The University of North Carolina in association with the
California State Archives and the
California Natural Resources Agency

eLegacy: Engaging Participation in the Preservation and Management of Sacramento-San Joaquin Bay Delta Scientific and Environmental Data

Final Report
June 30, 2010

Kevin Brown
Steven Dambeck
Renée Taylor Consulting, Inc.
Table of Contents

1. Executive Summary ........................................................................................................... 3
2. The Interagency Ecological Program ............................................................................... 4
3. Current Preservation and Management of “IEP Data” ............................................... 14
4. Strategic Technical Architecture .................................................................................... 19
5. Attitudes toward Change ................................................................................................ 27
6. Transformation Planning and Implementation of a DRSE ............................................ 33
   Appendix A: IEP Stakeholder Power Grid ................................................................. 37
   Appendix B: Adaptive Management Process Inputs and Outputs.......................... 38
   Appendix C: Sample Memorandum of Understanding ........................................... 40
   Appendix D: Service Pattern “Provide Advisory Encounters” ............................ 45
   Appendix E: Data Exchanges in the IEP Organization Domain ............................. 50
   Appendix F: Repository Architecture for DWR ......................................................... 52
   Appendix G: Bibliography ............................................................................................. 53
1. Executive Summary

This project was designed to study the status of data preservation within a representative group of cooperating conservation organizations within the California Natural Resources Agency (CNRA), and to analyze the extent to which this status quo serves the on-going data needs of that entity. It was further designed to ascertain a methodology for engaging stakeholder participation in a more coherent set of federated information management practices as a platform to improve information synthesis and engage these partners in a sustainable federated data preservation network.

The CNRA entity selected for this study was the Interagency Ecological Program (IEP), which is tasked with studying the factors that affect the ecological, environmental, and water resources in the Sacramento-San Joaquin Bay-Delta (“Bay-Delta”) and to provide that information to State and Federal management and policy decision-makers. In this role, IEP finds itself at the heart of Bay-Delta scientific and environmental data flow: it generates its own data, consumes data generated by others and transforms data into scientific information.

In the current state of data preservation, IEP staff faces challenges in determining the existence and location of all germane data and, once known, in accessing and assessing it. Further, because collection protocols and metadata tags are not standardized, the meaning and quality of much of this data is uncertain. Due to these factors, and to an absence of a data integration function, IEP’s ability to analyze, synthesize and evaluate data in a timely, credible and replicable manner – that is, to transform Bay-Delta data into meaningful information -- is compromised.

On the positive side, data preservation and presentation systems and practices which could fully meet IEP needs do exist and are currently being successfully employed by other organizations. If one such system and set of practices were implemented, Bay-Delta data could be effectively transformed into the information needed to make critical management and policy decisions based on sound science.

Conditions are ripe for such a change. True, there is skepticism of IT initiatives of any kind, born of previous experiences; there is also legitimate concern related to data ownership and governance; and there is the normal human tendency to cling to any status quo. Ranged on the other side, however, is a strong recognition of the benefits that would accrue from a truly effective and agile data management structure -- not only among IEP staff, but also among the recipients of IEP efforts, those scientists and legislators whose management and policy decisions will result in far-reaching impacts on wildlife, the economy and the quality of human life in California.
2. The Interagency Ecological Program

The California Natural Resources Agency (CNRA), host for this study, has as its mandate to “restore, protect and manage the state’s natural resources...using creative approaches and solutions based on science, collaboration and respect for all...interests involved” (emphasis added). The more solid the science available to it, the more effectively it can play this complex and highly-charged role. In turn, the more accessible, trustworthy and interoperable its data, the more solid its science will be.

CNRA proposed the Interagency Ecological Program (IEP) as the subject for this study, because, as a nexus for the flow of scientific data and information, it illustrates the complex information synthesis challenges that arise in an adaptive management context, and because the large number of participating data exchange partners provide a broad sample of the diversity of roles in conservation management organization networks for digital preservation and other collaboration.

IEP’s Shifting Mandate

The IEP recently celebrated its 40th anniversary, and is justly proud of its consistency in addressing its mission to “provide information on the factors that affect ecological resources in the Sacramento-San Joaquin Estuary that allows for more efficient management of the estuary”. Member entities have maintained long-term scientific and environmental monitoring programs in their particular areas of expertise, and have worked together on a number of directed research projects (most notably over the last ten years to understand “pelagic organism decline” (POD) in the Bay-Delta).

In the last few years, IEP has become increasingly counted upon not only to provide such raw data related to the Bay-Delta ecosystem, but also to provide reliable and relevant analysis, synthesis and evaluation of that data. In this role it directly supports the efforts of the Delta Science Program (DSP) and others to elevate meaningful scientific information to those who determine Bay-Delta management practices and policy. In the vision of the IEP Lead Scientist, “the future IEP has a clear, rightful place in the emerging Bay-Delta governance structure and a clear role in the various Bay-Delta management and conservation plans. Sustained long-term monitoring will always be complemented by shorter-term monitoring and research projects, often in adaptive management settings.”
Organizational Structure

As an interagency program, IEP is comprised of representatives from all nine of its member agencies – **State:** Department of Water Resources (DWR), Department of Fish and Game (DFG), Water Resources Control Board (SWRCB); **Federal:** National Marine Fisheries Service (USNMFS), Army Corps of Engineers (USACE), Bureau of Reclamation (USBR), Environmental Protection Agency (USEPA), Fish and Wildlife Service (USFWS), and the U.S. Geological Survey (USGS) -- as well as a number of non-member partners and stakeholders.

Since each of the member agencies has its own mission and spheres of activity, it is not simple to comprehend the inter-relationships of IEP’s constituent parts or its scope of action. The context diagram, below, is one perspective on the Program, viewed in relation to its data and information flows. The universe of data is depicted in the lower half of the diagram, and the universe of knowledge is depicted in the upper half.

The red box in the middle of the model represents IEP. Some member agencies (DWR, DFG, USBR, USFWS and USGS) both contribute data and are consumers of information that emerges from IEP; others (USEPA, USNMFS, USACE and SWRCB) do not contribute data, but consume resultant information. A significant number of other science partners also contribute data and/or consume information. While contributing to IEP, all of these entities at the same time maintain an existence separate from it within the world of Bay-Delta science (green box).

Below the green line delimiting the Bay-Delta scientific community are found entities and individuals which either contribute to or consume Bay-Delta data. Above the green line are the ultimate customers of the scientific work done by the scientific community at large and IEP in particular: those who will make decisions which will fundamentally impact the nature of this ecosystem.
In Appendix E of this report will be found charts breaking down the data exchanges of the organizations found in the context diagram into specific environmental domains.

