United States Environmental Protection Agency

EPA/600/R 16/259 | September 2016 | www.epa.gov/research

EPA Scientific Knowledge Management Assessment and Needs

Office of Research and Development
National Health and Environmental Effects Laboratory/ Western Ecology Division

EPA/600/R-16/259
September 2016

EPA Scientific Knowledge Management Assessment and Needs

by

C. Richard Ziegler[1], Allen Brookes[1], Kerry Burch[1], Ann Vega[1], Andrew Yuen[2], Gerry Laniak[1], Ethan McMahon[2], Paul Harten[1], Bhagya Subramanian[1], Wendy Blake-Coleman[2]

[1]US EPA Office of Research and Development
[2]US EPA Office of Environmental Information

Project Officer:

Allen Brookes

Western Ecology Division
National Health and Environmental Effects Laboratory
Corvallis, OR, and 97333

Notice/Disclaimer Statement

The views expressed in this report are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

This document refers to supplementary material available on a Microsoft SharePoint site. Accessing this site requires login credentials. If you are interested in seeing some of this material and do not have credentials please send a request to COR_webmaster@epa.gov.

Abstract

A series of activities were initiated by a core group of EPA scientists from across the Agency starting in 2012 to increase the reuse and interoperability of science software at EPA. The need for increased reuse and interoperability is linked to the increased complexity of environmental assessments in the 21st century. This complexity is manifest in the form of problems that require integrated multi-disciplinary solutions. To enable the means to develop these solutions (i.e., science software systems) it is necessary to integrate software developed by disparate groups representing a variety of science domains. Thus, reuse and interoperability becomes critically important. This report briefly describes the chronology of activities conducted by the group of scientists to provide context for the primary purpose of this report, that is, to describe the proceedings and outcomes of the latest activity, a workshop entitled "**Workshop on Advancing US EPA integration of environmental and information sciences".**

Foreword

The U.S. Environmental Protection Agency (EPA) is charged by Congress with protecting the Nation's land, air, and water resources. Under a mandate of national environmental laws, the Agency strives to formulate and implement actions leading to a compatible balance between human activities and the ability of natural systems to support and nurture life. To meet this mandate, EPA's research program is providing data and technical support for solving environmental problems today and building a science knowledge base necessary to manage our ecological resources wisely, understand how pollutants affect our health, and prevent or reduce environmental risks in the future.

The National Health and Environmental Effects Research Laboratory (NHEERL) within the Office of Research and Development (ORD) conducts systems-based, effects research needed to achieve sustainable health and wellbeing. Research encompasses both human and ecosystem health, in that they are inextricably linked.

Table of Contents

Acronyms and Abbreviations

AA	Agency Administrator
AAAS	American Association for the Advancement of Science
ADC	Application Deployment Checklist
AGU	American Geophysical Union
API	Application Programming Interface
ATO	Authorization to Operate
BYOD	Bring Your Own Device
CIO	Chief Information Officer
CoP	Community of Practice
CRADA	Cooperative Research and Development Agreements
CREM	Council for Regulatory Environmental Modeling
CTO	Chief Technology Officer
DDES	Development and Deployment of Environmental Software
DQA	Data Quality Act
eCPIC	Electronic Capital Planning Investment Control
EDG	Environmental Data Gateway
EGU	European Geosciences Union
EMOD	Environmental Modeling
EPA	U.S. Environmental Protection Agency
ESIP	Earth Science Information Partners
FISMA	Federal Information Security Management Act
FITARA	Federal Information Technology Acquisition and Reform Act
FOIA	Freedom of Information Act
FTE	Full Time Employee
HHS	Health and Human Services
GIS	Geographic Information System
GSA	Government Services Administration
IEMMS	International Congress on Environmental Modelling and Software
IQA	Information Quality Act
IM	Information Management
IM PMO	Information Management Project Management Office
ISC	Information Services Steering Committee
IT	Information Technology
LCO	Lab, Center, Office
MARC	Mobile Access Review Committee
MOU	Memorandum of Understanding
NASA	National Aeronautics and Space Administration
NCC	National Computing Center
NHD	National Hydrography Dataset
NHHERL	National Health and Environmental Effects Research Laboratory
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
OAR	Office of Air and Radiation
OEI	Office of Environmental Information
OCFO	Office of Chief Financial Officer
OCSPP	Office of Chemical Safety and Pollution Prevention

OGC Office of General Council
OITA Office of International and Tribal Affairs
OMB Office of Management and Budget
OPM Office of Personnel Management
ORISE Oak Ridge Institute for Science and Education
ORD Office of Research and Development
OSIM Office of Science Information Management
OSTP Office of Science and Technology Policy
OW Office of Water
PARS Performance Appraisal and Recognition System
QA Quality Assurance
QAPP Quality Assurance Project Plan
RCS Reusable Components Services
READ Registry of Environmental Applications and Databases
REST Representational State Transfer
RMS Research Management System
ROI Return on Investment
RTP Research Triangle Park
SAS Statistical Analysis Software
SDM Scientific Data Management
SHARP SharePoint App Review Panel
SHC Sustainable and Healthy Communities
SIO Senior Information Officer
SLCM Software Life Cycle Management
SOAP Simple Object Access Protocol
SR&I Software Reuse and Interoperability
UN United Nations
UNEP United National Environment Programme
USGS United States Geological Survey
XML Extensible Markup Language

1. Introduction

Earth sciences and sustainable development organizations from around the world – including US Government agencies – have achieved various levels of success in taking advantage of the burgeoning Information Age. Concepts like participatory web, software interoperability, technology transfer, big data, scalability, cloud, re-use and open science are no longer "new and emerging." They have emerged and – in some cases – are tied to US Government directives. Efficiency, transparency, reduction of risk and ability to innovate – in the context of Earth sciences and sustainability – align with an organization's digital maturity and effectiveness.

This report reviews the history of activities related to an original Sustainable and Health Communities (SHC) Research Program task focused on the reuse and interoperability of science software and presents materials, outcomes, and recommendations resulting from the most recent workshop entitled "Advancing US EPA Integration of Environmental and Information Sciences Workshop". As the reader will notice the scope of and motivations for these activities has evolved to include consideration of the relationship between science software and information technology (IT) within the Agency.

For purposes herein, EPA's "digital science"[1] efforts shall broadly include data, software and applications, interfaces and dashboards, computational models, information repositories, decision support tools, APIs, ontologies and more – across platforms and devices (e.g., desktop, mobile, web, etc,). Further, the subject matter or scope of EPA's digital science shall broadly include environmental science, Earth sciences, sustainability science and human health science. As such, we explore the intersection and integration of environmental and information sciences and technology.

This report is organized as follows. Section 2 presents a brief chronology of activities related to the Community of Practice (CoP) for the development and deployment of environmental software (DDES). Section 3 provides a summary of materials used in connection with the October 2015 Workshop and Section 4 presents outcomes and recommendations resulting from the workshop. Additional material referenced in the report is provided in the Appendices. The interested reader can find a comprehensive set of documents, audio files, PowerPoint presentations, workshop materials, etc. related to the CoP for DDES at the following EPA DDES SharePoint Site. This site requires login credentials. If you are interested in seeing these materials and do not have credentials, please send a request to COR_webmaster@epa.gov.

2. Background (What led us to this point)

The workshop reported on herein represents the most recent activity in a sequence that dates back approximately four years. To accommodate readers who may not be familiar with previous activities we have included this background section that provides a brief summary of the most important motivations, activities, insights, and recommendations that explain what led us to this point. The reader who has followed the activities as they occurred may wish to skip this section.

[1] Notably, some of the authors disagree on semantics surrounding the phrase "digital science." Other potential phrases include "science software", "environmental software" and "cyber science." In this report we predominantly use digital science or science software, but we use each of these terms with the meaning defined in the text above.

Problem Definition, Context, and Solution Approach

The original technical focus for the sequence of activities was to increase the reusability and interoperability of science software. The reason for this focus was centered on the increasing complexity associated with environmental problems and the related need for multi-disciplinary integrated modeling research and assessment capabilities. Characteristics of the emerging class of problems (Brookes et al., 2014; Laniak, 2012) include:

- Complex interactions between human engineered and environmental systems
- Diverse stakeholder groups with conflicting and competing interests and values
- Geographic scales ranging from local, to regional, to continental, to global
- Temporal scales ranging from real-time to decadal and beyond
- Topics requiring integration of knowledge from multiple science domains
- No single “correct” or “optimal” solution

Formulating answers to these problems and rendering sustainable solutions requires:

- Holistic systems thinking
- Value-based negotiations among stakeholders
- Balance among social, environmental and economic impacts
- Dynamic decision-making that adapts as new knowledge surfaces
- Teams of people from multiple disciplines

Examples of such problems include:

- Natural disasters and human engineered failures
- Energy exploration and extraction
- Land use impacts on environmental quality
- Water, soil and air quality
- Climate change mitigation and adaptation

Science software is the principal means by which to express and apply knowledge of the environment to such problems. To build the requisite multidisciplinary modeling systems requires combining software developed for different science domains. Because science software is not intrinsically interoperable and is designed and implemented in myriad ways by the various developers the construction of integrated modeling systems requires that new criteria be considered in the design and implementation of individual software components (e.g., models, databases, tools). Moving the environmental science software development communities toward design and implementation strategies that enable reuse and interoperability was the original focus of the group coordinating this series of activities. As the reader will see this scope has necessarily expanded to address a number of community and organizational IT issues that cannot be separated from achieving the original goals of reuse and interoperability.

The following section (and referenced Appendices) provide a brief chronology of the activities that preceded the 2015 Workshop that is the main focus of this report.

A. **Chronology of Major Activities and Events**

1. 2012 EPA Workshop "Interoperability, synergy, modernization of ORD mechanisms"

The first workshop brought together EPA ORD scientists who explored the concept of interoperability and shared experiences related to building and applying modeling systems requiring the interoperation of multiple components.

Discussions centered on how to improve our in-house skills and mechanisms for achieving interoperability, and related cost effectiveness, cutting edge solutions, and trans-disciplinary approaches.

A primary outcome of the workshop was the formalization of a research task within the Sustainable and Healthy Community (SHC) research program to further the ideas set forth at the workshop.

2. Report: "Environmental Software Reuse and Interoperability: An Initial Review and Recommendations for the Office of Research and Development"

The first product of the SHC research task (1.1.2.2 "Interoperability, Cross-ORD and Beyond") formally defined the terms reuse and interoperability, described results of an investigation of ORD research programs to shed light on the extent of software development and the potential for science software reuse and interoperability.

As a result of this effort it was clear that ORD's reuse and interoperability issues could not be addressed within the boundaries of a single research program. Because science software components that comprise integrated modeling systems required components from across the Agency (not only ORD research programs but also across EPA Regions and Program Offices), it was necessary to broaden the scope of the effort to be cross-Agency. Also, in ORD alone, the investment in science software development required millions of dollars per year and involved greater than 50% of all research tasks.

Appendix A provides an extended summary of the report and the full report can be found on the DDES SharePoint Site.

3. Coding Conversations

To begin to address the scope of community that needed to be engaged regarding the topic of science software reuse and interoperability a series of 'coding conversations' were held at locations across the Agency and included over 250 people representing ORD, Program Offices, and Regions. In addition, several outreach efforts involving key people and groups throughout the Agency were conducted. The objectives of these efforts were to 1) increase awareness and understanding regarding science software reuse and interoperability; and 2) to gather feedback from Agency stakeholders (i.e., those who develop, apply, fund, monitor, and manage science software) concerning their relevant perspectives and experiences. Community contributions to the conversations and outreach took the form of lessons learned, needs, suggestions, questions, statements of 'the way things are', descriptions of challenges, barriers to progress, and practices that enable or inhibit production of reusable and interoperable software.

The coding conversations clearly demonstrated that a singular focus on the software engineering aspects of reuse and interoperability was not sufficient. Resolving related issues that span the software lifecycle process and the concept of a science software enterprise were shown to be tightly coupled to reuse and interoperability. A framework was developed to capture this expanded scope and consisted of four overarching topics; the software lifecycle process, managing the software enterprise, community development, and an on-line community portal.

The following materials resulting from the conversations are available the DDES SharePoint site

- Background Materials for Coding Conversations & Webinars
- Notes and Presentation Files from December Coding Conversation Webinars
- Synthesis of Audio recordings of December 2014 Coding Conversation Webinars
- Audio-based PowerPoint presentation describing the concept of Interoperability

4. 2014 Workshop ("Development and Deployment of Environmental Software (DDES) Workshop") and subsequent report

During the time following the coding conversations and outreach the team leading the effort (recall this effort began as a research task and expanded to include participants from across the ORD and Agency) worked with the framework of overarching topics and initiated a series or webinars to introduce the framework to the EPA community and begin to move toward understanding and prioritizing actions items. This material formed the basis for the 2014 workshop that had the goal of determining how the community could catalyze an EPA-wide effort to advance the set of goals related to science software reuse and interoperability.

Insights gained from the workshop included 1) that solutions to the reuse and interoperability challenge would require both a top-down (organizational/management) and bottom-up (practitioners) approach, 2) that to achieve the goals would require significant investment, 3) that a number of relevant, albeit uncoordinated and inconsistent, standards, best practices, and support technologies existed at EPA, and 4) that the existing software development culture at EPA presented not so subtle barriers to progress.

The workshop also provided an organization and prioritization of action items to pursue per topical area of the previously developed framework.

All workshop materials can be found on the DDES SharePoint Site. The report that resulted from the workshop is summarized in Appendix B and the full report ("Developing and Deploying Environmental Software at EPA Phase II: Framing the Issues and Building Community can be found on the 2014 Workshop page.

5. Science Software Reuse and Interoperability Demonstration Project (2014 – 2015)

In an effort to demonstrate both the operational aspects of science software interoperability and provide a pathway for establishing community-wide standards and practices a group designed and executed a project. Notable aspects of the project included addressing the requirement that interoperable software must be designed and implemented for automated discovery, understanding, evaluation, access, and integration. To satisfy this requirement the demonstration project experimented with "packaging" a software component, that is, developing specific content that would accompany the publishing of a software component. The content of the package provides 1) a set of metadata files, in the form of an ontological framework, that describes the science embedded in the science software, 2) the standards-based software integration elements (API, code, I/O schema, and class diagrams, example datasets), and 3) means by which to access the software via the world wide web.

The results of the demonstration project can be found on the DDES SharePoint site.

3. 2015 Workshop Overview and Summary of Materials

This report synthesizes information from the latest activity, the 2015 workshop entitled "Advancing US EPA Integration of Environmental and Information Sciences". In this section we provide a summary of content that was prepared for and discussed during the workshop. Extended summary content is provided in the Appendices and the full set of content for the workshop can be found at the following DDES SharePoint Site. Section 4 describes the key insights and discussions, outcomes, and recommendations.

A. Purpose and Agenda

The focus of the workshop centered on the following goals:

- To increase the efficiency, effectiveness, transparency and innovation of US EPA's digital science-related efforts to protect the environment and human health,
- To foster the evolution and merger of environmental science and information technology for US EPA, in consideration of 21st century information culture and cyber advancements.

An important aspect of these goals is that the scope of the activities has further expanded, coupling solutions to the original reuse and interoperability challenge to virtually all things digital. This may seem like overreach, and for this particular team it may be, but it has become clear that to achieve the original goals of reuse and interoperability (which are to provide the best available science to inform Agency decisions) we must find a way to fully merge the conduct of science with the IT world, which facilitates the expression of that science in reusable and interoperable form.

Specific objectives of the workshop included:

- Understand and appreciate the respective policies, OMB directives, practices, needs and barriers of environmental and information scientists
- Understand what got us to this point – brief review of history of circumstances and events(above)
- Become aware of exemplary efforts and approaches that leverage modern practices to re-use, scale, transfer, design and build interoperable science software; demonstrations and/or prototypes of interoperability and/or circumnavigation of challenges
- Make interpersonal connections (among leaders, developers, scientists, etc.) to cross-pollinate scientific digital efforts, increasing reuse, interoperability, scaling and tech transfer

- Identify ideal future characteristics for EPA's digital science portfolio/enterprise and next steps for how to achieve those characteristics.

The workshop agenda is provided in Appendix C and the participants' package can be found DDES SharePoint site.

B. Supporting Materials

The following sections provide a summary of supporting materials used in the preparation and execution of the workshop. The reader should note that the information included is recognized as incomplete from the perspective of all relevant information that may exist. It was produced under constraints of time and access and intended to add substantive information to the workshop discussions and recommendations. We attempt to simply provide the information here without significant comment, reserving interpretation and application of the information to the final section of the report.

