CONCEPTUAL STUDY OF INTELLIGENT ARCHIVES OF THE FUTURE

August 23, 2002

A Report Prepared By:

H. K. Ramapriyan, Gail McConaughy, Christopher Lynnes, Steve Kempler, Ken McDonald

NASA Goddard Space Flight Center
Code 423, Greenbelt, MD 20771

Bob Harberts, Larry Roelofs, Paul Baker

Global Science & Technology, Inc.

Ramapriyan@gsfc.nasa.gov

tel 301-614-5356
fax 301-614-5267

TABLE OF CONTENTS

FIGURES

EXECUTIVE SUMMARY

Sponsored by NASA's Intelligent Systems Project within the Computing, Information and Communication Technology (CICT) Program, a conceptual architecture study is under way to address the problem of getting the most societal value from the large volumes of Earth and space science data that NASA expects to accumulate in the future. The particular focus of this study is on the next generation of archiving systems, the "Intelligent Archive," and its role within an end-to-end context. Beyond the obvious needs for more efficient storage and access to data that should be mostly met by improvements in hardware technologies, advances are needed in concepts and tools to enable intelligent data understanding. Large and ever-growing quantities of data pose a challenge to basic data management and utilization. However, there are several other challenges that need to be considered:

- Data acquisition and accumulation rates tend to outpace the ability to access and analyze them
- The variety of data implies a heterogeneous and distributed set of data providers that serve a diverse, distributed community of users
- Unassisted human-based manipulation of vast quantities of archived data for discovery purposes is difficult and potentially costly
- If NASA is to migrate its technologies to operational agencies' decision support systems, it is necessary to demonstrate the feasibility of near-real-time utilization of vast quantities of data and the derived information and knowledge
- The types of data access and usage in future years are difficult to anticipate and will vary depending on the particular research or application environment, its supporting data sources, and its heritage system infrastructure

The objective of the study is to formulate ideas and concepts and to provide recommendations that lead to research by the computer science community in the near-term, prototyping to demonstrate feasibility in the mid-term, and operational implementation in the period from 2012 to 2025. The approach consists of the definition of future usage scenarios and needs for data usage in applications, projection of advances in technologies, and an abstraction of an intelligent archive architecture. The abstracted architecture is defined without regard to physical implementation and is considered from the point of view of the functions that need to exist in support of usage scenarios. This is because the functions of an intelligent archive are more stable than the physical architectures and technologies used to implement them. By abstracting entities and processes from scenarios into functional elements, we can explore application strategies of technologies and system resources for future intelligent archives.

An Intelligent Archive (IA) includes all items stored to support "end-to-end" research and applications scenarios. The stored items could be anywhere in a highly distributed end-to-end system, including on-board caches with the sensors and those in the client systems with the users. Stored items include:

- Data, information, and knowledge
- Software needed to manage holdings

- Interfaces to algorithms and physical resources to support acquisition of data and their transformation into information and knowledge, storing the protocols to interact with other facilities

The phrase IA generalizes the term archive from a simple repository of data to one that supports and facilitates derivation of information and knowledge. An IA has greater ability to operate more autonomously than conventional archives and provides better service to users (as an intelligent assistant) with less operator intervention and, hence, with lower cost.

An IA offers new capabilities that distinguish it from archives of today. Some examples of such capabilities are:

- Storing and managing full representations of data, information, and knowledge
- Building intelligence about transformations on data, information, knowledge, and accompanying services involved in a scientific enterprise
- Performing self-analysis to enrich metadata that adds value to the archive's holdings
- Performing change detection to develop trending information
- Interacting as a cooperative node in a "web" of other systems to perform knowledge building (where knowledge building involves the transformations from data to information to knowledge) instead of just data pipelining
- Being aware of other nodes in the knowledge building system (participating in open systems interfaces and protocols for virtualization, and collaborative interoperability)

Many variations and permutations of operational scenarios are possible given the architecture model of elements, component objects, and associations for an intelligent knowledge building system. The conceptual architecture establishes a framework for how future intelligent archives operate in a context of cooperating systems and infrastructures. Furthermore, the architecture is a necessary guide where opportunities for integrating intelligent systems, data understanding, and machine learning can be identified and mapped to goals for automation that increases effective data utilization.

This study of the IA demonstrates the possibility of a new synthesis of ideas and technology. This new synthesis promises unique benefits for science, but they will require more research and effort. Some key technical questions that should be addressed are:

- Throughput requirements of the Intelligent Data Understanding (IDU) algorithms in the "context" of an IA
- Appropriate placement of grid technologies within the overall "knowledge building system"
- Likely physical locations (hardware allocations) for the functions of the IA in the light of projected enabling technologies, and the flows of data, information, and knowledge. (For example, could a decision support system be hosted on a hand-held device?)
- Design choices regarding integration of e-commerce with science software, for command control vs. peer-to-peer negotiation paradigms, for data- vs. software- mobile paradigms, for micro-sensor data collection, for automated quality assessment, etc.