IEP’s role in Bay-Delta Adaptive Management

The scientific community is increasingly called upon to inform policy and management strategies. As an example, a number of scientists from the Bay-Delta community have, over the last few months, been called to consult on the environmental impacts of the Gulf of Mexico oil spill in an effort to bring scientific rigor to an emotional subject.

As competition for the limited natural resources of the Bay-Delta increases, so too does the pressure of special interests from all sides to determine how those resources will be used. Well-accepted science can and must provide the necessary counterweight to these powerful and conflicting voices, and result in policy decisions that balance the needs of all parties most effectively.

To arrive at well-accepted science, nature must be understood not with relation to its individual components only, for each special interest can build effective cases for conflicting policy decisions based on single factors. The science which
informs policy and management must study nature as an integrated whole; it must, in current parlance, be *ecosystem-based*.

It must also be supported by a framework that permits intelligent and coordinated iteration. Every experiment and monitoring effort contributes knowledge which -- once digested -- yields deeper understandings which will, in turn, add precision to the *next* generation of experiments and monitoring procedures. This knowledge and the iterative understandings that it yields will also fuel policy and management decisions, which in turn propose fresh problems and objectives for science to address. The interdependencies of science, policy-making and management, and the iterative nature of the process, are well expressed in an Adaptive Management (AM) Framework, such as the one below:\(^1\):

This process flow highlights the interdependencies of different types of data consumers and contributors, as well as of different professional disciplines that must interact based on the best information available. The flow of data/information from one step is a critical success factor for the next step. In the absence of data/information, assumptions must be made and the quality of the process suffers as a result. Bay-Delta data preservation and management should be capable of supporting the entire range of data requirements within the AM framework, including the knowledge base, status of projects and management activities, raw scientific research data and findings, and the on-going data collection from the conservation measures and monitoring programs.

The implication behind an adaptive management approach to science is that inquiry is not for science’s sake alone, but also to inform policy and management decisions. In the diagram above, policy and management involvement enters primarily in the upper left half of the diagram, with particular emphasis on defining the problem and establishing the goals and objectives which will respond to that problem (boxes 1 and 2). What we tend to think of as “science”, the formulating of experiments and the research and monitoring efforts that follow, take place in boxes 3 to 7. The subsequent data is collected and managed in box 8. See
Appendix B: Adaptive Management Process Inputs and Outputs for details about the data flows between the various AMF process steps.

Data collection and management is currently the weakest link in Bay-Delta AM. As can be seen, effective management is necessary for subsequent analysis, synthesis and evaluation of knowledge (box 9). A common observation in ecology-based management environments (and strongly commented upon in the context of the Bay-Delta) is that the value of research and monitoring data is ultimately determined by how successfully it is *used*. As a recent report on the status of Bay-Delta conservation planning noted: “Monitoring adds no knowledge without a dedicated process for data management and analysis”.  

The adaptive management framework illustrates that the effective preservation of data is important not only for itself, but is also a critical success factor for ongoing utilization of that data for practical use. Without effective and flexible preservation, data cannot be efficiently presented and thereby leveraged for those critical purposes which it is designed to serve. In the case of Bay-Delta science, effective data preservation and management is the key to providing credible scientific opinion to policy-makers and managers, through an enhanced capacity to access, analyze, synthesize, evaluate and assimilate that data.

**Information Customers of IEP**

Customers of IEP data and information products are, as illustrated in the context diagram above, the management and policy arms of individual IEP members, as well as partner agencies, legislators, resource managers, the larger scientific community (agency and academic scientists and consultants), and the general public.

Of these stakeholders, the Delta Science Program (DSP), as the successor to the CALFED Science Program, is emerging as the most direct IEP customer and can be studied as a representative primary customer.

In 2008, CALFED stated a number of goals central to the concept of adaptive management and supportive of the need for effective data preservation and management; among them:

- Integrating the program more fully into statewide and national networks of information sharing and instrumentation to support ecosystem management

---


³ See Error! Reference source not found. for details about the complex relationships to IEP, their impacts on change and how they should be managed during change efforts.

⁴ The State of Bay-Delta Science, 2008
• Increasing the capacity for cross-disciplinary modeling of ecosystem behavior, and positioning the program to serve as a node or catalyst for the development of integrative models.

• Strengthening multi-agency and multidisciplinary integration of Bay-Delta science

• Addressing the weaknesses of existing databases and monitoring exposed by recent research projects (Pelagic Organism Decline, in particular)

• Encouraging and strengthening the integration of disciplines and the integration of science into management

These goals have been inherited by the IEP and Delta Science Program (DSP). DSP sees itself as a “boundary organization”, at the interface of science and policy. It is careful to define its role not as making management recommendations, but rather as stating management implications if certain actions are taken or not taken. They inform themselves of the upcoming information needs of policy-makers, and proactively support those needs by assimilating germane scientific information and translating it into language that policy-makers can understand and assimilate.

In terms of the Adaptive Management framework, DSP operates in boxes 9 (analyze, synthesize, evaluate) and 10 (assimilate and recommend), but might more precisely be said to move between 9 and 10\(^5\). It is their goal to become more effective in box 9, and to strengthen their relationship with IEP in that endeavor. Since they also make use of consultants and researchers not affiliated with the IEP, it would be of additional benefit to house the data/information issuing from those efforts in an appropriate location, from which they could effectively interact with IEP data.

From this perspective, they share IEP’s need for effective data management and increased agility, speed and capacity to utilize modeling tools in the processes of analysis, synthesis and evaluation. When asked what one thing would most improve the Delta environment, the DSP Lead Scientist responded, “more global understanding, therefore more data access”.

An additional need of DSP in its “boundary” role is increased capacity to utilize existing presentation tools that can visually communicate complex information so that it can be quickly grasped. In their efforts to translate scientific understanding into the “language of policy”, presentations must be customized to effectively engage the attention and nourish the understanding of a range of stakeholders. The tools required to accomplish this must be flexible and expressive, and ideally take advantage of semantic analytics and discovery.

---

\(^5\) DSP strategic objectives embrace other activities within the adaptive management framework, including supporting research and facilitating independent peer review
The Delta Science Program, in turn, serves a number of customers, none as significant as the Delta Stewardship Council (DSC). Water use is an extremely contentious issue in California, and the stakes are very high. The Delta is not only the largest estuary on the West Coast, but also the hub for the State’s massive water distribution system. Tremendous pressure is directed at decision-makers responsible for the proper allocation and management of resources.

The DSC was created by legislation in 2009 to develop and adopt a comprehensive management plan for the Delta (“Delta Plan”) by January 1, 2012, and depends heavily on DSP for the scientific underpinnings for the development of this plan.