1. US Government calls to action

There have been a series of Federal actions, directives, executive orders, memos, and guidance documents, dating to the 1996 Information Technology (IT) Management Reform Act, that target aspects of IT relevant to science software issues reported herein. The US Government has also recognized the slow adaptation to these initiatives among Agencies – notably, across two political administrations. This workshop, in part, discusses the status and rate of adaptation at EPA. A complete list of those Federal actions and references to the slow adaptation rate among Agencies, reviewed in the context of the workshop are provided in Appendix D.

2. US EPA's digital science landscape

It is difficult to accurately understand EPA's digital science portfolio. This challenge was identified in the January 2012 ORD Interoperability workshop (see the DDES SharePoint site), more recently called out by EPA's Office of Inspector General in 2014 (US EPA Office of Inspector General, 2015) and became a focus of this workshop. At present, there is no mechanism – or combination of mechanisms – that enables a complete and accurate assessment of EPA's digital science software portfolio. To address the question to the extent possible co-authors of this report reviewed the OEI Reusable Components Services (RCS), a database that provides a central point of access to a broad range of IT resources, components and services used in various EPA and partner systems. Not all products included in the RCS are science-based, for example, administration tools and products are also included. We attempted to divide Agency digital products into science and non-science categories. Appendix E provides the results of this review, and is organized to display the number of products per digital science category (resource type as defined by RCS) along with a brief description and examples.

It is clear that EPA develops thousands of digital science products, including data repositories, mapping interfaces, mobile applications, etc., for both regulatory and non-regulatory purposes. From the review conducted it is clear that there is significant potential benefits of increasing the reusability and interoperability of both existing and future software products.

3. Digital Science Stakeholders

A stakeholder is someone with an interest or concern related to, in this case, science software. To address the wide scoping challenges related to science software reuse and interoperability it will be necessary to involve a similarly wide range of EPA personnel, from developers to managers to QA personnel to senior management. Appendix F provides two views of the stakeholder community relevant to science software at EPA.

4. Existing Protocols and Procedures for Cataloging and Managing Digital Products

Of interest in the context of the workshop goals is understanding current Agency protocols and procedures for managing the science software development, deployment, and archiving process. Although there is no unified Agency-wide approach to "managing the science software enterprise" there are several examples of protocols and procedures within individual offices to either manage the enterprise or, more commonly, track (for approval) the process by which products are developed and deployed. The Office of Water (OW) is the one organizational unit within the Agency found to have formal procedures for managing their collection of science software. Appendix G provides a summary of the OW procedures as well as other existing protocols and procedures for the following classes of science software:

- Mobile Applications
- Web Applications
- Non-Web Software
- SharePoint Applications

5. Digital Science Communities of Practice (CoPs)

Lave and Wenger (http://www.learning-theories.com/communities-of-practice-lave-and-wenger.html) describe CoPs as "groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly." Lave and Wenger go on to explain that CoPs require three elements:

- **Domain:** Members have a commitment/passion for a shared domain of interest
- **Community:** Members willingly engage with and help one another increase knowledge in the domain of interest
- **Practice:** Members are practitioners in the domain of interest

Digital science communities of practice at EPA arose on top of – or superimposed on – the existing US EPA organizational chart. They often begin as rogue efforts, sometimes failing and other times achieving various levels of formal integration within US EPA. CoPs have evolved as an attempt to facilitate communication and collaboration across the Agency around specific areas of expertise, needs and/or interests. Specific areas include, but are not limited to:

- Modeling
- Data analytics and visualization
- R
- GIS
- SAS

- Statistics
- Scientific Data Management
- Developing and Deploying Environmental Software
- Computational Toxicology
- Strategic Foresight
- Records Management
- Project Management

Appendix H provides summary descriptions of CoPs that were reviewed for or included in the workshop.

6. *Digital Maturity and Related Metrics*

We have seen that reuse and interoperability of science software is a project level issue that requires community level solutions, solutions that necessitate the development of and adherence to key standards and practices. It also requires a community level approach to 'managing the enterprise'. As mentioned previously there is no unified Agency-wide science software enterprise management at EPA. The role of enterprise management, in the context of software reuse and interoperability, is principally to guide and fund the development of policies, standards, and practices; provide tools to facilitate the software lifecycle process; and to organize repositories of discoverable and easily accessible metadata about the science and the actual code expressing the science.

It turns out that there is an entire discipline associated with software enterprise management. As shown in the figure below there are gradations, or maturity levels, associated with an organizations approach to enterprise management. At the low end of maturity (where EPA resides) enterprise management is essentially non-existent, products are developed by independent autonomous groups with little or no relationship with products already developed, in development, or planned for development. This status reflects a project level focus where designing and developing for reuse and interoperability is a low priority and constrained by budget and schedule. The products place within the enterprise and relationship with other products is not substantively considered. As an organization proceeds up the maturity scale the enterprise comes into view as a single organizational asset and managed accordingly. Collaboration and automation increase, shared decision making and accountability, central processes and automation tools surface, and standards and best practices are developed and applied.

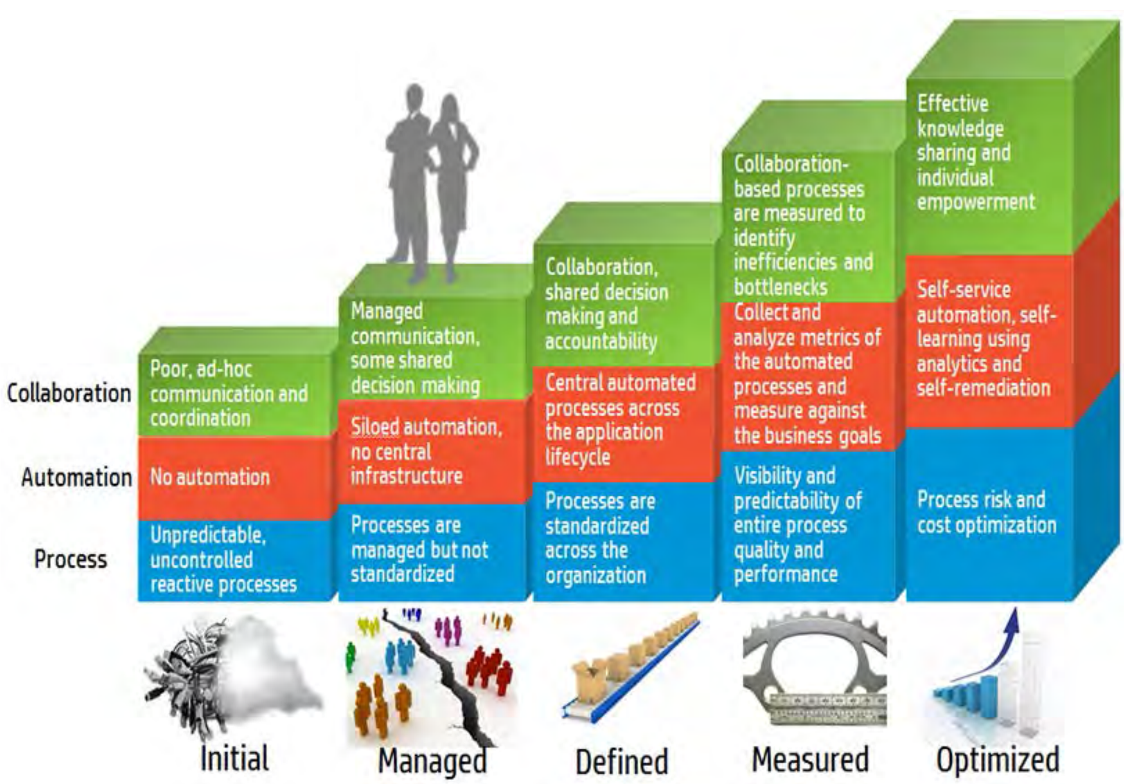


Figure 1. Gartner Digital Government Maturity Model

The importance of this idea of digital maturity in the context of our goal of increasing science software reuse and interoperability cannot be understated. Making progress toward higher levels of digital maturity will require a cultural transition at EPA. As is discussed in Section 4 of this report this topic is considered among the most important to the community of EPA stakeholders.

7. Practitioners Perspective

A practitioners' panel was convened at the workshop with the objective to advance community-wide awareness and knowledge of current and future protocols, practices and challenges associated with the software life cycle management (SLCM) process and to promote a cross-Agency approach for reuse and interoperability of science software products. For this panel, a "practitioner" is someone involved in the development, maintenance and/or deployment of science software products, including:

- software engineers (IT components/systems)
- software engineers (science components/systems)
- scientists who develop software (i.e., write code for components/systems)
- scientists who do not develop software but manage/fund its development

Each panel member was asked to prepare a 5 to 7 minute presentation focused on one or more of the following questions:

- What type of software are you associated with (e.g., tools to support science software, databases, models, etc.)?
- Briefly describe the process that is followed for software development you are associated with (e.g., waterfall, agile, ad hoc, contractor driven, etc.)?

- How do you address (if at all) reuse and interoperability in terms of preparing your software for R&I by those outside your development/user community?
- What is your experience with reusing and integrating software produced by others? What would you do to improve the experience?
- What are the challenges facing practitioners with respect to science software development in general and the goal of increasing reuse and interoperability in particular.
- What are practitioners doing to address these challenges (e.g., strategies to avoid/adapt to/work around/etc)?
- What is the role of contract support and how should software development by contractors be managed to ensure reusability and interoperability by others?

Among the more interesting insights provided by the panel are:

- Web services is a game changer with respect to interoperability. Web services removes constraints related to programming languages, platforms, and operating systems and allows direct access to the functionality of software through an API.
- Open source development and deployment facilitates reuse and interoperability in the following ways:
    - by providing no cost access to the best tools to support development
    - by enabling collaboration when developing in a public version control system such as GitHub
    - by providing well documented packages to help understanding and reuse
- Contractors need to be clearly informed of and held accountable for the IT standards they should follow.

A complete summary of insights from the panel is provided in Appendix I.

8. Related Efforts at Other Organizations

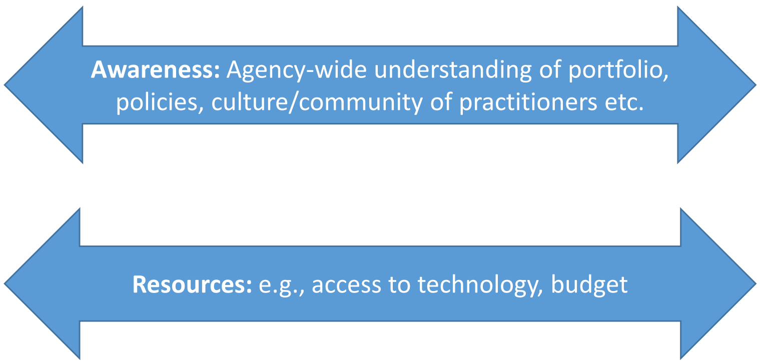
EPA is not alone in its need and desire to advance the integration of environmental and Information Sciences. Appendix J provides a brief description of several relevant efforts being pursued by organizations outside the EPA.

4. A View to the Future from the October 2015 Workshop

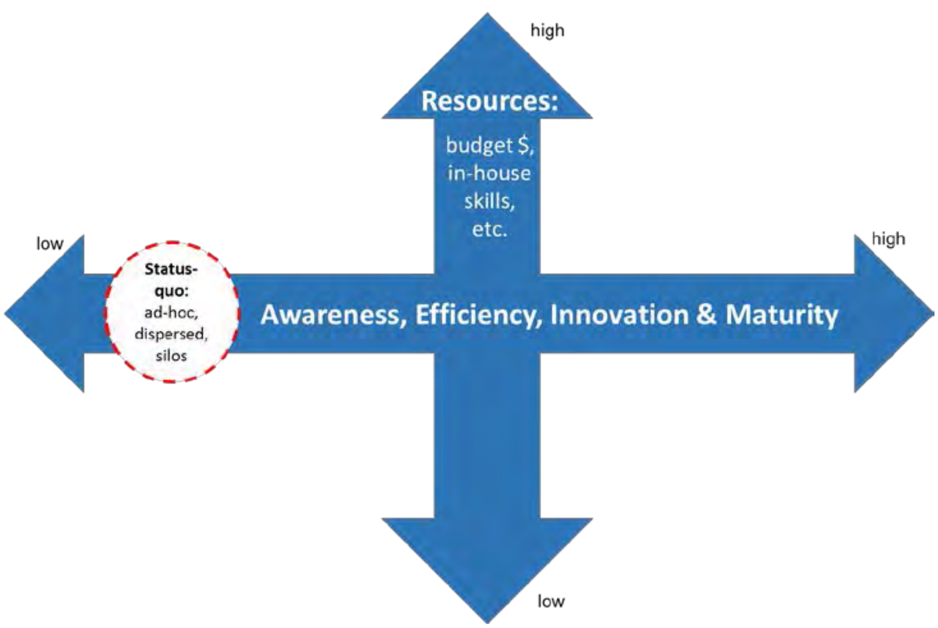
With the workshop materials and related discussions as background (see Section 3) the highlight of the workshop was engaging the participants in an open discussion of the future of digital science (aka science software) at EPA. This was done in two steps. First, a series of hypothetical future scenarios were presented and discussed in breakout groups. Each breakout group was asked to formulate important characteristics associated with a hypothetical future scenario. Secondly, the premise was established that regardless of the actual scenario that may unfold at EPA over time the characteristics identified in the breakout sessions would be fundamental to its success. Given this premise the collection of characteristics from across the breakout sessions (i.e., the discussions of hypothetical future scenarios) were organized together and presented to the participants during a plenary session. The participants were then asked to select the most important five characteristics (from their individual perspective). The results of the voting were then compiled and discussed with all participants. The next two sections summarize the hypothetical future scenarios and the characteristics that resulted from there discussion. Section 5 then presents a series of investment opportunities related to the most important characteristics.

A. Hypothetical future scenarios

Twenty years from now, what might digital science efforts look like at the US EPA? In this segment of the workshop participants were asked to ***hypothetically*** consider how the US EPA's science software efforts would be different in the future as a function of two variables:



We loosely modeled the exercise off of Rockefeller's 2010 futures analysis: "Scenarios for the Future of Technology and International Development" (The Rockefeller Founding & The Global Business Network, 2010).



Workshop break out groups considered multiple future scenarios for EPA's digital science efforts, including:

- EPA digital science as a **start-up**: Digital science projects and teams are independent with few constraints on tools and process. Similar efforts can co-evolve and/or diverge. Silos are leveraged to support novel ideas.
- **Centralization** of all EPA digital science: Projects are coordinated by a central EPA organization that specifies tools and process, and provides resources.
- EPA focuses on **technology transfer**: EPA leverages tools, code, interfaces, dashboards, etc., developed by other organizations and institutions; EPA minimizes its own development.
- EPA as **science consultant**: EPA drops the role of science software development organization (we no longer do it in house, we no longer contract it out; partnerships are used for the software extension of EPA's mission; EPA sticks to science, data, etc., but no software development (no websites, no decision support tools, etc.)
- **DevOps** EPA: Collaboration across digital, environmental science, IT infrastructure and administration to create digital science; imagine environmental scientists sitting beside developers in house at all phases, where iterations happen over seconds, as opposed to months, new concepts can be quickly tested in house, workflows (code development, etc.) are open to the public.
- **Personalization** of digital science in the Agency
- **Data-centric** EPA: Agency focuses on data as the primary assets of the Agency, and fosters but does not focus on development; democratization of data and wider data literacy pushes powerful knowledge outside of once-heavily controlled decision making processes.

B. Ideal future characteristics

Each break out group developed hypothetical future characteristics for being successful in their particular scenario. Several characteristics were identified by multiple groups, despite underlying differences in scenarios. Workshop organizers combined and synthesized all characteristics into one master list of 19 characteristics, and workshop participants voted on what they thought were the most important for the future of EPA's digital science efforts (Table 1).

The voting resulted in three groups; high, medium, and low importance. It is interesting that the community recognized 'culture change' as the most important characteristic of a future digital science environment at EPA. Without a change in the way we individually and collectively think about science software the necessary technical changes will not occur. Also of high priority is a movement toward the formation of multidisciplinary teams to conduct science software related projects. This does not simply mean multiple science disciplines but more importantly experts in software development, security, enterprise infrastructure, acquisition, etc. This will require additional IT expertise to be brought into the Agency. Interestingly, only after the culture change and organizational adjustments are addressed does it become important to focus specific attention on the topics of reuse and interoperability along with the science software enterprise.

Of mid-level importance are several operational characteristics including increasing partnerships with external communities, tracking, exploring, and introducing innovative technologies, streamlining calls for information, facilitating open communication inside/outside Agency, and embracing the best new digital development models.

Table 1. Ideal future characteristics for EPA digital science efforts, identified and ranked by 2015 Workshop participants.

