrictive nature of classic data management strategies and security protocol can hamper collaboration and stakeholder interaction.

For example, requiring FGDC metadata from outside agencies or contractors can cause situations where they simply won't send it or require funding (when in reality the most important thing is just to get the data). The author knows of one internal case where a Reclamation program was asked to put FGDC metadata on some GIS data that was to be sent to another agency and the program decided not to send it citing it took too much time.

What the planning phase of a CPDM would do is bring the various concerns to the table. The programs know they need to do something regarding data in general. The data management side wants to help them out, and on some issues certainly can. It seems that the various concerns could develop an overall strategy to make CPDM work, if agency management were to invest in that direction.

One approach that could come out of the discussions is that the CPDM will use less structured and less bureaucratically restricted methods for data at certain stages or status (e.g. temporary, incomplete, currently in collaboration). Under this strategy, finalized or authoritative datasets are restricted from modification and adorned with final metadata, and working datasets are less restricted where it makes sense. A similar strategy is currently being employed in Reclamation's BORGIS System's data tiers and the approach has clear advantages.

## **Document Management Systems**

Reclamation has struggled with document management systems (DMS) since the 80's. The problems seem to be:

1. Data management systems are inherently complex
2. One size does not fit all
3. DMSs can be unintuitive when compared with traditional file system storage
4. Management is hesitant to endorse the use of systems which require expensive migration efforts
5. Management is hesitant to 'lock in' to one system

There are obvious benefits to formal DMS's, for example searching, versioned editing, centralized storage, and web base access to name a few. It would be the job of CPDM personnel to access what options are available and develop a strategy. It may be the case that SharePoint or Google Docs is the best fit for collaborating Word documents, however some document storage hierarchies are better left in file systems.

Whatever approaches are used, CPDM personnel will be tasked with using or developing application programming interfaces (APIs) so that web based applications can access the document stores. Handling document collections and establishing links to geospatial constructs will be one of the CPDM's most important functions.

## **Funding**

It is too early in the process to formulate how a CPDM would be funded because its framework and staffing has not been established. However, all of the program managers interviewed were receptive to providing some level of funding.

There is general consensus that some level of base funding needs to be in place to ensure the longevity of the center. It does not make sense that a long term data center be totally reliant on potentially sporadic programmatic funding.

It was suggested that directed S&T funding was a possibility early on.

## **Conclusion**

It is clear from interviewing a cross-section of agency program personnel that Reclamation should back the existence of some kind of unit to support mission needs. This unit could take the form of a CPDM as outlined in this paper.

The infrastructure would include servers that reside internally and also in the DOI cloud platform to facilitate external access. Appropriate database, application, and collaboration technologies would be put in place.

The CPDM would be staffed with professional level personnel in the fields of core IT, data management, geographic information systems and web mapping, database architecture and general web application design.

This would require a certain level of base funding for the maintenance of core operations, long term data stores and application lifecycle management. The programs would also contribute. If set up properly internal Reclamation personnel, external agency personnel, consultants and stakeholders would come into our CPDM to collaborate. This would create an environment that we centrally manage. The desired data management, lifecycle management, and stewardship controls would be implemented, which would ensure a sustainable system.