In particular, the DSP will contribute to the Delta Plan by pointing to “best available” science. While the definition of “best” is quite controversial, “available” is usually assumed as a given. Because the availability of data and information is a critical first step in the application and evaluation of “best science”, this assumption needs to be challenged, and, as we will see in the next section, it is in fact an inappropriate assumption under which to be operating.

IEP Data Preservation and Management Requirements

As stated above, ecosystems are extremely complex and cannot be encompassed by a single mind or mastery of a single discipline. In order to use data to continually deepen our understandings of ecosystems, that data must be accessible, trustworthy and interoperable.

Access:
A tremendous amount of Bay-Delta scientific and environmental data exists, but there is no overview of what precisely comprises this data, or of where it is stored, or of where the critical gaps in knowledge lie. Scientific research and monitoring activities – within CNRA and elsewhere – are not highly coordinated, nor is the storage of the resulting data. Any department, conservancy or commission within CNRA might have its own data collection protocols and customized databases, as might other State and federal entities. Universities and private individuals also undertake research and monitoring projects.

A scientist, then, tasked with providing thoughtful scientific insight to a pressing policy or management question, operates under the significant disadvantage of potentially knowing very little about germane work that has already been done, and must allocate a significant percentage of his/her time to finding what might be “out there”.

Effective preservation of data implies creation of single point of entry, through which IEP and other Bay-Delta scientists might have a thorough view of all existing data that relates to their mandate. This includes datasets from all agencies, State and Federal, as well as from regional entities, universities and private individuals. The ability to see the full range of salient data would render
more obvious where gaps in knowledge lie, and permit more rapid and complete assimilation of existing knowledge before initiating new projects. It would also allow IEP to comply with its mandate to make data publically accessible.

For IEP to effectively perform its mission it must be able not only to know that a particular dataset exists, but also to be able to locate it, and, wherever possible, to directly link to the data site. Existing datasets that are not accessible electronically would at some point be assessed for value, and high value datasets would be transformed into electronically retrievable forms.

In addition to raw data, IEP also needs access to all components of the Knowledge Base, including published and unpublished papers, technical reports, articles, expert presentations and comments, and status and product tracking for IEP-funded and related projects.

Trust
Once the universe of Bay-Delta datasets is known and rendered accessible, it is still not possible, due to lack of agreed-upon standards of metadata, quality, scale and semantics, to confidently determine the values contained within much of that data.

Is “ammonia” in this dataset the same as “NH4” in that one? Upon what detection limit is “undetected amount of contaminant” based? Were the sampling techniques, methods and timeframes functionally similar for these two apparently identical datasets? What other datasets are needed for context before making conclusions based on this one in isolation? What levels of QA/QC were employed? In the absence of standards agreed upon and adhered to in these areas, a large percentage of Bay-Delta scientific and environmental data cannot be considered trustworthy.

Use
Once all of the above conditions are met -- knowing about the data, finding it, and confidently understanding its meaning and context -- the individual scientist can analyze and synthesize that data into meaningful information. But unless the data is also interoperable, the scientist is still limited in his/her ability to effectively manipulate and communicate it.

Interoperability in this context might be defined as the ability to automatically compare and interpret the information exchanged meaningfully and accurately in order to produce useful results as defined by the end users of all systems involved. To achieve interoperability, all sides must defer to a common information exchange reference model so that the content of the information exchange requests are unambiguously defined: what is sent is the same as what is understood.

Once such a common reference model is in operation; that is, once the body of knowledge in a particular area is known and semantic content, transfer formats, metadata and quality/scale definitions are in place, various data analysis,
synthesis and modeling tools can be brought into use. Such tools provide results that are transparent, predictive, replicable and, if data is sufficient, credible.

A variety of presentation tools can also be brought into play for preparing data displays and presentations, both to the scientific community and to those who subsequently determine policy and management strategies. Further, normalized data will permit data mapping, essential for presentation and study purposes for all stakeholders in the Sacramento-San Joaquin Delta, including the general public.

### IEP Critical Success Factors related to data =

- Cooperation between entities and databases

**Access:**

- Comprehensive catalogue of existing data
- Data management system that provides a comprehensive overview of salient data and information
- Access to data of contributors which is isolated, hidden and/or not in usable electronic form

**Trust:**

- Agreement on and adoption of standards related to
  - Methods and techniques for analysis and synthesis
  - Mandated timeframe for sampling and analysis
  - Metadata for QA/QC levels and standards
  - Metadata for detection limits
  - Metadata for other datasets needed for contextual understanding
- Enforceable and enforced data access rules

**Interoperability:**

- Master Data Definitions for semantic harmonization
- Exportability of data to researchers and scientists for analysis, synthesis, and evaluation
- Data in forms that can be analyzed and synthesized.
- Capacity to utilize conceptual or simulation models to predict outcomes of conservation actions
- Services and tools to aid in managing the entire data process

### DSP and DSC Critical Success Factors depending upon data =

- “Best available science” at the highest level possible
- “Best available science” responsive to current policy needs and timetables
- Clear and effective presentation of complex scientific concepts and findings
3. Current Preservation and Management of “IEP Data”

With so much resting on the capacity of IEP to manage its data and to transform it into meaningful and credible information, what is the current state of data within the Program, and what does the future state look like?

Although IEP staff makes use of salient scientific and environmental data whatever its origin, there does exist a concept of “IEP data”: long-term monitoring data (biological, water quality and hydrological) collected by DWR and DFG under the auspices of the Program, as well as research data from internally funded projects such as POD. A DWR-managed database, Bay Delta and Tributaries (BDAT) was designed to house these datasets.

This data can be, and is, used to good effect. At the 2010 IEP Annual Workshop, for example, a convincing demonstration was made of how certain biological communities respond to ocean climate patterns in the San Francisco Bay, using some 40 years of accumulated IEP data. Shrimp, crab, clam and fish datasets from BDAT were extracted and superimposed on graphs based on federal water temperature data. Clear patterns emerged of how population changes in one species related to both population changes in other species and to rising and falling water temperatures.

Bay Delta and Tributaries Database

BDAT was envisioned to be the data repository for the IEP. Funded to a large extent by the IEP and managed by the Interagency Information Systems Services (IISS) group within DWR, the concept behind BDAT was to gather all IEP-related data in one space. In accordance with then-current thinking and technologies, the structure of BDAT was integrated rather than federated in structure, its only web function being to search for specific data or subsets for download as flat files or spreadsheets for the consumer to process. Additionally, it contains summaries of studies and projects that provide some guidance as to what is available. There was the intention to include metadata, but this did not come to full realization. A geographically based web interface was also never realized.