Characteristic	Votes
Culture change: increased accountability, being open to change, aligning/creating awards for those who can adapt, incentive/disincentives, retention salaries.	41
Integrated, cross-functional, multidisciplinary teams (Scientist, Developer, Security, Ops/Infrastructure, Acquisition, Budget, HR-hiring, Legal).	38
EPA's in-house skillset matches our digital science portfolio (Developers/Coder, Architects, Security, Managers experienced running software portfolios, Operations, Software Engineers)**.**	36
Re-use and interoperability of hardware and software within EPA/External Partners/Industry: No duplicate software, Use of open source code, Modular development, Iterative practices, Providing a development and production environment that are the same (code playground), Terminology harmonization/data tagging; increase clarification/use of terminology, Federated data across the Agency, EPA leverages the semantic Web and its technology.	36
Improved portfolio management Agency-wide: Optimization of budget resources forecasting. Able to sketch, prototype and justify resources quickly and easily, Awareness upfront of total cost ownership, Aligning mission and performance measures with funding.	30
Increased partnerships/more integration with external communities (national and global): Reliant upon other entities doing software development, Software sharing.	22
EPA tracks, investigates, tests and explores new and emerging innovative technologies: horizon scanning, identify opportunities toward innovation.	21
Streamlined Information Calls: When someone enters metadata and project data into a system, it spreads to all systems, Open and transparent registry systems (e.g., the Wikipedia model).	16
Facilitate open communication inside and outside of the Agency.	15
Proactively embrace best digital development models (e.g., Agile, DevOps, evolving development techniques, what's next? When DevOps is obsolete will we still be using it?)	12
There is a new EPA Digital Science Innovation Team.	9
Openly and transparently track the ROI and environmental impact of all digital science efforts.	7
Risk-based security/segmented management of data and workflows for zero- and no-security risk digital science efforts.	5
Open and transparent digital science maturity tracking: Models, metrics, etc. are used to quantify Agency progress toward higher levels of digital maturity. .	5

Centralized resources for IT Security and SLCM compliance (included staff and funding).	4
More digital science efforts are self-sustaining (i.e., profit making or self-maintaining digital tools)	4
Bring Your Own Device (BYOD) (e.g., provide a stipend for personal computer and maintenance; no more network unless the scientists really need it).	4
Transition plans for digital science leaders.	3
There is a new Federal Software Development Office.	2

5. Investment needs and mechanisms

This section describes US EPA gaps, investment needs and potential implementation mechanisms. Some entries below are synthesized from the workshop, with particular focus on desired future characteristics, breakout sessions and workshop follow up activities. We acknowledge that some parts of the Agency and/or individual staff may be working on some or part of the areas outlined below. However, these areas require cross-Agency buy-in, and they are not being broadly implemented to an optimal degree. Greater upfront and inclusive attention to these needs will ultimately:

- Save government resources,
- Improve EPA's positive impact on the environment and human health, and
- Bring EPA into alignment with US Government directives (see Appendix C).

The "***potential***" implementation mechanisms" shown under each section below are ***ideas***. There are certainly other ways of accomplishing the inherent goals of the specific investment needs or Agency gaps.

A. Culture change

Gap analysis

Culture change is listed first by no accident:

1. Culture change was identified and voted among 2015 workshop participants as the most important need for the Agency's digital science efforts,
2. Cultural considerations and barriers cross-cut many of the remaining investment opportunities identified further below,
3. EPA has witnessed "a decrease in information sharing across IT and science."[2]

Investment opportunities[3]

- EPA should conduct more focused, face-to-face meetings. Outcomes of the meetings should be regularly reported to the Agency Deputy Administrator or other senior managers to ensure visibility and commitment to these conversations.
- Information must be shared about how EPA can acquire Agile and DevOps talent (via the Agency Chief Technology Officer (CTO) or otherwise).

[2] Excerpt from October 2015 workshop, specifically from "Breaking Down Barriers" breakout session
[3] Section paraphrased from October 2015 workshop, specifically from "Culture Change" breakout session

- EPA's CoPs should be empowered to address Agency issues. Expectations of their work should be well documented, including whether executive sponsorship is needed and the role its members will play.
- EPA projects should be more transparent. The use of project portfolio tracking across the Agency may help, including business justifications and budget planning. This could also support the FITARA review process (see also below under Portfolio Management)
- EPA needs to better determine what we value and want to reward. This could include applying resources to those projects which improve data and code transparency rather than simply seeing how many articles were published or downloads of data made.
- EPA needs to utilize more conferences like the October 2015 workshop

Potential implementation mechanisms

Several of the above investment opportunities are integrated in more specific sections immediately below. However, the following mechanisms are specific to culture change:

- **Adapt and extend the DDES CoP as a cultural change instrument for the Agency's digital science activities:** Determine those investment opportunities that DDES can assist and formalize for the Agency. While DDES's origin and name may not currently align with the broader goals of the October 2015 workshop and the goals implied in this suggested implementation mechanism, DDES has proven to be capable of confronting and defining ongoing problems for the Agency's digital science efforts, including for example lack of critical skills across management and staff, inefficiencies, redundancies and the need for greater adherence to US government directives. If DDES were to evolve as an instrument of cultural change, EPA senior leadership would need to formally acknowledge this and assign Agency staff to run DDES as part of their official duties (e.g., performance elements and PARS).
- **Repeat the October 2015 workshop:** The October 2015 workshop could be repeated annually, as an agnostic (no official EPA office sponsor perhaps organized again with help from DDES), along with digital maturity assessments and report-out of workshop activities and findings to senior management and in OEI developer meetings.
- **Publish this report and related information openly:** The findings and outcomes of the October 2015 workshop should be refined, continue to be made publicly available (e.g., Ziegler, Brookes, et al., 2015) and reported to the Agency Deputy Administrator and other senior managers. This report is a start. Further dissemination of information can be coordinated with OSTP (based on their interest in the workshop and follow up documentation).
- **Reject silos and expand accountability across the Agency with dedicated staff (not solely from OEI and OSIM):** All offices, divisions and branches of EPA must be held equally accountable for tackling the investment needs and challenges reported herein. Digital technology has become a common thread and device for achieving EPA's mission and must be recognized as such. It is unacceptable for Agency staff and management to place all digital accountability in the hands of, for example, EPA/OEI and EPA/ORD/OSIM. While OEI and OSIM may have a lion's share, within ORD for example, every lab and center must dedicate multiple FTE and acquire technology skills necessary for implementing a better digital science future. A piecemeal approach will not work (e.g., assigning 10% FTE for a SharePoint or ScienceHub contact, having a single webmaster, allowing one or two staff to help with digital CoP facilitation – those are good things, but this sort of activity must be expanded considerably). Dedicated FTE (e.g., multiple staff from each ORD lab and center) would support digital science efforts on-the-

ground, increase re-use, scalability and interoperability, transfer knowledge on agile/DevOps[4] mechanisms across the Agency, ensure that all environmental science products are machine readable, coordinate with OEI and OSIM, make sure larger external collaborations and efforts are known to avoid duplicative efforts across federal agencies.

B. **Partnerships and external coordination**

Gap analysis

US EPA is not alone (see Appendix J). Several organizations (government agencies and NGOs) with similar missions are working toward similar ends (greater interoperability, more innovative use of online technology, etc.). US EPA must collaborate with these groups to achieve more – more efficiently and more effectively. Further, US EPA must be "at the table" to inform and influence the international technology community in developing and implementing standards for environmental sciences as they continue to evolve (e.g., ontological standards and harmonization). We must look to share technology, to achieve more solutions using less resources (Fig.2).

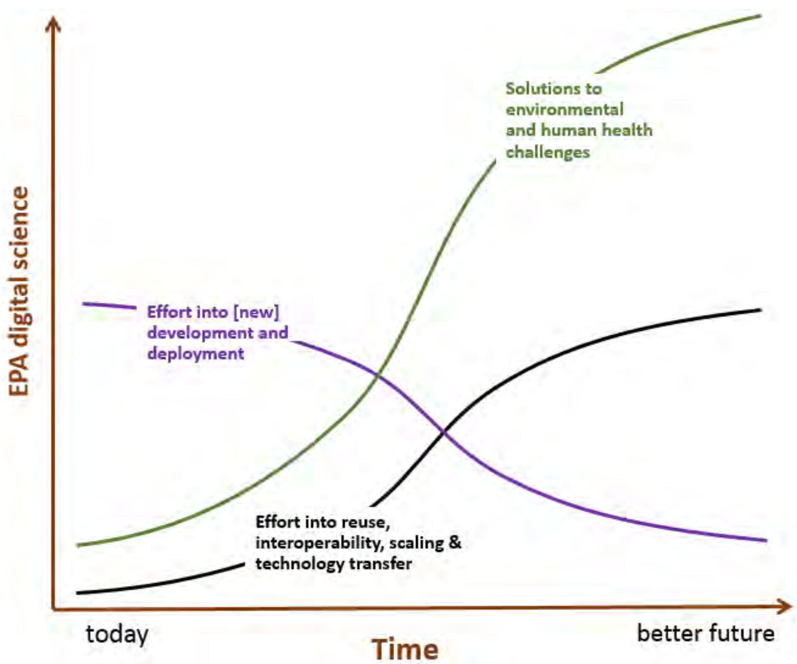


Figure 2. Improved digital future through greater re-use, interoperability, scaling and technology transfer

Investment opportunities

- Look for common goals, and set up collaborative arrangements and projects across organizations, both science-specific and specific to organization portfolios. For example, each federal agency does not need to develop its own portfolio tracking mechanisms; rather, they can

[4] Agile software development is a set of principles for developing software, including evolutionary development, adaptive planning, and continuous improvement. DevOps is a compound of development and operations that advocates for continuous communication between software developers and IT specialists.

share those tools and save resources. Further, those mechanisms can share information, benefitting the scientists, technologists and managers attempting to re-use and scale digital tools.

- Make sure EPA is represented at meetings and in groups with sister organizations and other external partners (immediately above).

Potential implementation mechanisms

- **Global Cyber Science of Technology Transfer Team:** Assign resources to the establishment of a global cyber team consisting of one or more FTE, minimally from each of EPA/ORD, EPA/OEI and EPA/OITA. This team would be dedicated to liaison with the national and international digital science community (See Appendix J), and then work within the Agency to ensure that EPA is not developing redundant digital science tools, but rather capitalizing on pooling resources across organizations working toward similar ends. The team would work closely with OITA to establish international working relationships with organizations that share common goals.
- **Dedicated Points of Contact:** EPA must establish and make known "points of contact" with external organizations that share similar environmental and sustainability goals, and organizations that are trying to improve interoperability, efficiency and innovation (e.g., ESIP, NSF EarthCube, Eye on Earth Alliance, OSTP, USAID Global Development Lab, Research Data Alliance, Belmont Forum, etc.). This must not be a territorial effort; namely, EPA staff member would not be limited in reaching out to external organizations if another EPA staff is already considered a point of contact. More importantly, EPA staff should at least know who among us is currently reaching out, and EPA managers (or the above "Global Cyber Science Team") must ensure that critical alliances are being nurtured.
- **Critical Meetings:** Maintain a central list of critical meetings that are focused at the nexus of science and IT (e.g., AGU, ESIP, EGU, IEMMS, etc.); EPA attendees must have vehicles (e.g., DDES webinars) to report out on meeting activity.
- **Collaborative Pilots:** EPA could set up digital science projects with interagency groups to pool resources, scale/re-use technologies and ensure interoperability of data and tools. The aforementioned "Global Cyber Science Team" might identify and shortlist several pilot opportunities within EPA and with external peers and collaborators.

C. Track maturity and impact via metrics (return on investment)

Gap analysis

We must track how well we leverage information technology to achieve our mission. In preparation for the 2015 workshop, the workshop organizers and sub-group of DDES conducted a pilot Gartner maturity assessment aimed at EPA's digital science activity (see Section 3.B.6). Gartner and other maturity metrics must be utilized on a regular basis by the Agency.

The success and return on investment of EPA's digital science efforts must also be tracked by attempting to associate such efforts with positive impact on the environment and human health. As described above, this is more difficult than conducting maturity assessments, but advances are being made external from EPA at the science and IT nexus.

Investment opportunities

- Identify, develop and routinely conduct maturity assessments:
    - Gartner should be replayed every year in the digital science context to track progress – The Gartner maturity assessment provides the basis for answering the question, "where are we now in our ability to develop and manage scientific software?" With the information from the maturity assessment described briefly above, the Gartner pilot provided a baseline that can be used to grow maturity in specific disciplines. Conducting more assessments using Gartner's structured and repeatable evaluation method would validate the pilot results, and enable management to focus improvement efforts where needed based on objective evidence.
    - Other maturity assessments should be explored and implemented on a regular basis in addition to Gartner assessments. For example, an open government maturity model (Lee & Kwak, 2012) and interoperability maturity model (Gottschalk, 2009) have been developed and would highlight specific areas for improvement and help track progress.
    - EPA would benefit by working with other federal agencies to develop and adapt a maturity model specific to digital science efforts in the federal government.
- Identify and develop impact metrics:
    - Tasks and projects must be developed with resources dedicated to tracking success (links to environmental outcomes)
    - EPA should explore and begin using altmetrics, ImpactStory.org, Depsy.org and similar (H. A. Piwowar, Becich, Bilofsky, & Crowley, 2008; H. Piwowar, 2013; e.g., Priem, Taraborelli, Groth, & Neylon, 2010; Priem, 2013)
    - EPA should strive to "connect the dots" starting from choice of research, all the way to measurable environmental improvements and outcomes; as it stands, we often are forced to associate success with outputs, such as publications, as opposed to outcomes (Fig. 3).

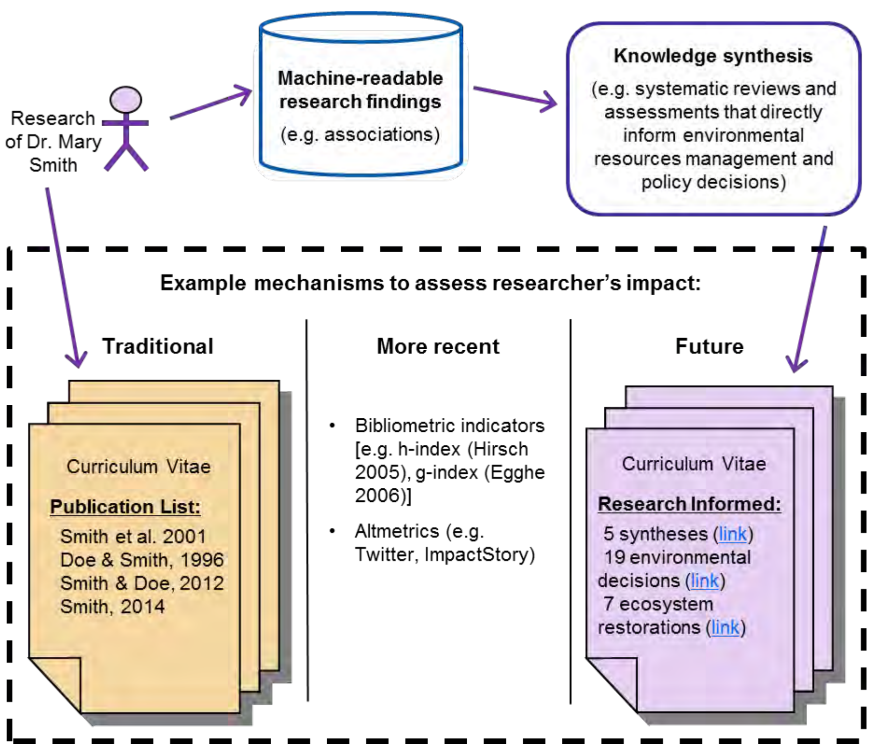


Figure 3. From measuring outputs to actual outcomes or improvements in environmental quality (source: Ziegler, Webb, Norton, Pullin, & Melcher, 2015)

Potential implementation mechanisms

- **Digital Science Metrics Strategy and Team:** EPA would dedicate cross-Agency FTE and/or a team of cross-Agency FTE to develop and implement a metrics strategy. Initially, a cross-Agency group could be pieced together as a combination of detail assignments, but this effort would be ongoing, and so dedicated permanent FTE would eventually be needed. The new "Metrics Team" would develop a digital science metrics plan for the Agency, including both maturity metrics and formal, objective attempts to quantify environmental impact. The team would collaborate across agencies to avoid duplication and ensure that the plan and tools can be transferred to federal agencies working on digital science efforts (EPA, NOAA, NASA, USGS, etc.). The plan may include, for example:
    - Conduct and repeat Gartner, open government and interoperability assessments
    - Work with other agencies and similar organizations to develop and conduct new maturity assessments if appropriate
    - Explore altmetrics technology (e.g., ImpactStory) to develop and promote EPA stories of data/tool use and value of data tools
    - Collaborate with one or more digital science asset teams (e.g., EnviroAtlas, Envision, ExpoBox, etc.) to pilot the use of altmetrics (e.g., ImpactStory, Depsy, Vivo, etc.)
    - Institutionalize use of Google Analytics and altmetrics for all public-facing data and tools.
    - Develop educational modules where appropriate, capturing educational use in analytics and altmetrics
    - Prototype the connection between EPA's task management systems (e.g., ORD's RMS) to metrics and trends showing environmental outcomes – in other words, demonstrate the connection between resource allocation and achieving the mission to protect human health and the environment.
    - Identify specific environmental topics and decisions for which success can be tracked from research choice to outcome (see Fig. 3), through the use of online tools; pilot and document one or more such efforts.

D. **Modernize portfolio management**

Gap analysis

A number of key business disciplines, associated processes, and dedicated staff with the appropriate skill set as well as senior leadership are needed to identify solutions to the difficulties of developing and managing scientific software in the EPA. Although there are many parallel improvement activities that can be occurring at the same time, implementing Application Portfolio Management practices and processes is an essential first step that is needed if the Agency is going to be able to take advantage of digital technology that enables scientific and environmental disciplines.

Investment opportunities:

In November 2015, the House Oversight and Government Reform FITARA Implementation Scorecard list EPA as having a grade of F in IT Portfolio – which includes all IT assets – infrastructure, major service contracts, and applications. Having a greater understanding of the inventory – that is to say a listing of applications that can be categorized as a "scientific application" would help improve FITARA implementation, reduce application redundancy, and provide a wealth of cost and technology information useful in the context of managing the larger portfolio.

Goals of application portfolio management include:

- Providing visibility and transparency by creating a single inventory of all existing and proposed scientific applications and data products
- Facilitating executive decision-making and investment review by providing reliable, easily analyzed data on cost, impact, value of current applications and proposed future scientific applications
- Prudent resource allocation using systematic analysis of applications from a holistic perspective. For example, being able to understand how much funding is going to specific types of scientific applications versus business administrative systems is important. And furthermore, what classes of scientific applications are getting the most Agency funding. This kind of analysis provides a holistic view – or the big picture (particularly in combination with impact metrics – see above)
- Identifying redundant applications and recommending decommissioning of applications. How to do we identify redundant applications? Are we doing competing application development work with other Agencies? If redundancies are identified, how do we reconcile strategies for moving forward or decommissioning?
- The strategic development and management of scientific applications with Application Portfolio Management activities supports the mission of the Agency. The benefits of application portfolio management are numerous by providing scientific applications that encourage more partnerships with business, non-profits, state and local governments, and tribes to share applications and information.
- Reduce IT spending, while accomplishing more.

Potential implementation mechanisms:

- **Independent audit and dedicated internal EPA team:** EPA would hire an external independent audit team to assess the current state of portfolio management. Then, with the findings, if deemed appropriate by the audit team, EPA would dedicate full time resources (staff) responsible for working with management to improve the scientific software portfolio, the management processes and project planning processes. The audit team and newly dedicated EPA staff would explore and adopt appropriate portfolio management software.

E. **“Secure Enough” for EPA science and research**

Gap analysis

EPA computer networks prioritize security over all other considerations. Security is a very important issue for computers and computer networks. Without sufficient security we cannot maintain confidentiality, integrity and accessibility. However, with too much security we compromise accessibility and scientific progress. Much of EPA’s work product is not confidential and benefits from being easily accessible by other researchers and the general public. Separating this work to networks that are more accessible will allow for easier access, better collaboration, and even allow truly confidential work to become more secure.

Investment opportunities:

- Create multiple networks that allow for varying security needs.
- Provide external access on least secure networks, allowing external collaboration opportunities.

Potential implementation mechanisms:

- **"Secure Enough" Pilot Program:** The below sub-bullets might be considered a proposal that could be tested by setting up pilot sites within the agency with a more open network. This could be as easy as providing a completely open network through a commercial service provider. Existing machines could be used, including ORD owned machines and machines from before the current refresh, with admin rights given to the users. In addition, servers could be set up outside the firewall with password protection for testing server applications, web services, and databases. Alternative network environments already exist within the agency and could be used or enhanced to provide pilot sites. Evaluation of the pilots could then determine the feasibility of agency wide adoption.
    - *Alternate EPA computers:* EPA would dedicate science-specific computers that attach to a more open network and that have greatly reduced security needs, so the default would be that users of the computers will have full Admin rights. These computers will have virus detection software and may have EPA authentication, but users would be able to install the software they need and configure the computer to run in ways that are consistent with the needs of their work. The computers will not have a standard image and will not receive pushed software updates. Users would be informed of OS and software security updates and be given reasonable time frames for installing updates. The machines will not have standard configurations, and refreshes should be as needed rather than Agency- or office-wide. Users needing higher performance machines should be able to have higher performing computers and refresh more often than those requiring less performance.
    - *Personal computers:* Personal computers would be acceptable for use on the open network. Stipends would be provided for purchasing and maintaining personal computers to use for most agency computing needs. Since most EPA employees already maintain one or more personal computers, this may be advantageous for both employees and save Agency resources.
    - *EPA Servers:* Various types of EPA servers should be accessible through the aforementioned more open network. We currently provide public access to a variety of EPA data, including the Geo Platform and READ. But we don't have easy ways to provide access to data, databases, web services, or server applications that haven't been vetted or completed. We need to be able to collaborate with people outside the agency and having more open servers is a big part of this.
    - *Access to Production Network:* Computers on the open network cannot be connected to the EPA production network, but there needs to be a mechanism for sharing data and documents between the networks. One possible mechanism would be to provide a site that is write only externally and read only internally. Items copied to this site would be scanned for viruses and then deleted within 24 hours. Within the network scanned items could then be copied to other shared drives.
    - *E-Mail:* One of the most important items within the production network that would need to be accessed externally is email. Although there is currently a mechanism for accessing email outside the network, it is awkward and there are varying interpretations among EPA managers on how and if this e-mail access should be used. One must authenticate

multiple times and deal with frequent time outs. It would be much easier to be able to set up Outlook on one's computer and be able to access email any time without additional authentication. This is much less secure than the current system, but most emails do not need the level of security that they are currently given. Instead of trying to secure all emails it would be better to secure only those needing additional security. One way to do this might be to identify secure emails as they are created and not deliver those when reading outside the secure environment. Instead, a notice could be sent indicting the existence of a secure email. The user could then use the more secure access route to read the email.

F. In house skills, capacity building and education

Gap analysis

We need in house skills that reflect the efforts being put toward the intersection of environmental and information sciences. We also need to cultivate skills among existing staff. One of the workshop breakout groups ("Barriers for change") specifically noted, "There are unclear pathways for advancement, including prevention to go on detail to learn new skillsets… Many organizations have difficulty finding and hiring the staff they need, including IT professionals."

Investment opportunities

- EPA should leverage hiring mechanisms beyond USAJOBS to improve the hiring process, including the use of Schedule A, Pathway and veteran hires.
- Organizations should prioritize the skillsets they need and conduct better succession planning to ensure expansion of skills.
- Details should be encouraged to help build new skillsets and share skills across organizations (internal and external of EPA).
- Hire people with the skills of helping Agency scientists meet their software needs. This includes acting as an interface between the scientist and the software developers.
- OEI needs to work with Agency scientists to determine how Student Services Contractors, ORISE Fellows, AAAS Fellows, etc. will work with new fellowships (e.g., Presidential) and agile/integrated teams
- Security policy - Improvements could be made to the Agency's Role-based Training Program to ensure managers and staff not only understand their role in the process, but get the right information they need to know and questions to ask to consistently apply and implement the policy
- Software Managers should be required (see also Culture Change above) and distributed just as QA Managers and Communication Managers are required and distributed within Organizations.

Potential implementation mechanisms

- **Incentivize digital skills building at the AA level:** Senior leadership must establish tangible rewards for staff and managers that participate in and allow for example continued learning (e.g., PhD programs, internal details), respectively. Encouraging this is not enough. Offices and divisions throughout the Agency may fear FTE reductions or workforce reductions by sending staff outside of their offices. Rather, long term Agency-wide benefits need to usurp those fears.

Mandating staff exchanges within EPA organizations and with those organizations with whom EPA signs MOUs may remedy this. The exchanges must somehow reward (e.g., cash, additional leave, etc.) those who “compete” to be considered, along with their supervisors.

- **Details to external organizations:** This must also be encouraged, since many of the technologies being worked on by EPA are similar to those being developed elsewhere.
- **Technology staffing, brainstorming and knowledge sharing workshop**: It would benefit EPA to host an internal and/or external federal government workshop on how to acquire and retain different types of IT skills. Many times, managers use the mechanisms and existing job descriptions they already have, forcing staff scientists to outsource IT tasks. Knowledge sharing across the Agency and other federal agencies, specific to IT skills for science, may illuminate alternate staffing mechanisms.
- **Rethink and track mechanistic approaches:** Alongside portfolio management, EPA would track how it develops digital science and catalogue methods and best practices. For example, we could be using more code-a-thons, hackathons, CRADAs, and similar less traditional mechanisms (in addition to internal staff and contractors).

G. **Emerging areas, horizon scanning, digital innovation**

Gap analysis:

The Agency must catalyze digital innovation to more effectively and efficiently achieve EPA’s mission. It’s important to leverage emerged technologies (e.g., crowdsourcing, social media, citizen science, etc.), but we must be positioned to leverage those technologies on the horizon, including those coming into view and those not yet imagined. This means setting up cultural and institutional mechanisms capable of experimenting and implementing rapidly. Several groups within the Agency are working at the edge of this gap, including OCFO’s Strategic Foresight Community of Practice, ORD’s Innovation Team, RTP’s Cutting Edge seminar, and OEI’s internet of things/citizen science effort. It will be important to collaborate with these groups and others to fill this gap, and more formally recognize horizon scanning and innovation specific to digital science.

Investment opportunities:

- Add a cross-Agency innovation element to managers’ PARS.
- Develop a process for innovation to make it effective and efficient. This should include the ability to present the idea to management of a cross-Agency group (including senior leaders or former leaders) for feedback. If the feedback is positive, then a brief proposal about the idea (including an alternatives analysis) would go through an approval process based on its costs (low/no-cost activity goes to a line manager for approval; medium/higher cost activity would require higher levels of approval). Results would need to be presented regardless if they are positive or negative.
- Collaborate with similar groups within and outside the Agency that are doing similar things
- Build list of emerging IT focal areas and track progress, leaders, skills, etc. online -- maybe just for EPA internally or for federal agencies with OSTP
    - For example, gaming has shown tremendous untapped potential for the environment (Costanza et al., 2014; Ziegler et al., 2013) and was mentioned more than once at the 2015 workshop as an area needing more attention at EPA. Notably, computer gaming

efforts to date among federal agencies have not necessarily yielded good return on investment; they often focus on "serious game" development approaches which have serious disadvantages (Azadegan et al., 2013). A group of EPA staff have begun to brainstorm this area and potential ideas, but more dedicated attention is needed to help EPA staff avoid pitfalls encountered in the past.

Potential implementation mechanisms:

- **Digital Innovation Team:** This new team would have dedicated staff from across the Agency and work closely with OCFO's Strategic Foresight Community of Practice, ORD's Innovation Team and OEI's internet of things/citizen science effort. As innovation must be ubiquitous and systemic within the Agency, some positions on the team might be term-based (1 to 2 years); others would be permanent. The team would:
  - Maintain and curate a living "horizon" list of digital science ideas (champions, interested leaders, external collaborators, potential sources of resources, etc.)
  - Seek implementation methods and mechanisms for assigning innovation champions or new staff to permanently work on innovative ideas.
  - Help new efforts (such as the gaming workgroup mentioned above) get off the ground and test potential directions using best practices
  - Establish new funding mechanisms and/or profit generating mechanisms for environmental digital science innovation

H. **Open information and workflows**

Gap analysis

EPA information flows often adhere to greater security rather than openness. One of the 2015 workshop side conversations focused on how things might change if IT projects were developed in full, open view of the public. While focused on digital science, much of the 2015 workshop and many recent digitally-focused US government directives (see Appendix C) are centered conceptually around openness. The struggle to modernize technology within EPA specific to science repeatedly confronts the notion of openness in the historical shadow and burden of liability.

Investment opportunities:

Imagine, for example, not just using open source code and code developed outside of EPA in real-time, as it's developed, but also providing real-time updates on interface development, removal of bugs, new features, etc. in the creation of scientific decision support tools. A development team would broadcast their progress, and welcome interactive feedback and code edits from anyone. Such a practice is not just about development of knowledge management tools, but also the dissemination of information about the tools. The public would be able to see the process. Such a conceptual approach to proactive information flow applies to other EPA functions, also in the digital science context. For example, EPA's use of social media, eRegulation effort and FOIAonline effort often provide retro-active dissemination of information. In contrast, subscribing to the new mantra of "open" as the new default, information dissemination would become proactive, and those tools are poised to open new doors for public participation and collaboration. This is made possible with online technology.

Potential implementation mechanisms:

- **Digital Open Government Team:** EPA would dedicate staff to an open government team that focuses attention on digital science efforts. Among other tasks the team would:
    - Assess EPA's level of compliance on relevant US government protocols, specific to digital science, perhaps in conjunction with conducting an open government maturity assessment (Lee & Kwak, 2012; also see above)
    - Identify and prototype digital tools for making ***all*** EPA workflows (code development, impact metrics tracking, decision making) more transparent, collaborative and open. For example, EPA could analyze past FOIA requests that could have been avoided by publishing information pro-actively; this type of exercise would inform which EPA activities and workflows could be made open immediately to save resources.
    - Look for mechanisms to change the publication paradigm of workflows, creating meaningful public involvement (R. Farina, Epstein, B. Heidt, & J. Newhart, 2013) earlier in every process
    - Compare the investment in cyber infrastructure and personnel for pro-actively releasing information as it's used, versus supporting resources for keeping information secure (see "Secure Enough" pilot above) and administering retroactive information dissemination (e.g., FOIA responses, including legal staff and expenses).

Develop a cross-agency workgroup to confront challenges and share technologies associated with open government mandates, specific to shared workflows and processes in the science realm (similar to the eRegulation direction).

In conclusion, as indicated in the first part of this section, the above ideas were generated as an outcome of the workshop and should be considered by the Agency if it is serious about achieving a "merger of the sciences" (information, digital, decision, environmental, sustainability, human health, etc.). While there are several possible approaches described, perhaps the most important is that these ideas must be embraced by a group of accountable individuals, a group of champions in authority who can affect change. Without these champions present throughout the Agency, as we have learned over the last four years, we will be unable to achieve the vision presented by Ann Dunkin (EPA CIO) as expressed in her keynote address at the workshop.

> *"While we can maintain the EPA's strong culture of autonomy, it is time to give up this particular practice of starting every system from scratch. The tools we build are better as a system than as individual tools. The whole will be greater than the sum of its parts and it will allow us to attack deeply complex multidisciplinary problems that we cannot attack if our tools will not work together. If we get a handle on our scientific applications and models across the EPA, not only can we share data and tools within the EPA, but we can share data and tools across the federal government, with citizen scientists and with the private sector."*

Appendix A: Summary of Phase I Report

Laniak, G. (2012). *Environmental Software Reuse and Interoperability : An Initial Review and Recommendations for the Office of Research and Development*. Retrieved from [DDES SharePoint site]

Phase I Report Summary

[See Full Report on DDES SharePoint site]

Key Activities and Accomplishments

- Defined terms and benefits of reuse/interoperability in context of science software along with attributes of R&I software
- Reviewed ORD research plans and quantified extent and cost of science software development
- Described in practice with current environmental software systems (data systems, modeling systems, and decision support systems)
- Described a framework and standards for metadata
- Described levels of interoperability and existing relevant standards

Key Findings

- **The volume of environmental software produced at EPA represents a significant investment.** There are more than 400 environmental software products existing within ORD alone and more than 200 ongoing research tasks that involve additional software development. On an annual basis, this translates into tens of millions of dollars in combined FTE and extramural funding.
- **EPA environmental software includes a wide and diverse range of science-based components** (i.e., individual or standalone databases, models, and tools) along with an array of integrated systems for data management, modeling, and decision/policy support. The software is deployed on several platforms and is developed with nearly limitless combinations of operating systems, computer languages and coding practices.
- **Software reuse and interoperability is a project-level issue that requires community-**level solutions. Individual projects need to combine software components developed by disparate groups into coherent and quality assured system solutions.