The system was built on a foundation of local DWR water flow and quality data. Efforts to encompass heterogeneous datasets (for example relating to fish, benthic invertebrates, bathymetry, etc.) within this homogeneous data structure resulted in significant data quality issues. Some data-types required intensive manipulation for input, with the result that the output looked different than when it went in, and with some links missing; it therefore became suspect even to those
consumers who initially contributed the data. The current state of much of the
data within BDAT is not considered reliable.

Because of these factors, and also because the IISS group is not sufficiently
resourced to provide technical support, many previous contributors to the
database no longer provide data to it. Home-grown local databases have
proliferated, with no centralized overview of available datasets.

Salient Bay-Delta data, then, is difficult to access through BDAT and not
considered consistently trustworthy. It is also not structured for interoperability.
Due largely to resource constraints, spatial or temporal disparities of data are
entered as captured and cannot be harmonized for overarching in-depth
analysis. Adjustments are not made to account for inconsistent sampling
methods or inconsistencies in time series monitoring.

Further, rules for data formats and metadata elements of contributed datasets
are not adhered to, with the result that even when DWR itself is sponsoring a
monitoring or research effort, the DWR-based IISS group does not always feel
empowered to impose requirements on how the data is delivered. The lack of
harmony at the point of data collection and provision contributes to poor
integration of the data within the system. IEP, which “owns” the data, does not
currently have the capacity to exercise any control over it.

Weaknesses in the existing system can be observed in the above-mentioned
DFG study related to temperature effects on San Francisco Bay aquatic
communities. The study provides clear value for understanding the past, but is of
little value for projecting the future or for informing management or policy
decisions. In this case, it did not matter that shrimp, clam, crab and fish data
arose from different protocols on different scales or that they were stored in
different databases. Manually overlaid on scale-adjusted graphs they tell a clear
and compelling story. As they are currently preserved, however, they cannot be
easily adapted for use in modeling tools.

Department of Water Resources and “Data Web”

The DWR IISS staff is currently engaged in replacing BDAT. The new system,
referred to by its developers as the DWR Data Web, very directly responds to
many of the perceived shortcomings of its predecessor. It will take a metadata
(mash up) approach rather than a monolithic one, using a web interface to find
the data. Working methodically from within DWR, it will focus on local scientific
data and be slow to accept other datasets for inclusion. The concept for the
system is shown in Appendix F: Repository Architecture for DWR.

As currently envisioned, Data Web will provide preservation for only a subset of
what needs to be preserved in order for IEP to meet its mandates. In its reticence
to store a wider range of Bay-Delta data, its current direction is to simply point to
where that data can be found. If this direction is actualized, IEP staff will be little
better off in terms of time spent accessing needed data, and with no gain as
regards working with fully interoperable data.
If the Data Web mandate were widened to serve as a single point of entry for IEP data, or, alternately, as the model for a number of linked databases to jointly comprise IEP data, IEP critical success factors would more likely be met.

**Department of Fish and Game**

DFG staff assigned to the Bay-Delta ecosystem is tasked with tracking and documenting species trend data and how species are impacted by water exports. Since the exporting of water from the Delta is a highly charged subject, solid scientific evidence of impacts caused by this practice are critical for effective policy and management decisions.

DFG has found that the BDAT structure does not support its data needs well due to lack of responsiveness in system support, lack of internal skill sets to make desired changes and increasing distrust in the integrity of previously-contributed data, and has reverted to the use of personal ad hoc databases. Some DFG scientists use BDAT entry screens, but route the data to their personally constructed databases, avoiding direct interaction with BDAT. There is currently no mechanism for staff to know of, or be able to directly access, the work of other scientists, even within the same DFG field office.

**Other Data Contributors of the IEP**

The memorandum of understanding (MOU) signed by each partner as a condition of membership in the IEP does not include any requirements related to data management. It is not, therefore, surprising that DFG and – currently – DWR do not manage their data with a strong sense of responsibility toward the IEP as a customer. This is also true of the other contributing partners. As shown in the context diagram, three federal IEP partners also collect and monitor data which contributes to the aggregate of Bay-Delta scientific data:

- **US Bureau of Reclamation (USBR) datasets:** pelagic invertebrates, sediment, phytoplankton, benthic invertebrates, fisheries, atmosphere and water quality
- **US Fish and Wildlife Service (USFWS) datasets:** water quality and fisheries
- **US Geological Survey (USGS) datasets:** water quality, atmosphere, benthic invertebrates, bathymetry and contaminants

Although these datasets contribute substantially to the current understanding of the Bay-Delta ecosystem, they are not always conceived of as “IEP data”. IEP scientists, like all other data consumers, go to the federal websites to access data.

Because the MOU that binds members to the IEP has no data management requirements, there is no operative sense of the IEP as a customer of that data.
There is no demand upon these agencies to become responsive to IEP needs, nor has the IEP articulated these needs. There are, for example, no metadata detection limit standards for sampling methods, meaning that scientific analysts have difficulty in determining whether data from USBR, USFWS or USGS matches (“apples to apples”) with similar-looking data from DWR or DFG.

The other four members of the IEP do not contribute data. Again, since the MOU does not speak to data, there is currently no mechanism in place that guarantees sharing of data sets collected by IEP members with others within the Program.

Sponsored Studies

IEP supports a number of research studies, conducted by staff members of members of partner entities, to gain scientific knowledge in areas of particular concern in the Delta. Some of the data and information from these studies are managed in BDAT.

Since BDAT lacks the capacity for modeling-tool-enabled analysis, IEP studies must export the data elsewhere if such analysis is desired; and since the standards for data within BDAT are not clearly defined, exporting of salient data must be done manually. In the last two years, for example, IEP-sponsored projects at the National Center for Ecological Analysis and Synthesis (NCEAS) required manual annotation, transformation, and exporting of data that had been generated within IEP and housed in BDAT.

DSP-sponsored studies are done on a contractual basis, with universities or individuals. DSP can determine how and where study results are housed, but since there is currently no satisfactory data management structure, researchers currently are left to determine their own data standards and locations for the raw data and QA/QC data. There is currently no centralized comprehensive index of data and information derived from these studies.

Other Sources

A great deal of data beyond that contributed by IEP members is required for a full understanding of this complex and highly-stressed ecosystem. An individual’s data-accessing success depends on two key strengths: data searching savvy and personal connections. Even with these skills highly developed, there is significant risk that, for any given study, not all salient datasets will be found. Many important data sets, especially belonging to independent researchers, remain underutilized and are lost over time. Some of these datasets were at one time accepted into BDAT. This may, indeed, have contributed to the failure of BDAT due to the heterogeneous nature of the accepted data; the fact remains, however, that lacking a satisfactory preservation strategy, much of this data will eventually be lost.

It should also be noted that new data searching skills and personal relationships will need to be established if, as is possible in the very near future, the
geographical boundaries of the IEP mandate are expanded. Data from entities not currently consulted (for example, the Bureau of Land Management, the U.S. Forest Service, the National Park Service, other divisions of DFG, etc.) will need to be incorporated if IEP is to bring adjacent wetlands, the Pacific Ocean and upstream riparian habitats into its sphere of study.

<table>
<thead>
<tr>
<th><strong>IEP Critical Success Factors</strong></th>
<th><strong>Current State</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperation between entities and databases</td>
<td>Poorly developed</td>
</tr>
</tbody>
</table>

**Access:**
- Comprehensive catalogue of existing data
- Data management system that provides a comprehensive overview of salient data and information
- Access to data of contributors which is isolated, hidden and/or not in usable electronic form

**Trust:**
- Agreement on and adoption of standards related to
  - Methods and techniques for analysis and synthesis
  - Mandated timeframe for sampling and analysis
  - Metadata for QA/QC levels and standards
  - Metadata for detection limits
  - Metadata for other datasets needed for context
- Enforceable and enforced data access rules

**Interoperability:**
- Master Data Definitions for semantic harmonization
- Exportability of data to researchers and scientists for analysis, synthesis, and evaluation
- Data in forms that can be analyzed and synthesized
- Capacity to utilize conceptual or simulation models to predict outcomes of conservation actions
- Services and tools to aid in managing the entire data process

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not exist</td>
<td>Is not currently possible</td>
</tr>
<tr>
<td>Does not exist</td>
<td></td>
</tr>
<tr>
<td>No agreement or implementation</td>
<td></td>
</tr>
<tr>
<td>No agreement or implementation</td>
<td></td>
</tr>
<tr>
<td>No agreement or implementation</td>
<td></td>
</tr>
<tr>
<td>No agreement or implementation</td>
<td></td>
</tr>
<tr>
<td>Not fully realized</td>
<td></td>
</tr>
<tr>
<td>Not fully realized</td>
<td></td>
</tr>
<tr>
<td>Currently possible only via manual effort and only then for some datasets</td>
<td></td>
</tr>
<tr>
<td>Do not exist</td>
<td></td>
</tr>
</tbody>
</table>
4. Strategic Technical Architecture

4.1 Digital Repository “Service Ecology”

The concept of preserving data through federating discrete databases in a Digital Repository (DR) is increasingly understood as an effective means of making related yet heterogeneous datasets available, while avoiding the pitfalls that inevitably arise when attempting to force such datasets into a single database. The experience of BDAT (shared by IT developers world-wide over the last two decades) shows that coercing a large amount of heterogeneous matter into a single database does not lead to effective data management.

A federated repository provides participants with the key benefits of a single centralized database -- such as a single point of entry, searchability and access to system data -- while significantly expanding the amount of data available. In contrast to a more traditional centralized database, however, it also permits the individual contributing databases to maintain control of their internal data management and governance, including decisions regarding which datasets to include and what standards will apply to them. In fact, the first requirement for improving data access in the DR is for each entity within the system to take responsibility for the management of its own data.

What then is a “Service Ecology”?  

- Service ecology is a system of actors and the relationships between them that form a service. The service ecology takes a systemic view of the service and the context it will operate in.  

- Service ecologies include all actors affected by a service, not only those directly involved in production or use. By analyzing service ecologies, it is possible to reveal opportunities for new actors to join the ecology and new relationships between them.

- Ultimately, sustainable service ecologies depend on a balance where the actors involved exchange value in ways that is mutually beneficial over time.

---


8 Designing Product/Service Systems: A Methodological Exploration, Nicola Morelli in Design Issues: Volume 18, Number 3 Summer 2002
A Digital Repository Service Ecology (DRSE) then is much broader concept than a single federated repository. It consists of networks of actors (participants) bringing their repositories together to create a whole that is greater than its parts.

When a digital repository is able to respond to more dynamic consumer needs -- for example, enabling them to share information about on-going or proposed collection efforts and ensuring that datasets are interoperable -- we can begin to speak of a service ecology.

An information network of this type, as happens in an ecological system in nature, establishes relationships between multiple interacting “organisms” in such a way that their connections and dependencies can be recognized and acted upon. Such a system does much more than bring data to a consumer: it facilitates the consumer’s efforts to transform that data into information, to make it actionable and meaningful.

Technical solutions to implementing network-based systems are well-developed, particularly in the field of data preservation. Efforts such as the Washington State Digital Archives Project (http://www.digitalarchives.wa.gov/) have demonstrated the capacity to preserve and provide access to records of enduring significance through a central repository. Records are captured in a variety of formats and metadata tags are applied; these records are then searchable through a web interface with a consistent presentation.

Implementers of digital repository service ecologies are faced with significant additional challenges in planning and managing their service in relation to the “ambient” information environment (see Appendix D: Service Pattern “Provide Advisory Encounters” for sample services that a DRSE would provide and appropriate metrics to apply to those services). They face particular challenges when:

- Articulating needs
- Identifying opportunities
- Expressing complexity
- Managing development⁹

**DRSE supporting science**

Digital repositories and the supporting presentation services make data accessible and searchable; they should also provide enough information about data sources and content to minimize misunderstanding and misinterpretation. By effectively providing this raw data, or providing pointers through which consumers can access it, the repository enables scientists to begin the

---

⁹ Robertson R. J., 2007
necessary work of scientific analysis. But how can data become interoperable, so that modeling tools can be brought into play?

With the emergence of holistic and semantic data integration technologies, not only structured, but also semi-structured (Web) and unstructured data sources can now be integrated. This capacity to integrate (mash-up) data is optimized in lightweight and flexible architectures [which] extract data from heterogeneous sources, build semantic relationships across them, and provide real-time access to the composite virtual data services\(^\text{10}\).

A digital network which is able to “leverage the immediacy and relevance of unstructured and Web data to enrich enterprise applications and processes”\(^\text{11}\) can be said to provide the technical elements of a service ecology. In the case of science, this means a greatly enhanced capacity to provide solid and supportable input to policy-makers and managers, and to maintain the flexibility demanded by the iterative, adaptive nature of ecology-based science/policy/management.

Limiting factors for the IEP and the Bay-Delta science community to perform their work effectively are access to data, access to data tools and coordination of data standards. A well formulated and implemented Digital Repository Service Ecology will meet these needs. Such a DRSE must be capable of supporting the entire range of data requirements of the AM framework, including the knowledge base, status of project and management activities, scientific raw data and findings, and on-going data collection from conservation measures and monitoring programs.