Recommendations

- **Establish an enterprise level focus and role for environmental software reuse and interoperability at ORD**.
- **Create and empower a software reuse and interoperability (SR&I) group**
  (composed of scientists, software developers, users, and managers) with the charge to:
    - increase awareness and understanding of the concepts and need for software reuse and interoperability
    - represent community issues and needs to the organization establish a formal collaboration among ORD, OEI, OSIM, and CREM
    - partner with other national and international organizations
- **Conduct community-based activities directed at increasing software reuse and interoperability**
    - review existing Agency and OMB IT guidelines and requirements to determine relevance and applicability to scientific software reuse and interoperability
    - develop and publish consensus best practices and standards for reuse and interoperability
    - populate repositories of reusable environmental software components and systems
    - develop tools for discovery, access, and integration of components
    - develop demonstration projects to explore issues and challenges related to reuse and interoperability (e.g., how to refactor legacy software for reuse)

Appendix B: Summary of Phase II Report

Phase II Report Summary

(See Full Report on DDES SharePoint site)

Key Activities and Accomplishments

- **Established Agency-wide dialogue regarding the need for, benefits of, and issues related to increasing the reusability and interoperability of environmental software.** This was achieved via a year-long series of interactions that included 7 open discussions (referred to as "coding conversations"), 13 outreaches to specific individuals and groups conducting DDES related activities, 4 webinars and follow-on discussions to present and discuss DDES action areas, and a workshop to build consensus on a path forward.
- **Identified DDES stakeholder community and associated roles and responsibilities.** The community includes, of course, software developers, but also individuals who fund and manage software development, establish software-based policies and QA guidelines, provide access to and support for tools and technologies that facilitate software development and deployment, and those that use the software to conduct research and regulatory-based assessments in support of Agency decisions and policies.
- **Assimilated community discussions and developed a DDES framework to foster awareness** and provide a holistic understanding of and context for DDES issues and future efforts**.** The framework includes four overarching topics as shown below.
- **Using this framework, identified 18 action areas,** i.e., sub-topics for which specific activities are to be identified and executed. Formed community consensus that the following 7 action areas are of highest priority in the near term (3 – 18 months).

Key Findings

- **Focus on DDES community building.** Why should we focus on community-building rather than on nuts and bolts of software standards for developing reusable and interoperable software? The short answer is that everything about increasing reuse and interoperability requires community-level coordination, cooperation, and collaboration. Benefits from participating in the larger community will be maximized only if we develop a community-based view of EPA's environmental software enterprise.
- **DDES solutions require top-down and bottom-up approaches.** From the bottom to achieve the goals of the DDES initiative will require that the community of move toward standards and practices that facilitate the design and construction of reusable and interoperable software. Equally important, and grossly underappreciated, is the requirement that environmental software be viewed as an organizational asset and managed with formal software enterprise business principles and practices.
- **Interoperability requires investment.** Several aspects of committing to the goals of DDES require new resources. Initially, it costs more to design and build reusable software. The primary economic benefit of preparing software for reuse and interoperability occurs when developing new software is avoided by reusing existing software. An underappreciated fact is that an inventory of reusable software must be available before the benefits can occur.

Appendix C: 2015 Workshop Agenda

Participant's Packet
for the

Advancing US EPA Integration of Environmental and Information Sciences Workshop

(complete package can be found here)

US EPA Campus
Research Triangle Park, NC

WORKSHOP GOALS

1. Increase efficiency, effectiveness, transparency and innovation of US EPA's digital science-related efforts to protect the environment and human health
2. Foster the evolution and merger of environmental science and information technology for US EPA, in consideration of 21st century information culture and cyber advancements

WORKSHOP OBJECTIVES

1. Understand and appreciate the respective policies, OMB directives, practices, needs and barriers of environmental and information scientists

2. Understand what got us to this point – brief history of circumstances and events:
   a. Overview of challenges identified over the last 3 years, including drivers and reasons for change (thru DDES and related activities; see above)
   b. Overview of US EPA's environmental science software portfolio – that is, digital projects and products, including data, interfaces, dashboards, models, repositories and decision support tools
   c. Overview of stakeholder roles and responsibilities

3. Become aware of exemplary efforts and approaches that leverage modern practices to re-use, scale, transfer, design and build interoperable science software; demonstrations and/or prototypes of interoperability and/or circumnavigation of challenges

4. Make interpersonal connections (among leaders, developers, scientists, etc.) to cross-pollinate scientific digital efforts, increasing reuse, interoperability, scaling and tech transfer

5. Identify ideal future characteristics for EPA's digital science portfolio/enterprise and next steps for how to achieve those characteristics.

Agenda

Day	Time	Event	Location; Potential Lead(s)
DAY 1: Mon, Oct 19 (PM only)	12 – 1p	Registration	Atrium B
	1 – 1:30p	**What do we hope to accomplish in this workshop? Drivers for change!!** Logistics, ground rules, goals, agenda - Brief timeline of DDES and interoperability efforts, leading up to this workshop - Issues, inefficiencies and barriers emerging from Coding Conversations, DDES webinars, August 2014 virtual workshop, ESIP 2015 Winter session and related activities (ontologies, security, staffing needs, EPA dichotomy between IT and science, contractors, awareness, conflicts of interest, etc)	Auditorium, plenary; Elstein and Ziegler
	1:30 – 2:15p	**Keynotes (Ann Dunkin and Tom Sinks)** - Vision for EPA's digital science efforts - Speakers might consider big picture changes and how things might look in 2025 or beyond, as a function of changing environmental challenges, quickly evolving digital landscape, and emerging / horizon technologies	Auditorium, plenary; 10 min each for presentation, plus 5 min Q&A
	2:15 – 2:45p	**Structures and roles summary (presentation)** - Overview of software development at EPA; includes OEI and two example program offices (OW and ORD) with the following roles highlighted:	Auditorium, plenary; OEI, OW, and ORD tag team (Yuen, Allen

		- Developer/Application Owner - Support personnel - Approvers/Management - EPA organization chart showing where different roles occur across OEI, OW, and ORD - Jumpstart discussions for the rest of the workshop. Do other program offices/regions do things differently? Are there alternatives to how we do things now? What roles are missing? What would make software development more efficient and effective?	& Vega)
	2:45 – 3:15p	break	
	3:15 – 4:30p	**Practitioners Panel – Current Practices and Future Directions** - See below for details	Auditorium, plenary panel; led by Laniak and Brookes
	4:30 – 4:45p	**Plenary close out; homework; What's next?**	Auditorium, plenary; Ziegler
	4:45 – 6p	**Science software portfolio (poster and demo session)** - See below for more information - The posters will remain standing for the workshop, supporting a portfolio synergy exercise (where we will seek specific case studies, demonstrations and investment opportunities)	Atrium B and break out rooms; organizers: Yuen
DAY 2: Tues, Oct 20	9 – 9:15a	Opening: re-cap day 1, and plan for day 2	Auditorium, plenary; Elstein
	9:15 – 10a	**"Superimposed" communities of practice (and similar groups) panel** - Panel members coming from "ad hoc" groups superimposed on EPA's structure : GIS groups, EMOD, R user group, DDES, scientific data management, DataViz, SAS User Group, Web Council, CoP for statistics, etc.) - How do these groups fit with org chart? Who runs them? Why do they exist? How many are there? Commonalities? Potential collaborations? Synergies? - See below for more information	Auditorium, plenary; Vega and Subramanian; cross-Agency panel consisting of digital CoP leaders and/or prominent participants of CoPs
	10 – 10:20p	**Science software portfolio summary (presentation)** - Summary of all Agency science related digital efforts (data, dashboards, decision support tools, websites, models, ontologies, repositories, etc.) - Scope (subject areas, end-users, platforms, etc) - List/table of projects (active, sunset, maintained, other) - Budget estimates (extramural and FTE) for all past, active and planned science software efforts under development, retired and/or in use	Auditorium, plenary; OEI (John Harman)
	10:20-10:40	Break	
	10:40 – 11:40a	**IT *for* Science Panel** In an ever changing technology landscape, this panel will offer critical perspective on modernizing key disciplines and innovative practices, and skills sets that cut across traditional functional and programmatic lines. Check your perceptions at the door and come learn about architecting scientific applications and data that deliver secure, open and transparent data, applications, and service to the American people. Come learn from this diverse group of experts about the practices, skills, and management support that are needed to create an information-centric model that cuts across in EPA where scientific applications and data are architected security, interoperability and openness.	Auditorium, plenary; led by Kerry Burch
	11:40 – 12	**Digital Science Maturity** Maturity assessments are valuable tools that provide specific guidance on current and future actions that will guide Agency IT reform. Actions in multiple disciplines that contribute to scientific application development and management will reinvent not only what applications we develop, but how we do it.	Auditorium, plenary; Kerry Burch
	12:00 – 1:30	Lunch	
	1:30 – 2:00p	**Panel synthesis and open discussion** - Review, synthesize and discuss major topics from each of the preceding panels	Auditorium, plenary; led by facilitators from 3 preceding

			panels
	2:00 – 3:20p	**Future: scenario planning exercise (break outs)** - N.B. this feeds the "Future characteristics panel" and "Let's go there together" break out session and report out - See separate instructions (below) for more information on the "Future" sessions - Workshop participants will be required to vote on ideal future characteristics after scenario flip charts are finalized and before "Future: characteristics panel"	Auditorium, plenary instructions, then to break out rooms; Ziegler facilitates
	3:20 – 3:40p	Break	
	3:40 – 5p	**Science software portfolio continued (fire talks session)**	Auditorium, plenary; organizers: Yuen; cross-Agency presentations
	5 – 5:30p	End of day assessment	Auditorium, Elstein
	5:30 – 9p	Dinner / social (optional, TBD)	
DAY 3: Wed, Oct 21	8:45 – 9a	Opening: re-cap day 2, and plan for day 3 (vote on future characteristics)	Auditorium, plenary; Elstein
	9 – 10a	**Emergent solutions - presentations and panel** - Perspectives from 3 to 4 other federal agencies or from within EPA (e.g., 18F, OW business model, other) on how they've dealt with modern digital science challenges (3 to 4 short plenary talks followed by questions and open discussion with speakers; approx. 10 min each, delivered via web conference and/or video)	Auditorium, plenary (Vince Allen (OW), and others via teleconference setup)
	10 – 10:30a	Break (be sure to vote on future characteristics, if you haven't already)	
	10:30 – noon	**Future: Leaders/Managers future characteristics panel** - Before the panel discussion, panel members will have the ***option*** to deliver a 5 min presentation on existing challenges (e.g., accountability, implementation, etc.) and their vision for the future - The panel will then discuss characteristics (physically positioned on stage behind panel members) that result from the "Future: scenario planning exercise," ranked by workshop participants before this panel - N.B. This panel will feed the "Let's go there together" break out session and report out - See detailed instructions below for more information	Auditorium, plenary; panel members: Auditorium, plenary; Ann Dunkin, Greg Godbout, Tom Scheitlin (OEI), Phil Dickerson and Ron Evans (OAR), Jerry Blancato and Mike Slimak (ORD), Oscar Morales (OCSPP)
	12:15 – 1:30p	lunch	
	1:30 – 3:15p	**Future: Let's go there together! Actions and plans (break outs)** - break out sessions will be led by the preceding session panel participants; additional break out sessions will be formed as needed, depending on workshop attendance - N.B. the "Future: scenario planning exercise" and "Future: characteristics panel" will feed this session with specific information to be transformed into action items, case studies and blueprints for the future - See instructions below for more information on the "Future" sessions	**Auditorium C only**, start as plenary, then to break out rooms; Harten, Ziegler and Laws facilitate
	3:15 – 3:30p	break	
	3:30 – 4:15	**Report out -- Let's go there together! Actions and plans**	Auditorium C only, plenary; Elstein
	4:15 – 5:30p	**EPA's science software portfolio continued (fire talks session)**	Auditorium C only, plenary; organizers: Yuen, cross-Agency presentations
	5:30 – 6p	End of day assessment	Auditorium, Elstein

	6 – 9p	Dinner / social	
DAY 4: Thurs, Oct 22 (AM only)	9 – 9:15a	Opening: re-cap day 3, and plan for day 4	Auditorium, plenary; Elstein
	9:15a – 9:45a	Exercise report out and close out (synergy, challenges, others as needed)	Auditorium, plenary; Yuen, Burch, Harten
	9:45 – noon	**Action formulation (combination of open discussion and table discussions/exercises)**	Auditorium, plenary; Elstein
Post workshop, optional: Thurs, Oct 22 (PM)	2 – 5p	**Assignments, next steps, tasks, documentation & publications** - Revisit each session from workshop, parking lot items, assimilate information, develop rough proceedings, determine if information is missing and follow up - Look to fine tune and document investment strategies, recommendations, synergies and/or case studies - Document action plan in proceedings and one-pager - Revisit and update DDES action items from the August 2014 meeting (SharePoint site, charter, etc) - Consider annual science software reuse and interoperability award - Continue to refine journal article and any other publications	Auditorium, plenary; meeting planners, DDES staff, senior leadership, and ALL interested parties are welcome to attend this optional session

Appendix D: US Government calls to action for IT

The following table provides a chronology of actions, directives, executive orders, memos, and guidance documents initiated by the Federal Government that target Information Technology (IT).

Table 1. US Government executive orders, directives, memos and guidance documents that apply directly to the intersection of environmental and information science

US order, directive, guidance	URL	Excerpt
Information Technology Management Reform Act of 1996 (AKA Clinger-Cohen Act)	http://govinfo.library.unt.edu/npr/library/misc/s1124.html	…promote and be responsible for improving the acquisition, use, and disposal of information technology by the Federal Government to improve the productivity, efficiency, and effectiveness of Federal programs, including through dissemination of public information and the reduction of information collection burdens on the public
Executive Order 13011, Federal Information Technology, 1996	https://www.gpo.gov/fdsys/pkg/FR-1996-07-19/pdf/96-18555.pdf	…A Government that works better and costs less requires efficient and effective information systems… Agencies now have the clear authority and responsibility to make measurable improvements in mission performance and service delivery to the public through the strategic application of information technology…
Information Quality Act (IQA) or Data Quality Act (DQA)	https://www.whitehouse.gov/omb/inforeg_agency_info_quality_links/	… ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies…
E-Government Act of 2002	https://www.gpo.gov/fdsys/pkg/PLAW-107publ347/pdf/PLAW-107publ347.pdf	… establishing an Administrator of a new Office of Electronic Government within the Office of Management and Budget… promote use of the Internet and other information technologies to provide increased opportunities for citizen participation in Government… promote interagency collaboration in providing electronic Government services…
FISMA - Federal Information Security Management Act of 2002	http://csrc.nist.gov/drivers/documents/FISMA-final.pdf	…provide a comprehensive framework for ensuring the effectiveness of information security controls over information resources that support Federal operations and assets…
White House memo - Transparency and Open Government, 2009	http://www.whitehouse.gov/the_press_office/Transparency_and_Open_Government; http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_fy2009/m09-12.pdf	…ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government.
OMB 25 Point Implementation Plan to Reform Federal Information Technology Management, 2010	http://www.dhs.gov/sites/default/files/publications/digital-strategy/25-point-implementation-plan-to-reform-federal-it.pdf	This plan is divided into two sections: Achieving Operational Efficiency and Managing Large-Scale IT Programs Effectively. The first section outlines the steps being taken to adopt cloud solutions and leverage shared services. The second section covers the structural areas that impact the success rates of large IT programs across government.

OMB Federal Cloud Computing Strategy, 2011	https://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/federal-cloud-computing-strategy.pdf	… Cloud First policy.
OMB Contracting Guidance to Support Modular Development, 2012	https://www.whitehouse.gov/sites/default/files/omb/procurement/guidance/modular-approaches-for-information-technology.pdf	…provide agencies with contracting guidance to support modular development, as required by Information Technology (IT) Reform Action 15: Issue Contracting Guidance and Templates to Support Modular Development, 25 Point Implementation Plan to Reform Federal Information Technology Management.
OMB Digital Government Strategy, 2012	http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html	New expectations require the Federal Government to be ready to deliver and receive digital information and services anytime, anywhere and on any device. It must do so safely, securely, and with fewer resources. … innovate more with less, and enables entrepreneurs to better leverage government data to improve the quality of services to the American people.
White House memo - Building a 21st Century Digital Government, 2012	https://www.whitehouse.gov/sites/default/files/uploads/2012digital_mem_rel.pdf	…ensure that agencies use emerging technologies to serve the public as effectively as possible.
White House Digital Services Governance Recommendations, 2012	https://www.whitehouse.gov/digitalgov/digital-services-governance-recommendations	…help agencies develop or strengthen their governance structures across all three layers of digital services: information, platform, and presentation.
OMB memo - Open Data Policy-Managing Information as an Asset, 2013	http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf	Information is a valuable national resource and a strategic asset to the Federal Government, its partners, and the public… executive departments and agencies (hereafter referred to as "agencies") must manage information as an asset throughout its life cycle to promote openness and interoperability, and properly safeguard systems and information.
Executive order: Making Open and Machine Readable the New Default for Government Information, 2013	http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-	Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation.
FITARA, H.R.1232, Federal Information Technology Acquisition Reform Act of 2014	https://www.congress.gov/bill/113th-congress/house-bill/1232	…make available to the public the cost, schedule, and performance data for each major information technology investment…

n.b. This table includes guidance, executive orders, memos, directives, etc. directly applicable to information technology, environmental science and the US Government. There are many unlisted materials that apply generally to our areas of interest, including for example, government operations and efficiency. For example, OMB's Capital Programming Guide (1997 – 2015; https://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/capital_programming_guide.pdf) provides strategic planning guidance, not specific to IT alone, but broadly applicable to budgeting and planning for life cycle management, reduction in redundancies, etc.