4.2 Interoperability of Data

Interoperability of information is a key principle in the information world today, from inter-dependent businesses, to multiple healthcare environments, and to disparate environmental repositories. What makes data interoperable? Interoperability is a three legged tripod; missing one of the legs, it cannot stand. The three legs of interoperability are “metadata, “master data,” and “data integration.” Each of these elements has its own discipline, body of knowledge, practices and procedures. Only when these disciplines are aligned and working together, will the information provided to customers be truly interoperable.

Metadata

Metadata encompasses all of the data and information that is needed by an organization to effectively and efficiently manage its data and information

---

\(^\text{10}\) Chandrasekaran, 2010

\(^\text{11}\) Ibid.
resources as well as to communicate to external parties what data the organization possesses for sharing. The four primary purposes of metadata are classification, description, guidance and control:

- Classification groups data by similarity of characteristics
- Description defines data meanings and documents properties of data
- Guidance helps people to find the data that they need
- Control enforces data constraints including rules, regulations, and service levels.

Over time, metadata standards have been developed to facilitate the sharing of information between organizations and to make it possible to search and combine information for review and analysis. An organization which bases its metadata on a standard has several key advantages. These include a strong foundation, for which much preparatory work has already been performed, and streamlined processes for both publishing and retrieving data.

A standard must also be extensible, as every industry and data domain will have its unique requirements and constraints that must be accounted for. The following are some common metadata standards.

- Machine Readable Cataloging (MARC)
  - A library sciences standard for structure & content designation
- Dublin Core
  - An ISO standard for cross-domain information resource description
- Resource Description Framework (RDF)
  - XML-based language for describing web resources
- Common Warehouse Metamodel (CWM)
  - The Object Management Group standard for data warehouse metadata encoding and exchange
- Metadata Object Description Schema (MODS)
  - XML-based description format less complex than MARC but more complex than Dublin Core
- Metadata Encoding and Transmission Standard (METS)
  - Complex XML-based metadata format for describing and exchanging media digital objects
- Content Standard for Digital Geospatial Metadata (CSDGM)
  - A US Government standard for description of geographic and geospatial data

**Master Data**

Master Data refers to the non-transactional data entities of an organization or industry that support transactional activities and are used by many groups and processes throughout the organization. For businesses it is customers, products, parts, services, suppliers, employees, accounts, etc. In the healthcare world it is patients, providers, services, drugs, diagnosis, allergies, etc. In the
environmental world it includes locations, waterbodies, contaminants, constituents, species, sample medium and methods, etc.

The value of effective Master Data Management (MDM) is that the repositories which contain the data are not required to adapt to some standard, usually from an external source. These repositories can continue to function as currently, and when a view across multiple repositories is required, the MDM will provide the ability to see the data as one source.

What cannot be overstated is the commitment of time and resources necessary to create and manage the Master Data by each organization wishing to participate in the larger world of sharing information. MDM implementation is a multi-step process that involves people, processes, systems, & technology.

Data Integration
Data Integration is the combining of heterogeneous data that resides in multiple and disparate sources, resolving conflicts and inconsistencies among that data, and providing a unified view to consumers. Data to be integrated comes from:

- different sources
- different users
- different meanings
- different technologies
- different structures
- different semantics & models

Data integration is accomplished by a series of steps taken by the system or applications that will integrate the data. The diagram below illustrates the following steps:

1. Map
   - Identify connections among sources. The integrator often accomplishes this by access to a metadata catalogue that will identify various sources of data and include the metadata necessary to access the data

2. Collect
   - Gather disparate data sources. Based on the metadata, the integrator can collect and correctly align data elements to the information request

3. Unify
   - Consolidate disparate data. The Master Data enables the integrator to align different ontologies and descriptors to a common language and identifiers enabling true comparison from multiple sources

4. Verify
   - Check for process errors & data defects.

5. Deliver
   - Publish or distribute consolidated data to the requestor.
In the above diagram, multiple information management tools work together in alignment to deliver harmonized information to a consumer as described in the data integration steps above.

4.3 Expanding the Network

Environmental DRSEs already exist and could serve not only as models for an emerging DRSE in the Bay-Delta, but could also act in partnership with it to expand domains and ranges of data.

The CUAHSI Hydrologic Information System (HIS) is an internet-based system for sharing hydrologic data, and has already established relationships with USGS, EPA and NOAA. It is comprised of databases and servers, connected...
through web services to client applications, allowing for the publication, discovery and access of data.

CUAHSI-HIS could support a Bay-Delta DRSE through

- Supporting data consumers in their efforts to find data, retrieve data, organize data, analyze data and/or use data for modeling
- Helping data publishers organize data, store data, and make data available to others
- Encouraging tool and technique developers create and share additional tools and techniques that are compatible with the CUAHSI-HIS system

Another partner of CUAHSI-HIS is the Long Term Ecological Research Network (LTER). LTER is comprised of 26 nodes, each one of which is committed at the agency level to system requirements (including a stipulation for a full-time data manager) and which is supported by the network office. It could also be solicited to participate in sharing environmental data, information and tools.

When sketching possible solutions, and again when formally designing the system, the response to each IEP CSF will be developed through the following functions.

**IEP Critical Success Factors =**

- Cooperation between entities and databases

<table>
<thead>
<tr>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata definitions</td>
</tr>
<tr>
<td>Master Data Definitions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Access:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive catalogue of existing data</td>
</tr>
<tr>
<td>Data management system that provides a comprehensive overview of salient data and information</td>
</tr>
<tr>
<td>Access to data of contributors which is isolated, hidden and/or not in usable electronic form</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expanding network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Integration</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metadata Catalogue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
### Trust:
- Agreement on and adoption of standards related to:
  - Methods and techniques for analysis and synthesis
  - Mandated timeframe for sampling and analysis
  - Metadata for QA/QC levels and standards
  - Metadata for detection limits
  - Metadata for other datasets needed for context
- Enforceable and enforced data access rules

<table>
<thead>
<tr>
<th>Metadata / Master Data Definitions</th>
<th>Metadata / Data Integration</th>
</tr>
</thead>
</table>

### Interoperability:
- Master Data Definitions for semantic harmonization
- Exportability of data to researchers and scientists for analysis, synthesis, and evaluation
- Data in forms that can be analyzed and synthesized
- Capacity to utilize conceptual or simulation models to predict outcomes of conservation actions
- Services and tools to aid in managing the entire data process

<table>
<thead>
<tr>
<th>Master Data Data Integration, etc. Metadata Metadata / Master Data Definitions Data Integration</th>
<th>Master Data Data Integration, etc. Metadata Metadata / Master Data Definitions Data Integration</th>
</tr>
</thead>
</table>
5. Attitudes toward Change

In order for IEP to meet its mandate to provide meaningful and timely scientific data and information to Bay-Delta management/policy-makers, it must be part of a data preservation/management system much like that described in the previous section. Such a Digital Repository Service Ecology requires the participation of all elements of the Adaptive Management Framework. Is there sufficient willingness to make meaningful changes in data preservation and management among this broad and heterogeneous group?