The US Government has also recognized the slow adaptation to these initiatives among Agencies – notably, across two political administrations. The following is a sample of expressions to this effect.

"The Federal Government has had uneven success in applying advances in information technology to enhance governmental functions and services, achieve more efficient performance, increase access to Government information, and increase citizen participation in Government."

– E-Government Act of 2002

"Information technology should enable government to better serve the American people. But despite spending more than $600 billion on information technology over the past decade, the Federal Government has achieved little of the productivity improvements that private industry has realized from IT."

– OMB 25 Point Implementation Plan to Reform Federal Information Technology Management, 2010

"The Federal Government's current Information Technology (IT) environment is characterized by low asset utilization, a fragmented demand for resources, duplicative systems, environments which are difficult to manage, and long procurement lead times. These inefficiencies negatively impact the Federal Government's ability to serve the American public."

– Whitehouse Federal Cloud Computing Strategy, 2011

"… it is time for the Federal Government to do more. For far too long, the American people have been forced to navigate a labyrinth of information across different Government programs in order to find the services they need. In addition, at a time when Americans increasingly pay bills and buy tickets on mobile devices, Government services often are not optimized for smartphones or tablets, assuming the services are even available online."

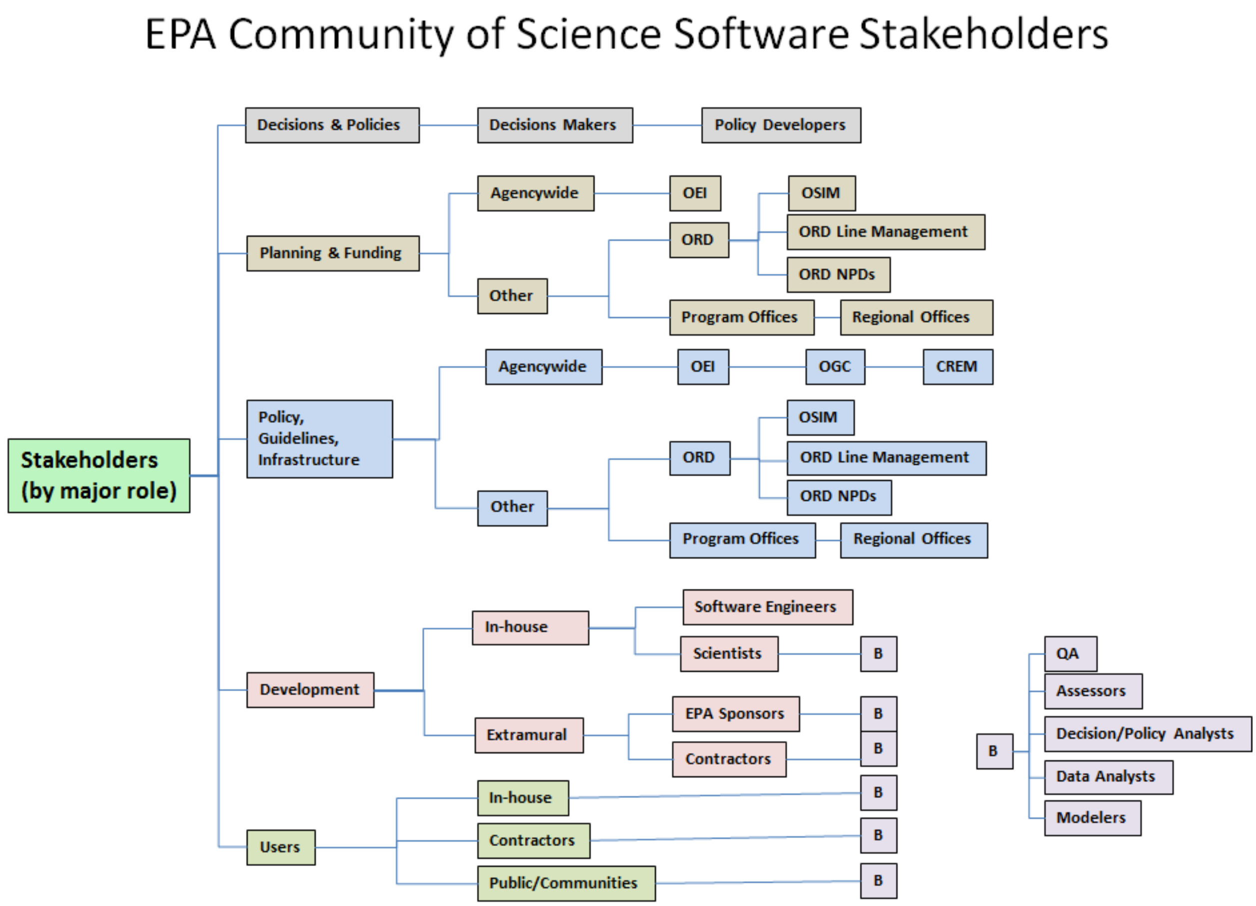
– Whitehouse memo on Building a 21st Century Digital Government, to executive branches and agencies, 2012

Appendix E: RCS-based summary of the EPA software portfolio (including both science-based and non-science-based products)

Product or resource type	#	Description	Examples
Geospatial dataset	3572	Datasets that includes geospatial information like longitude, latitude, and other special metadata. These datasets are usually accessed through geospatial Web services or geospatial (i.e., mapping) tools.	dataset Waste Water NHD Locations, Alaska coastline 1 – 63,360, 1990 Census Blockgroups for the PNW
Non-geospatial dataset	1511	Dataset that does not include geospatial information	Toxic Release Inventory dataset, Energy Star qualified residential clothes washers
Scientific models	262	Computer representation of environmental systems, problems or solutions	Atmospheric-Ocean model, Ecosystem Assessment Tool, Waste Reduction Model
Systems	349	Computer software that performs one or more specific tasks.	Environmental dataset Gateway (EDG), EnviroAtlas, Combined payroll redistribution and reporting system, Enviromapper, PeoplePlus
Custom tool	462	Any piece of software (utility, application, etc.) that performs one or more functions and that's been specifically built for a specific solution or need. These are not available commercially (Commercial Tools).	ePermit, Energy Star find a product, Discharge monitoring reports, NHDPlus append tool, PARIS III find greener solvents tool.
External developer resource	125	A resource or group of resources from an external institution or person who is not an Agency partner (state, tribe).	Bureau of Economic Analysis API, DATA.gov developer resources, HHS – Medicare developer resources
Mobile app	39	Software tool that runs on a mobile device (smart phone, tablet computer or similar device)	greenMeter for the Apple iPhone, AIRNow, Light bulb finder

		connected to the internet via wi-fi.	
Programming code Repositories	42	Resource that registers or contains reusable code, whether pieces of code, libraries, or code of complete projects.	GitHub, Computer Science Corporation has put together a library of programming code in order to reduce development time.
Misc		Data dictionary, Data exchange, Data model, REST/SOAP web services, web portal, widget tool, XML schema	

Appendix F: Science Software Stakeholders



From the organizational perspective the stakeholder community may be viewed as follows. EPA's Diversity Dashboard indicates that roughly 4% of total staff are considered "Information Technology" according to OPM's occupational classification (Fig A) (US EPA Office of Administration and Resources Management, 2015).

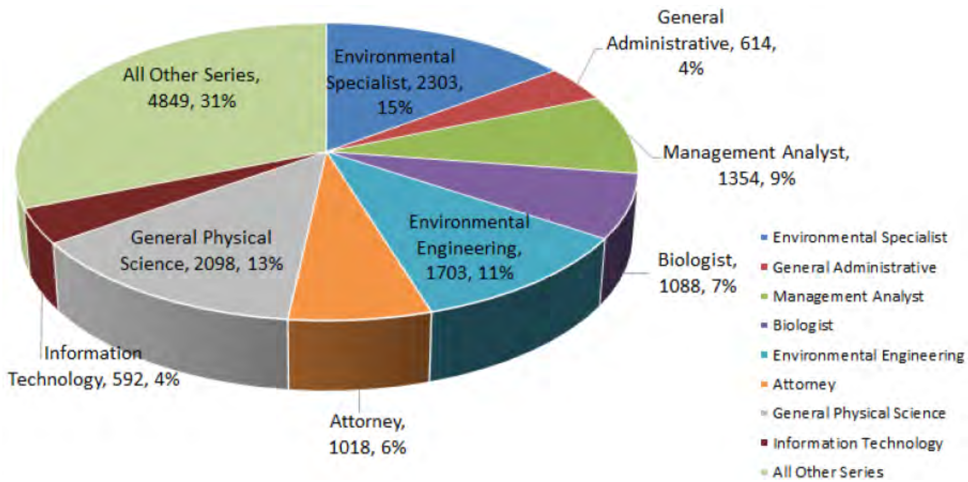


Figure A. 2015 US EPA staff occupation classifications; source: US EPA Diversity Dashboard report for FY 2015, 4th Quarter (US EPA Office of Administration and Resources Management, 2015)

The EPA consists of 12 headquarters Program offices and 10 Regional offices located across the United States (Table 1). The EPA has a Chief Information Officer (CIO) and a programmatic office (Office of Environmental Information or OEI) responsible for issuing OMB and Agency IT policies and procedures. EPA/OEI aims to identify and implement innovative IT solutions that cross-cut all of EPA. In addition to EPA/OEI's centralized authority, the Agency implementation of IT policy is delegated to SIOs in each Program Office and Region. This decentralized approach allows flexibility among SIOs – and therefore inconsistency – in cross-Agency implementation of policy. Very few Program Offices and Regions have skilled IT project level and management staff. Program and Regional offices, therefore, have unique science software development and management approaches that are not necessarily consistent with best practices referenced in Agency and OMB policy. Partly as a result, IT best practices and guidance (e.g., directives issued by OMB as listed above) have not been implemented.

Table 1. EPA Offices.

Headquarters offices	Regional offices around the nation
Office of Administration and Resource Management	Region 1 / Boston Serving CT, ME, MA, NH, RI, and VT
Office of Air and Radiation	Region 2 / New York Serving NJ, NY, Puerto Rico, and the U.S. Virgin Islands
Office of Chemical Safety and Pollution Prevention	Region 3 / Philadelphia Serving DE, DC, MD, PA, VA, and WV
Office of the Chief Financial Officer	Region 4 / Atlanta Serving AL, FL, GA, KY, MS, NC, SC, and TN
Office of Enforcement and Compliance Assurance	Region 5 / Chicago Serving IL, IN, MI, MN, OH, and WI
Office of Environmental Information	Region 6 / Dallas Serving AR, LA, NM, OK, and TX
Office of General Counsel	Region 7 / Kansas City Serving IA, KS, MO, and NE
Office of Inspector General	Region 8 / Denver Serving CO, MT, ND, SD, UT, and WY
Office of International and Tribal Affairs	Region 9 / San Francisco Serving AZ, CA, HI, NV, American Samoa, Commonwealth of the Northern Mariana Islands, Federated States of Micronesia, Guam, Marshall Islands, and Republic of Palau

Office of Research and Development

Office of Solid Waste and Emergency Response

Office of Water

Region 10 / Seattle
Serving AK, ID, OR, and WA

Appendix G: Examples of current EPA protocols for digital science software development

Digital science development at the EPA can be classified into 4 generic categories:

- Mobile Applications
- Web Applications
- Non-Web Software
- SharePoint Applications

MOBILE APPLICATIONS

EPA has established two review boards managed by OEI for approving mobile applications: The Mobile Access Review Committee (MARC) and the Mobile Application Approval Committee (MAAC). The MARC is set up to review mobile application concepts for apps that are developed by the EPA (external and internal). In addition, the MARC provides best practices, advice on what type of mobile app development strategy should be followed and standardized look and feel requirements for mobile application. All mobile applications developed by EPA should be listed in RCS and Developer Central (see Fig. A). The MAAC reviews mobile applications that are primarily developed by third parties for use on EPA mobile devices. There review consists of both a legal and security measures.

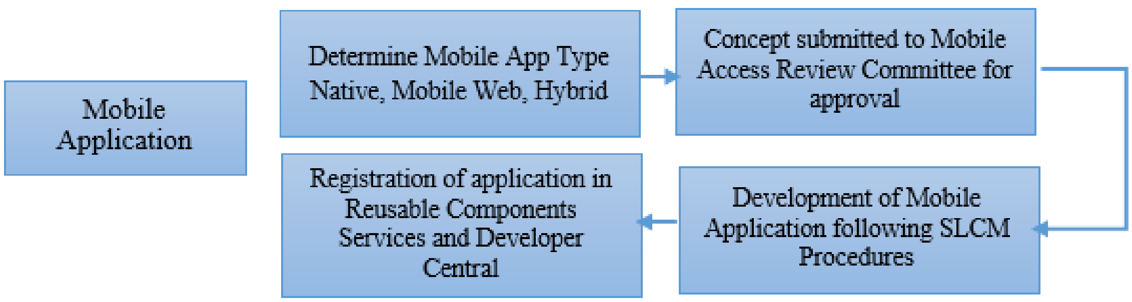


Figure A. Simplified Mobile app development; detailed diagram located here: https://prezi.com/yrgkvxuje0rp/epa-software-application-development/

WEB APPLICATIONS/NON-WEB SOFTWARE

There are numerous FITARA controls that apply to both web and non-web applications starting with budget formulation and planning controls. The budget formulation and planning controls are meant to engage the CIO in the budget process to ensure that IT needs are properly planned and resourced. For web applications that require a server, there is an application deployment checklist that must be filled out in order to acquire hosting services. Currently, all applications must be hosted at the National Computing Center (NCC). Prior to deployment of any application, a security plan must be developed. There are a variety of OEI enterprise systems that must be considered when developing software. These systems are shown in Figure B.

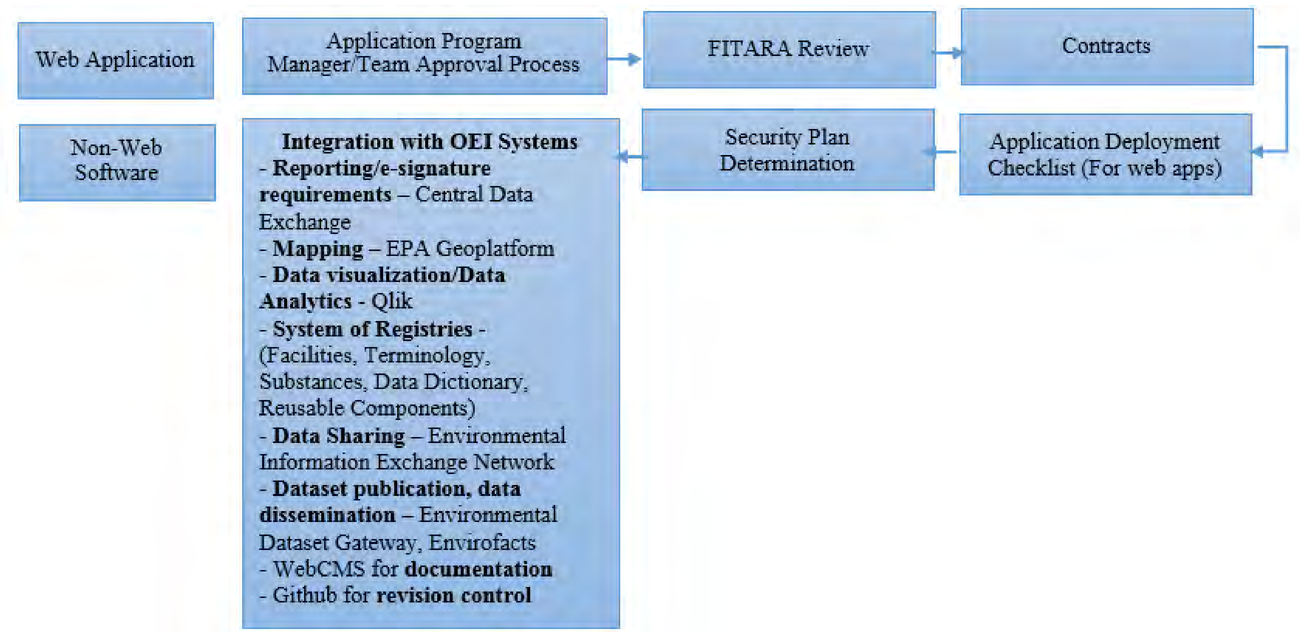


Figure B. Simplified Web Applications/Non-Web Software development; detailed diagram located here: https://prezi.com/yrgkvxuje0rp/epa-software-application-development/

SharePoint Applications

SharePoint applications can either be developed by EPA or by third parties. In cases where a SharePoint application is developed by EPA, there is a redundancy check to make sure that other similar applications do not already exist (Fig. C). For applications developed by a third party, there is a legal review prior to OEI SHARP review. After SHARP review, if a security plan is required, it is developed. Once the application is developed, the code for the application is scanned and reviewed by the NCC. After the application is made available, it is listed in SharePoint.