Previous studies related to engaging participation in digital repositories have used the terms "user", "partner" and "provider" to describe the various participant-types. This current study has substituted for "user" the more narrowly descriptive term consumer, to describe individuals or organizations utilizing repository resources and services. Partner here continues to refer to a database linked to the repository and providing data of agreed-upon standards. A provider is an entity which provides over-arching infrastructure for the system.

During the course of this study, it was found that the role which an entity plays – that of consumer, partner or provider – is a greater determinant of willingness and ability to change than any other factor. It was further found that yet another participant-type exists, with its own distinct concerns and challenges, that of contributor. This role describes one whose research or monitoring activities produce results that are of interest to the whole, even if that individual is not him/herself interested in the larger context.

In very broad terms, the right half of the Adaptive Management Framework is comprised of contributors. The left half, with the exception of box 8, is comprised of customers of the work done on the right side, that is, of consumers. Box 8 is where partners reside. A provider, if and when one is brought into play, would invisibly harmonize all roles within the framework.

---

12 The Need to Formalize Trust Relationships in Digital Repositories, Berman F, Ardys Kozbial, Robert H. McDonald, and Brian E. C. Schottlaender., 2008
The diagram below overlays the concept of participant-types on the specific AMF roles played by Bay-Delta entities. This view, which lends clarity to the customer relationships within the Bay-Delta AM framework, also underlines the impact of participant-type on the change readiness of each stakeholder. For example, as will be explored below, data contributors within DWR, DFG and IEP tend to have more similar attitudes toward change than do data partners or consumers, even within the same agency.
Consumers

DSC, DSP and those elements within IEP which are involved in synthesis, analysis and evaluation have, as consumers of Bay-Delta data and information, a strong need for all of the benefits which a DRSE would confer. In order for them to meet their mandates, they need a data system which has strong access, trustworthiness and interoperability of data.

This group shows no perceptible resistance to DRSE and is quick to recognize (and many among them are already aware of) the potential benefits. Opposition to a DRSE is unlikely to arise from this participant-type: but how far will they enable the process?

Currently, Bay-Delta data consumers do not behave as customers; that is, they do not place requirements on their direct service provider (data managers, or partners), nor do they feel responsible for ensuring that sufficient resources are allocated to data managers to perform those desired services.
Partners (Data Managers)

If data consumers see nothing but benefit from a DRSE, data partners (and potential partners) see difficulties in providing these benefits. They do not have resistance to the concept of a DRSE, and in general are quite aware of, and appreciative of, the promise in such a system. But there is resistance to moving toward such a change, largely due to previous IT endeavors.

Previous, well-intentioned, efforts to broaden and improve Bay-Delta data preservation and management made, in the experience of this participant-type, two classes of errors: first, there was excessive and inappropriate top-down direction; and second, there was inadequate funding to fully implement -- and then to sustain -- the desired changes.

Individuals with data stewardship roles in the Bay-Delta scientific community feel themselves in a service role, and have a strong wish – an idealism – to provide both consumers and contributors with an excellent system. They tend to be frustrated that outside constraints not only prevent them from providing such excellent service, but in fact place them in a bad light, as though they didn’t care.

These constraints arise partly from how data-improvement initiatives historically tend to begin. There is a strong burst of interest and ideas from an unseen executive level to make IT changes that do not grow from, or take into due consideration, conditions on the ground that data managers struggle with on a daily basis. Such a top-down initiative results in a situation where data managers need to implement changes that they do not feel ownership of. The result is not only ineffective implementation (how can you build what you do not believe in?), but also in a subtle and pervasive need on the part of data managers to “protect” the data environment from executive “meddling”.

Such initiatives also tend to be under-funded, in particular as regards on-going maintenance of implemented systems. This is true not only in the Bay-Delta community, but is a tendency everywhere. As Robertson and Barton have pointed out, the IT “community has reached a point where getting a repository to work is not the problem – the critical factor is finding resources to keep it working.”

For partners, then, a key issue is that they have not been sufficiently involved in the development of, and budgeting for, data systems. For future initiatives to be successful, this issue will need to be addressed; and not only for the project in question, but also to heal deep wounds from the past.

In addition to involving partners (data managers) more thoroughly and respectfully in future initiatives, it will be important to change the paradigm of directives being issued from the top with the more appropriate paradigm of downstream consumers defining needs and supplying adequate resources. That

---

13 Optimising Metadata Workflows in a Distributed Information Environment
is, the building of an effective DRSE must be customer-driven, rather than driven by agency leadership or by data managers/partners.

An important executive role, not hitherto fully performed, would then be to mandate and enforce cooperation by all parties with the rules and standards which the customers define and which the managers incorporate into the system. Partners should neither be enforcers of data standards, nor victims of non-compliance.

**Contributors**

This category includes all those whose research or monitoring data would appropriately either be input into a shared database, or stored individually and linked to a larger network.

As with the partners, the contributors to Bay-Delta data are wary of IT initiatives. In the past, their needs have not been sufficiently considered, nor have promised benefits fully materialized. Primary among these needs are ease of input and query, and control and security of input data.

Input and query needs include:
- Tools and services to upload, manage, review, and transform their data
- Helpful query options
- Technical support
- Input data returns unchanged, in the same form

Control and security needs include:
- Freedom from fear that data will be manipulated, misapplied or misinterpreted -- to their detriment and to the detriment of true science
- Freedom from fear that they will lose control of data that they consider proprietary
- Assurance that their information is given the appropriate respect and protection

Any attempt to build a Bay-Delta DRSE would need to fully involve representative contributors in the process of developing solutions for these needs.

Even without historical factors, however, data contributors will tend to be the most resistant participant-type to building a DRSE. This is due to well-understood human mechanisms of innate conservatism and consequent resistance to any change that comes from outside. This is compounded in the case of Bay-Delta scientific inquiry due to a deep culture of rugged individualism and self-sufficiency.

Scientists who work primarily in the Knowledge Base side of the Adaptive Management Framework tend to depend upon a personal network of interactions for knowing about and accessing salient data; they tend to accept the fact that a large percentage of their time will be allocated to finding this data, and to
manually transforming it into information; they tend to see their legacy in terms of studies they have completed, and not in terms of data that they leave behind for future generations of scientists to utilize. Preserving their legacy in a knowledge base and environmental data repositories can be a strong incentive, if their trading agreements include provisions for ensuring that they are cited when their data is used. This gives the contributors assurance that they will have the proper attribution when data is used, and allow them to know about their data the fundamental questions: who, why, how, and when.