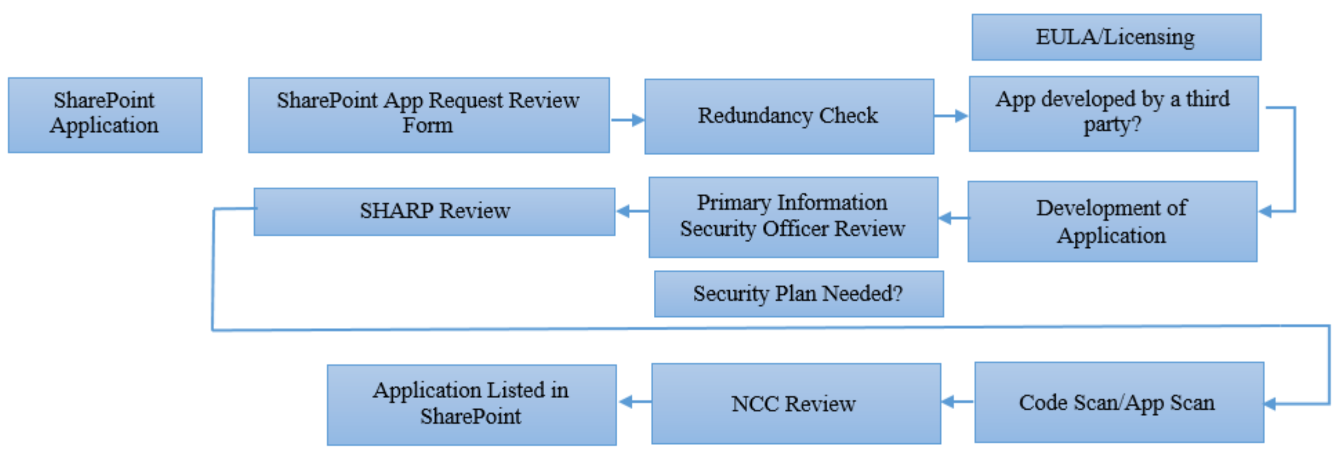


Figure C. Simplified SharePoint application development; detailed diagram located here: https://prezi.com/yrgkvxuje0rp/epa-software-application-development/

Governance

The Office of Water's governance framework consists of an Information Steering Committee (ISC) and an Information Management Project Management Office (IM PMO). The ISC provides a senior management governance construct for IT investments within OW. In addition, the ISC has a 5-year strategic plan to set direction for OW IT portfolio and program process improvement. The core mission of the ISC is to approve all OW IT investments, disinvestments and budget actions, and to coordinate with the EPA IT governing body and the CIO.

The IM PMO was established to lead IT portfolio change management and business improvement. In addition to providing support to the ISC, the IM PMO provides consultation on all IT investments and is the core team for managing IT project and enterprise architecture across OW. Figure D walks through the various components of the software application development process in OW. The IM PMO helps IT project managers within OW through each component of the software application development process by providing PM Procedures (documentation/templates/training) in addition to acting as consultants.

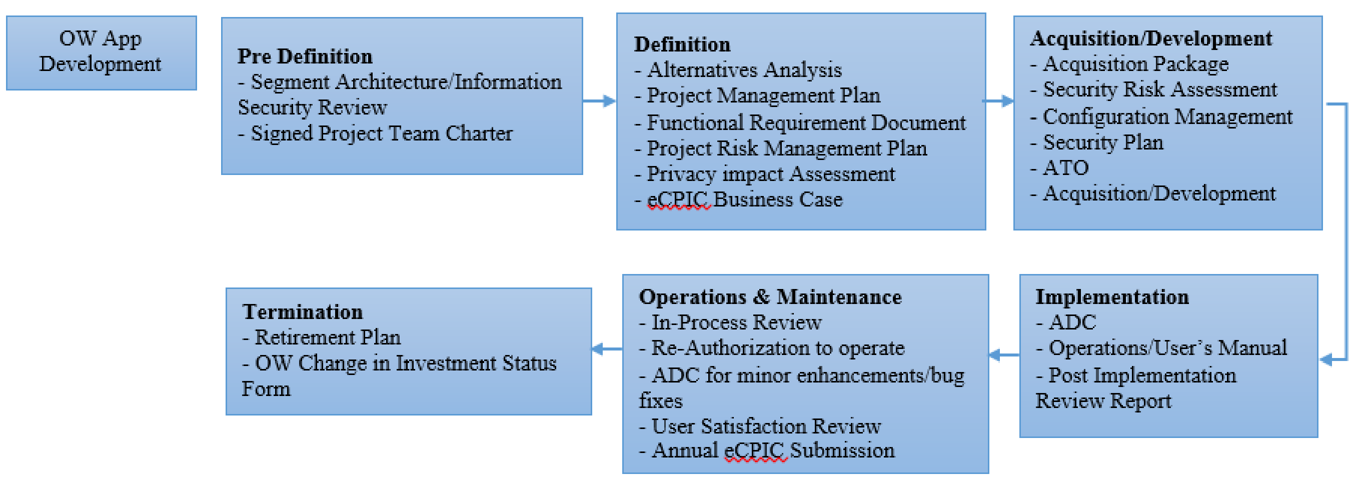


Figure D. Simplified OW application development; detailed diagram located here: https://prezi.com/yrgkvxuje0rp/epa-software-application-development/

The Office of Research and Development (ORD) has a less defined governance framework when compared with OW. A review board does not exist to approve/disapprove projects, set strategic goals or coordinate with EPA's CIO/IT governance body. In addition, a project management office does not exist to assist in walking IT project managers through the software application development process. Figure E walks through the various components of the software application development process in ORD.

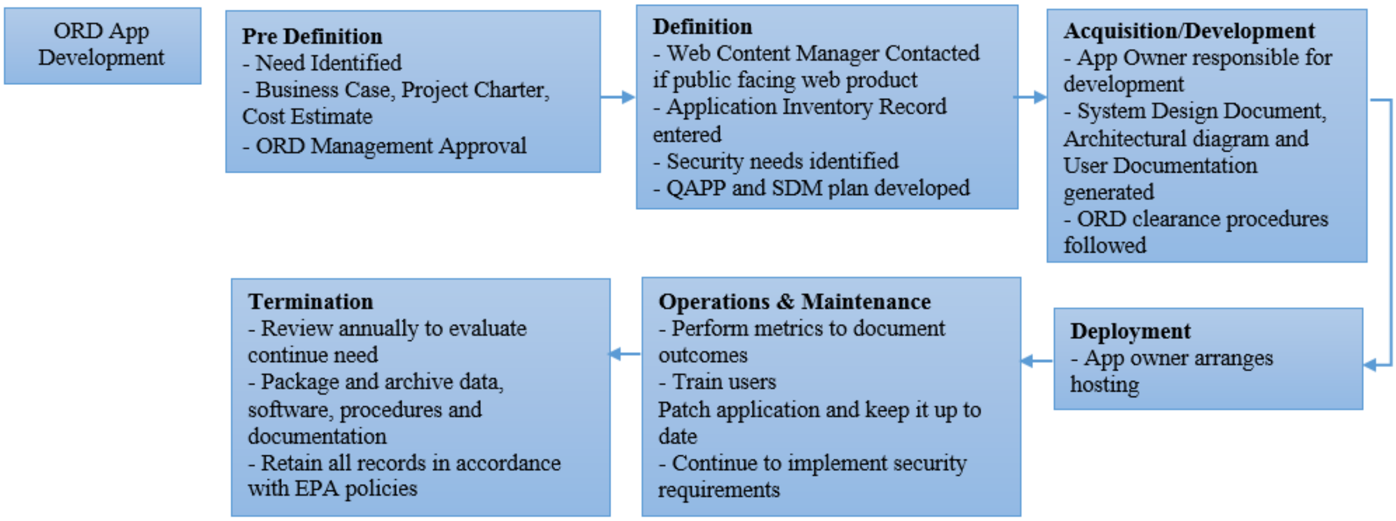


Figure E. Simplified ORD application development; detailed diagram located here: https://prezi.com/yrgkvxuje0rp/epa-software-application-development/

Appendix H: Current Communities of Practice at EPA

Lave and Wenger (http://www.learning-theories.com/communities-of-practice-lave-and-wenger.html) describe CoPs as "groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly." Lave and Wenger go on to explain that CoPs require three elements:

- **Domain:** Members have a commitment/passion for a shared domain of interest
- **Community:** Members willingly engage with and help one another increase knowledge in the domain of interest
- **Practice:** Members are practitioners in the domain of interest

In EPA, other "unwritten" requirements may include:

- **A champion:** Someone with the time and energy to organize and maintain the community
- **Organizational acceptance:** Management must enable practitioners to spend part of their working hours participating in the community

In the January-February 2000 issue of the Harvard Business Review (https://hbr.org/2000/01/communities-of-practice-the-organizational-frontier), Wenger and Snyder describe CoPs as "complementary" to the existing organizational structure. At the time, CoPs were just beginning to enter the business environment. They found that CoPs could not be dictated by management, but rather management needed to nurture them by providing the foundation for the communities to thrive.

At EPA, many CoPs arose on top of – or superimposed on – the existing US EPA organizational chart. They often begin as rogue efforts, sometimes failing and other times achieving various levels of formal integration within US EPA. CoPs have evolved as an attempt to facilitate communication and collaboration across the Agency around specific areas of expertise, needs and/or interests. Specific areas include, but are not limited to:

- Modeling
- Data analytics and visualization
- R
- GIS
- SAS
- Statistics
- Scientific Data Management
- Developing and Deploying Environmental Software
- Computational Toxicology
- Strategic Foresight
- Records Management
- Project Management

Table 1 provides a summary of some of the CoPs related to digital science that were presented at the 2015 EPA October workshop:

Table 1. US EPA digital science communities of practice

Community of Practice (Title)	Purpose, Relevant Links
EPA GIS Workgroup/Geospatial Community	Purpose: The purpose of EPA's GIS Workgroup is to promote coordination and planning necessary for the development of comprehensive, unified GIS and data integration programs within EPA. Technology: GIS Workgroup members generally take advantage of desktop, server, and mobile GIS software, and data and API licenses available under EPA's enterprise license agreements (Esri, Bing, Google, HERE and others. Increasingly users are taking advantage of the EPA's ArcGIS for Organization subscription (the collaboration component of the EPA GeoPlatform) to share web applications and data services with co-workers and the public. The Developers Subgroup of the GIS Workgroup is focused on standardizing spatial application templates and widgets and sharing code through EPA's Github account. GIS Workgroup SharePoint EPA GIS Intranet Resources EPA GeoPlatform SharePoint EPA Public Geospatial Resources
ORD Geospatial Sciences User Group	Purpose: The ORD Geospatial Science User Group facilitates collaboration and information exchange among all ORD staff interested and working with geospatial sciences and related fields, and using tools such as GIS. The Group functions as a liaison between the research staff and the OSIM subject matter experts for geospatial-related tools. Technology: The issues with GIS related ESRI and open source software are bought to the meeting. People discuss the issues in using different modules of these software. The issues and bought by the group here are moved up to either ORD management or EPA GIS steering committee. There usually is some discussions about to data hosting and data management. https://usepa.sharepoint.com/sites/ORD_Community/Geospatial-Sciences-User-Group/_layouts/15/start.aspx#/SitePages/Home.aspx
Developing and Deploying Environmental Software (DDES)	Purpose: This effort represents a cross-Agency initiative to bring together the community of stakeholders responsible for environmental software. Our collective purpose is to jointly surface, discuss, and resolve the challenges of DDES at EPA. The DDES CoP grew out of an ORD project focused on reuse and interoperability of scientific software. It is comprised of members who have widely varying roles and responsibilities with respect to environmental software. At the core are practitioners who design, develop, deploy, and maintain the software. The practitioners may be focused on science software or supporting IT software. More and more, these sub-communities work together to improve the overall environmental software enterprise from a grass roots (bottom up) perspective. Equally important are members who focus on management of efforts that include environmental software. Primary software-based concerns from this top down perspective include managing the enterprise, funding, facilitating collaboration, infrastructure development, and policies that ensure the quality and security of software. https://usepa.sharepoint.com/sites/ORD_Community/DDES/

EPA SAS Users Group	Purpose: To be the primary forum for our 500 users EPA-wide on SAS-related topics, ranging from software installation to common problems to training course information. Technology: Used for data processing, analysis, mapping and graphics. Primarily used as independent desktop application (i.e. "Desktop SAS"). Also used as an Internet service to deliver data and graphics on EPA's AirData website - http://www3.epa.gov/airdata/. **SharePoint Site is in approval process and will be rolled out in FY16.**
R User Group	Purpose: To provide a forum for R users to resolve issues, discuss R packages/code, and learn from each other by sharing expertise and experience, thus improving collaboration across locations. Technology: R is an open source software language and environment for statistical computing and graphics (source: https://www.r-project.org/). It is very widely used in government, academia and industry for statistics, visualization, general purpose computation, and is one of the most commonly used languages for data science. R is extended through the use of R packages which provide a common framework for developing new functionality and ensure interoperability between R users. R packages are distributed through repositories, primarily the Comprehensive R Archive Network (CRAN, https://cran.r-project.org/), but may also be reused via source files, served on GitHub, or via other repositories. Additionally, R integrates with other software and languages including C++, Python, ArcGIS, and javascript. R is freely available and installed on EPA computers so that a user can freely access and use R packages. Service Catalog: http://intranet.ord.epa.gov/administrative/itresources/service-catalog/r-software-ords-research Training: http://intranet.ord.epa.gov/administrative/training/e-learning/scientific/r-training SharePoint Site: https://usepa.sharepoint.com/sites/ORD_Community/R-User-Group/
Scientific Data Management	Purpose: The Scientific Data Management Community of Practice (SDMCoP) supports members of ORD's research community in sharing information and ideas about managing scientific data. This group is intended to grow over time into a long-term, productive, self-sustaining collaborative community where members learn about SDM topics, help each other solve data management problems, and provide input to enhance SDM practices across ORD. Technology: At one SDM CoP meeting, a demonstration was given of a very sophisticated tool for managing scientific data. It is anticipated that similar tools for managing scientific data will become more prevalent in the future. It is the goal of the SDM CoP to share such tools in the future.

	SharePoint Site: https://usepa.sharepoint.com/sites/ORD_Community/SDMCoP/_layouts/15/start.aspx#/SitePages/Home.aspx ORD@WORK Site: http://intranet.ord.epa.gov/science/scientific-data-management
Environmental Modeling (E-Mod)	Purpose: The Environmental Modeling (E-Mod) Community of Practice (CoP) has been established for modelers (i.e., developers, users, evaluators, and those with a passion for modeling [e.g., statisticians, software analysts] - likeminded individuals) to come together as a community with an open opportunity to regularly interact with one another, form new collaborative relationships, and possibly pursue a project(s) to advance environmental modeling at the US Environmental Protection Agency (EPA or the Agency). This is a cross-Agency opportunity to learn about the exciting modeling related activities ongoing or needed within (or potentially outside of) the Agency; collaborate with colleagues to solve an issue; learn about new modeling techniques, tools, or opportunities (i.e., internal or external to the Agency); talk with individuals that use the same language, among other things. Technology: There have been multiple presentations that involved the presenter discussing the software they used and a presentation specifically about interoperability. More will likely continue in the future. https://intranet.ord.epa.gov/p2/e-mod/home https://www2.epa.gov/modeling

Appendix I: Practitioners perspective and pain points

This section summarizes and synthesizes two panel sessions of the 2015 workshop: Practitioners and IT for Science.

Practitioners

The practitioners' panel was convened (Table 1) with the objective to advance community-wide awareness and knowledge of current and future protocols, practices and challenges associated with the software life cycle management (SLCM) process and to promote a cross-Agency approach for reuse and interoperability of science software products. For this panel, a "practitioner" is someone involved in the development, maintenance and/or deployment of science software products, including:

- software engineers (IT components/systems)
- software engineers (science components/systems)
- scientists who develop software (i.e., write code for components/systems)
- scientists who do not develop software but manage/fund its development

Table 1. Panel members, their EPA office affiliations and primary digital science role.

Panel Member	Organizational Office/Lab	Primary Role
Mike Galvin	ORD/National Exposure Research Lab	Software Engineer
Vince Allen	OW/IM/IT Project Management Office	IM/IT Project Manager
Marc Weber	ORD/National Health and Environmental Effects Research Lab	Scientist, R developer
Janet Kremer	Region 3, Environmental Assessment and Innovation Division	Modeler, software user
Paul Harten	ORD/National Risk Management Research Lab	Scientist, developer
Annie Neale	ORD/National Exposure Research Lab	Scientist, software development management

Each panel member was asked to prepare a 5 to 7 minute presentation focused on one or more of the following questions:

1. What type of software are you associated with (e.g., tools to support science software, databases, models, etc.)?
2. Briefly describe the process that is followed for software development you are associated with (e.g., waterfall, agile, ad hoc, contractor driven, etc.)?
3. How do you address (if at all) reuse and interoperability in terms of preparing your software for R&I by those outside your development/user community?
4. What is your experience with reusing and integrating software produced by others? What would you do to improve the experience?
5. What are the challenges facing practitioners with respect to science software development in general and the goal of increasing reuse and interoperability in particular? What are practitioners doing to address these challenges (e.g., strategies to avoid/adapt to/work around/etc.)? [Here we will use the existing list of pain points (see below) and ask panel members to add to/modify/prioritize the list and describe their strategies.]
6. What is the role of contract support and how should software development by contractors be managed to ensure reusability and interoperability by others?

The following is a list of main ideas and recommendations that surfaced during the presentations and open discussion. Many pain points were reiterated during these discussions, and can be found further below.

Interoperability (Practices & Standards)

- Interoperability is important at two levels; within systems and across systems. Interoperability generally refers to software components (e.g., tools, models, databases) designed to enable the sharing of their data and information with other components. Within system interoperability is the most common and refers to components that have been designed to exchange data and information using protocols designed for a specific system (aka framework). Inter-system interoperability, much less common but much more powerful, refers to components that are designed in a system independent manner, that is, the exchange protocols are not tied to a specific system (e.g., components designed as web services).
- The object oriented programming paradigm is conducive to producing reusable and interoperable software. For example, the idea of separating (or uncoupling) the presentation layer, business logic layer, and data processing layer in terms of functional code.
- Tying datasets to data standards (national scale standards) benefits both developers of individual datasets as well as others (scientists, other Federal agencies, public) in terms of:
    - accessibility
    - integration of data from multiple datasets
    - adding value (functionality) to content
- Web services is a game changer with respect to interoperability. Web services removes constraints related to programming languages, platforms, and operating systems and allows direct access to the functionality of software through an API.
- Several panelists have used and recommend the agile development process. Developers explained that Agile is helpful so they can continuously gather feedback from users and reflect it in the evolving software.

- Try to tie programmatic needs to national data standards so that different services are compatible with each other, both within the Agency and across agencies and organizations.
- This includes using standard software libraries and widget toolkits, version control and Communities of Practice.
    - Some try to develop code and data in such a way that anyone should be able to reproduce their work.
- Adaptability (i.e., the ability and willingness to adapt to new and better standards and practices) is important.
- Interoperability practices and standards are not yet mature (i.e., stable, universally accepted, etc.) and therefore can be a burden for developers.

Open Source

- Open source development and deployment facilitates reuse and interoperability in the following ways:
    - by providing no cost access to the best tools to support development
    - by enabling collaboration when developing in a public version control system such as GitHub
    - by providing well documented packages to help understanding and reuse
- When starting a new development effort, developers should determine if relevant software solutions already exist (as opposed reflexively pursuing new software).
- While some custom coding is usually necessary, some developers (e.g., seasoned software engineers) prefer to use open source material.

Resources & Tools

- It makes sense to put coding and software into GitHub because this is one of the most widely used repositories. GitHub also provides access to use cases that other developers have worked with in the past.
- There are several pre-existing applications developers could leverage.
    - ORD has been creating a scientific and administrative application inventory that researchers and developers can access to learn about what has already been created.
    - Reusable Component Services (RCS) is a catalog of IT services, Web services, widgets, etc. where people should register their tools and IT services.
    - The Environmental Data Gateway (EDG) is a public-facing platform, but it could be useful for developers as well.
    - The GeoPlatform has proven to be an excellent platform for developing geospatial data processing and display software.

Contracting & Contractors

- Contractors are necessary and should be managed well.
- With respect to contracts and contractor support there needs to be IT standards applied to contracts. At present different contractors choose from among many software development options and this results in software that is not efficiently and cost effectively reused. Contractors need to be clearly informed of and held accountable for the IT standards they should follow.

- OARM needs to be part of the conversation going on in this workshop.

Miscellaneous

- Some panelists believe the knowledge needed for software development should be institutionalized, but since there is a shortage of developer resources this may not be possible.
- Challenges that the panelists described included finding the correct infrastructure needed to support Web and mobile applications, and waiting for hardware and software installations.
    - Contractors handle software installations, which can be frustrating because developers have to wait and the contractors are unfamiliar with scientific software and unable to meet the needs of developers.
    - A lack of access to expertise can prevent developers from being able to use software.
    - Hosting charges can be a pain point for developers who host large data sets.
- Systems can be complex and continued access to original developers and science experts is important
- Marketing (making it known that software we produce is available) is important. We are not sales people or market people, we are scientists and developers. The inventories being built by OEI and ORD are a good start. There remain issues of access and usability associated with the individual inventory items.
- There is a need for more software engineers within the Agency. Current practice is, for example, to hire a science oriented post doc and expect them to produce reusable and interoperable software. Tools, practices, standards, and access to knowledgeable software engineers is needed by the scientists inclined to develop code.

Challenges and Pain Points

Application development and management within in EPA is challenging for both scientists and management at multiple levels within EPA. Researchers, scientists and regulators developing scientific software/applications (e.g., models, databases, decision support systems, etc.) in support of EPA's mission encounter both real and perceived obstacles. Funding and security are perceived as obstacles, but may also be the challenge because scientific software is often not developed in cross-disciplinary, integrated teams that include skilled software developers, security and architectural professionals. Additionally, the general lack of understanding at all levels within Program Offices and Regions of the programs of White House Executive Orders, OMB Memorandum, and Agency IT policies does not allow for scientific software developers to leverage the best practices and processes learnings of other software development efforts. What is perceived as administrative burden by scientists is really the result of the Agency not hiring staff with critical modern skills in IT project management that are needed to support the development and management of software (see also "Hiring Tenacity" section above).

Developing software always presents challenges. Developing software within a bureaucracy, especially one for which software development is not the main focus, presents additional challenges (Curtice et al., 2012), and this is true for EPA. The EPA, as a large government bureaucracy has its own set of challenges, which many believe are holding us back in both efficiency and innovation.

Out of the coding conversations came the following list of barriers to developing software within the agency.

- Lack of collaboration tools for code development
- Need to query and use datasets and models that aren't in EPA's firewall
- Need to install software and updates without having to call for assistance
- Need to optimize application development and improve efficiency including managing contractor-based development
- Need to adopt and allow the use of, new technology more quickly
- Need a more streamlined process for freeware/shareware approvals
- Need subject-matter experts to assist with application development (computer scientists, data architects, IT project management)
- Need assistance in following Agency processes and meeting compliance requirements so the scientists can focus on the science
- NCC hosting costs are prohibitive
- Need to share software licenses and obtain software on a temporary basis (can't locate it)
- Need training on new software (e.g., moving from SAS to R; GitHub)
- Need more (and consistent) information about process and roles and responsibilities of different groups (OSIM, OEI, OARM, LCO)
- Need help finding contacts (both data owners and their IT support) in Program Offices to facilitate data sharing
- Need guidance/authority to establish data sharing with external sources
- Need guidance on developing QA protocol for shared data system – Who "owns" the data and is responsible for its quality?
-

Some of the items listed might exist in any software development environment, for example, needing training on new software, but many are the result of working within a large bureaucracy.

Appendix J: Activities and Approaches at Other Organizations

EPA is not the only organization struggling with challenges at the intersection of information and environmental sciences (Be, 2014; Curtice et al., 2012). The global organizational "landscape" of digital science efforts is evolving and growing (Fig. A).

Some efforts and organizations have emerged across environmental science domains to build collaborations, improve culture of sharing and to improve scaling, interoperability and re-use of digital components; these include, for example:

- Belmont Forum
- Research Data Alliance
- ESIP
- Coopeus
- Eye on Earth Alliance
- NSF's EarthCube

Additionally, many organizations that share similar missions as EPA are making changes to better leverage IT for science, including for example:

- USAID Global Development Laboratory: staffing scheme focuses on short term positions/efforts for IT-related work: AAAS fellows, 1 to 3 year government employee terms; hackathons and code-a-thons
- UN Global Pulse: New UN entity to work with big data
- UNEP: Advancing their concept of an online platform called *UNEP-Live* to share information and engage stakeholders
- NSF: Has founded numerous digital science efforts over the last decade including the data centric hubs (datanet, dataone, etc), NEON Inc. (currently undergoing changes) and EarthCube
- GSA 18F: Fostering government-wide agile business models and the presidential fellows program

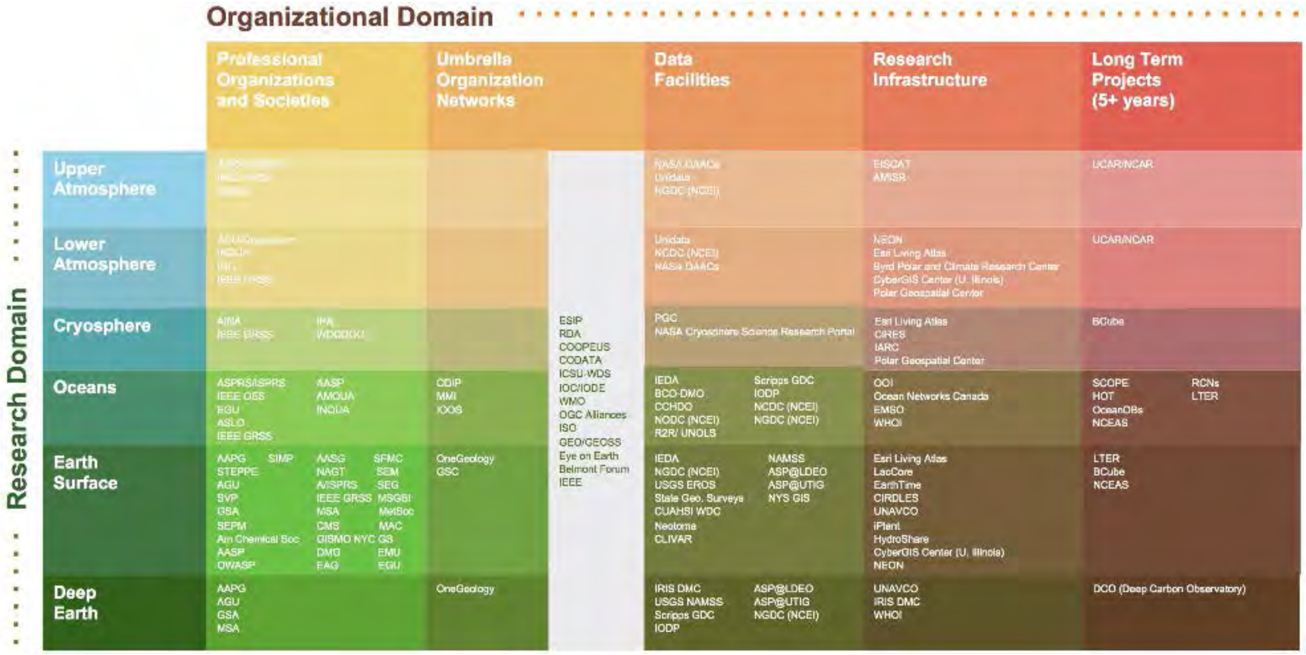


Fig. A. Landscape of organizations involved at the intersection of environmental and information sciences (source: Dawn Wright of Esri and the NSF EarthCube Liaison Team; http://www.arcgis.com/home/item.html?id=9bde7150da474c828d61a5e67e98855d)

Appendix K: References

Azadegan, A., Hauge, J. B., Bellotti, F., Berta, R., Bidarra, R., Harteveld, C., … Stanescu, I. (2013). The move beyond edutainment: Have we learnt our lessons from entertainment games? In *Proceedings of GaLA 2013 - 2nd Games and Learning Alliance International Conference*. Retrieved from http://www.aidaazadegan.co.uk/download/i/mark_dl/u/4011972943/4601296839/The move beyond edutainment.pdf

Be, W. (2014). Science and Cyberinfrastructure: The Chicken and Egg Problem. *Eos*. Retrieved from http://geomag.usgs.gov/downloads/publications/Kelbert_2014_EarthCube_EOS.pdf

Brookes, A., Burch, K., Elstein, K., Galindo, L., Harewood, C., Harten, P., … Ziegler, C. R. (2014). *Developing and Deploying Environmental Software at EPA Phase II: Framing the Issues and Building Community*. Retrieved from https://usepa.sharepoint.com/sites/ORD_Community/DDES/Shared Documents/Background Materials/DDES Phase II Report (Final Nov 2014).pdf

Costanza, R., Chichakly, K., Dale, V., Farber, S., Finnigan, D., Grigg, K., … Richard Ziegler, C. (2014). Simulation games that integrate research, entertainment, and learning around ecosystem services. *Ecosystem Services*, *10*, 195–201. http://doi.org/10.1016/j.ecoser.2014.10.001

Curtice, C., Dunn, D. C., Roberts, J. J., Carr, S. D., & Halpin, P. N. (2012). Why Ecosystem-Based Management May Fail without Changes to Tool Development and Financing. *BioScience*, *62*(5), 508–515. http://doi.org/10.1525/bio.2012.62.5.13

Gottschalk, P. (2009). Maturity levels for interoperability in digital government. *Government Information Quarterly*, *26*(1), 75–81. http://doi.org/10.1016/j.giq.2008.03.003

Laniak, G. (2012). *Environmental Software Reuse and Interoperability : An Initial Review and Recommendations for the Office of Research and Development*. Retrieved from https://usepa.sharepoint.com/sites/ORD_Community/DDES/Shared Documents/Background Materials/DDES Phase I Report (Final Aug 2013).pdf

Lee, G., & Kwak, Y. H. (2012). An Open Government Maturity Model for social media-based public engagement. *Government Information Quarterly*, *29*(4), 492–503. http://doi.org/10.1016/j.giq.2012.06.001

Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, *493*(7431), 159.

Piwowar, H. A., Becich, M. J., Bilofsky, H., & Crowley, R. S. (2008). Towards a data sharing culture: recommendations for leadership from academic health centers. *Plos Medicine*, *5*(9), 1315–1319. http://doi.org/e183 10.1371/journal.pmed.0050183

Priem, J. (2013). Scholarship: Beyond the paper. *Nature*, *495*(7442), 437–440.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. Retrieved from http://altmetrics.org/manifesto

R. Farina, C., Epstein, D., B. Heidt, J., & J. Newhart, M. (2013). Regulation Room. *Transforming Government: People, Process and Policy*, *7*(4), 501–516. http://doi.org/10.1108/TG-02-2013-0005

The Rockefeller Founding, & The Global Business Network. (2010). Scenarios for the Future of Technology and International Development. *Business*, 54. Retrieved from http://www.rockefellerfoundation.org/news/publications/scenarios-future-technology

US EPA Office of Administration and Resources Management. (2015). *US EPA Diversity Dashboard, Fiscal Year 2015, Fourth Quarter*.

US EPA Office of Inspector General. (2015). *EPA Needs to Improve Recording Information Technology Investments and Issue a Policy Covering All Investments (15-P-0292)*. Retrieved from https://www.epa.gov/sites/production/files/2015-09/documents/20150922-15-p-0292.pdf

Ziegler, C. R., Brookes, A., Burch, K., Harten, P., Kremer, J., Laniak, G., … Yuen, A. (2015). US EPA Digital Science: An Evolution. *American Geophysical Union*.

Ziegler, C. R., Miller, C. A., Kilaru, V., French, R. A., Costanza, R., & Brookes, A. (2013). Gamification – Environmental and Sustainable Development Organizations Could Do More. San Francisco, CA: American Geophysical Union - Annual Meeting (presentation).

Ziegler, C. R., Webb, J. A., Norton, S. B., Pullin, A. S., & Melcher, A. H. (2015). Digital repository of associations between environmental variables: A new resource to facilitate knowledge synthesis. *Ecological Indicators*, *53*, 61–69. http://doi.org/10.1016/j.ecolind.2015.01.003

EPA
United States
Environmental Protection
Agency

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT NO. G-35

Office of Research and
Development (8101R)
Washington, DC 20460