Contributors, then, are not particularly interested in the benefits that a DRSE would provide: they tend to be wary, and their natural conservatism prevents them from being active enabling agents in positive change. At the same time, they do not have the wish or capacity to actively block such an endeavor. They require only to be approached respectfully and to verify over time that their legitimate needs and interests are being met.
6. Transformation Planning and Implementation of a DRSE

Effective change issues from a single powerful vision that is shared with, and is eventually embraced by, the majority of stakeholders affected. In the case of Bay-Delta science, there is an apparent disharmony of interests between consumers, partners and contributors of scientific and environmental data. At base, however, each of these participants know that their ability to meet their individual mandate is compromised by inadequacies in the same existing system of preserving, managing and utilizing data. For positive change to happen, a vision must be shaped that brings into sharp focus the essential unity of need.

Once such a vision has sufficiently coalesced, the process of transformation can begin. We refer to it as “transformation” rather than “building”, because the existing data structure cannot simply be torn down and begun again from scratch. Each participant in the Bay-Delta scientific community must continue to do work while construction is in progress. To determine how most effectively -- and least disruptively -- to manage this complex rebuilding program requires comprehensive transformation planning. Once planning -- and funding -- is in place, the transformation itself can begin and must adhere rigorously to standard project management (PM) procedures.

RTC is recommending a three phase approach to this transformation process: courtship, planning and implementation, with each phase being managed as a discrete project.

**Courtship:**

The first phase is the smallest, yet requires the greatest sensitivity and diplomatic skill. The purpose of this effort is to align the various participant needs into a single vision, and to engage a sufficient number of key participants to play their appropriate roles. Unless and until this can be accomplished, no further efforts to enhance the existing data preservation system can effectively be made.

This project can be viewed as an enhanced Feasibility Study Report (FSR) which aims to engage participation and establish commitment in the very act of determining whether sufficient engagement and commitment is in place. As with any FSR, it would need to demonstrate the business case for change by setting out the reasons for undertaking the project and analyzing its costs and benefits. In contrast to a standard FSR, it would also contain a strong educational element (to help stakeholders understand the possibilities of a DRSE) as well as a significant change management piece (to assess change readiness and address reticence to change). In addition, this project would need to procure funding for the next phases or, at minimum, to locate potential funding sources and prepare the groundwork for securing it.

It might be said that the greatest impediment hitherto to establishing an effective data preservation/management system has been the lack of appropriate
customer involvement. Data consumers have not clearly and consistently stated their needs, nor have they committed to providing sufficient resources to data managers for delivery of those needs. The "courtship" phase needs to both determine whether these consumers will step into their appropriate role and to encourage them to do so.

At the conclusion of the courtship phase, a Go/No-Go decision will need to be made for further action, and it will need to be based on the following factors:

- Do the key consumers acknowledge their role in establishing a customer-driven process?
- Will an appropriate consumer step forward to act as sponsor for the next phase of the transformation project?
- Is there sufficient political will to carry out subsequent phases of this transformation (will other State and Federal participants actively support it)?
- Can the appropriate consumers work together effectively to set appropriate data and governance standards?
- Is it likely that sufficient funding can be established for the next phases?

**Transformation Planning:**

The second phase is a unified planning effort, driven by principles of Enterprise Architecture. Its aim is to establish a roadmap for implementation of a DRSE, showing the most effective and least disruptive path to arrival.

Built around a well-accepted sponsor and competent project manager, a broadly-based and inclusive project team will – following standard PM practices -- determine strategies for arriving at decisions regarding:

- Application and Information Architecture
- Data standards
- Data governance rules
- Role of partners and provider(s)

The project team will perform a detailed analysis of the “As-Is” situation of Bay-Delta data preservation/management, materials for which can be found in the appendices of this report. They will then design the “To-Be” system (elaborated from *Figure 3 - Interoperability Components and Integration Steps*), and identify the gaps between the current and future states. This will serve as the basis for a plan of action, beginning with a comprehensive view of all activities required to bring about the desired changes and resulting in the definition of the scope, schedule and budget necessary for its implementation.

Specific deliverables for this phase include:

- An Application and Information Architecture
• Stakeholder Analysis/Roles
• A Project Charter for Phase Three
• A Project Portfolio Plan for Phase Three, showing phasing of activities
• A Change Management Plan
• Metadata and MDM Standards for implementation of the DRSE
• Templates for Memoranda of Understanding (MOU) and Service Level Agreements (SLA) between participants
• Signed SLA with Provider(s)
• Funding

**Implementation:**
Implementation, the longest and most expensive phase of system transformation, is also the simplest if the first two phases have been properly and thoroughly conducted. Before initiating this implementation effort, it will be critical to commit to and establish a clear and firm project management process, following State guidelines for project management, including effective Communication and Risk Management Plans.

It will also be critical to ensure that a solid organizational and governance structure is in place, as well as sufficient on-going funding, to ensure that the system, once implemented, is sustained.

**Caveat: a critical risk**
FSRs are well-accepted as a preliminary to implementing IT systems, and therefore Phase One has a good chance of receiving adequate funding and support. Phase Three -- implementation of a well-designed and planned project -- is likely to attract strong support and funding if return on investment has been clearly demonstrated. Phase Two, however -- transformation planning -- is less likely to receive enthusiastic support and therefore runs a serious risk of being under-resourced.

Planning requires more time and money than people initially see the need for. Organizations, like individuals, are typically reacting to pressures around them, and find it hard to allocate sufficient time to strategic work of any kind. Organizations also tend to measure themselves in terms of output, which may or may not be effective output. They also tend to fund activities which have marketing value, that is, activities that can be seen and admired from the outside. Planning is not the response to an acute pressure, it is not a direct output, and it does not offer marketing value.

What Bay-Delta science so desperately needs (as is the case in so many organizations) is foundational, unseen by the outside eye and unglamorous. It needs to preserve and manage its data effectively, and it needs to perform the transition to an effective system with the utmost tact and agility. Careful and
inclusive planning is essential for this to happen. Any weakness in the planning process can lead to serious issues in the completed system, compromising the Bay-Delta science community’s mandate, with potentially grave consequences to all species affected by its ecosystem, including humans. It is imperative for full success of this endeavor that the Phase One final deliverable insist upon adequate funding for Phase Two; and that Phase Three not begin without assurances that comprehensive EA planning has been done.