1. INTRODUCTION

One of NASA's vision statements is "To improve life here on Earth." Derived from this are several strategic objectives for the Earth Science Enterprise (ESE). The ESE supports NASA's vision by providing to the research community, policy makers, and the general public, data and information products and the knowledge derived therefrom to enable/improve decision-making. Some of the benefits of this are conserving resources, increasing prosperity, improving the quality of life, reducing impacts of disasters, and saving lives. To enable this, ESE combines the NASA-unique capabilities for space-borne observations with research in various Earth science disciplines to address a number of scientific questions.[1] A suite of Earth observing satellites, funded research investigations, and the Earth Observing System Data and Information System (EOSDIS) along with the federation of Earth Science Information Partners (ESIPs), are now operating to provide an unprecedented amount of data and information products to a broad user community. In addition, ESE supports research in and development of technologies needed for future Earth observing systems,[2] research in and development of information systems' technologies [3] and an applications program.[4]

As expressed in NASA's plans referenced above, some of the key science and applications goals for 2010 are to improve predictive capabilities for:

- Weather
  - 5-day forecasts with over 90 % accuracy
  - 7-10 day forecasts with 75% accuracy
  - Routine 3-day forecasts of rainfall
  - Hurricane landfall prediction with +/- 100 km accuracy 2-3 days ahead of time
  - 3-5-day forecasts of air quality
- Climate
  - Routine 6-12 month seasonal predictions
  - Experimental 12-24 month predictions
  - 10-year experimental climate forecasts
- Natural Hazards
  - Continuous monitoring of surface deformation in vulnerable regions with millimeter accuracy
  - Improvements in earthquake and volcanic eruption forecasts
  - Improved post-eruption hazard assessment

Some of the areas where additional progress is hoped for as a vision for the future (2025) are[5]:

- 10-year climate forecasts
- 15- to 20-month El Niño prediction
- 12-month regional rain rate
- 60-day volcano warning
- 10- to 14-day weather forecast
- 7-day air quality notification
- 5-day hurricane track prediction to +/- 30 km
- 30-minute tornado warning
- 1- to 5-year earthquake experimental forecast

More details about current thinking, identification of research and measurements needed, and projections on what could be achieved are given for three key areas of significant societal impact: biological invasion and ecological forecasting,[6] understanding sea level changes,[7] and understanding and responding to earthquake hazards. [8] For society to derive benefits from these advances, it is necessary not only to conduct the research and development to enable such capabilities, but also to demonstrate such technologies in operational environments and transition them to operational agencies. In most cases, the operational agencies need to use the information and knowledge acquired through such advancements in real- or near-real-time environments to support decisions that have significant impact on society. Thus, for the scientific and technological advances from NASA to be truly applied for the benefit of the community, advances are needed in sensor technologies, satellite systems, scientific research, transformation of data to information and to knowledge, and dissemination of knowledge to support decisions in a timely manner.

Over the past decade, there have been significant advances in our ability to collect, archive, and disseminate data. In the 1980s, NASA's Earth science data were generally held by principal investigators or specialized data systems with little interaction or interoperability. Specific data of interest could be difficult to find. The access to data became easier, and the quality of services associated with data increased significantly, with the development of Version 0 EOSDIS[9] through interoperable, geographically distributed, data centers. The explosion of the World Wide Web (WWW) has since revolutionized the access to information. Considerable progress has been made in ingesting, archiving, and distributing large volumes of data using distributed databases with EOSDIS,[10,11,12] discovery of the existence of datasets and services through the Global Change Master Directory, [13,14] the area of interoperability through the EOS Data Gateway (EDG),[15] DIAL,[16] Distributed Oceanographic Data System (DODS),[17] Alexandria Digital Library,[18] EOSDIS Clearing House (ECHO),[19] and other efforts. Access to specialized data products and applications' development for focused user communities has been enabled by NASA through the Federation Experiment involving over 24 Earth Science Information Partners (ESIPs).[20,21] There are a number of efforts underway to take advantage of distributed computing and storage resources that are generally referred to as "Grid Architectures." Examples of these are: National Science Foundation (NSF)'s National Technology Grid,[22] NASA's Information Power Grid,[23] US Department of Energy (DOE)'s DISCOM,[24] GriPhyN,[25] NEESgrid,[26] and the Particle Physics Data Grid.[27]

Even with the above accomplishments, further developments in information sciences and technology are critical to the achievement of NASA's vision and its deployment for operational applications. Currently, NASA's Earth Science Technology Office (ESTO) supports the development of information system technologies[28] for near- to medium-term deployment in mission applications. In addition, development of more embryonic technologies for longer-term mission adaptation is supported through other NASA programs, addressing needs of all NASA enterprises. An example of such a program is the Computing, Information, and Communication Technologies (CICT) Program, a component of which is the Intelligent Systems Project (ISP).[29] One of the technical areas under the ISP is Intelligent Data Understanding (IDU), supporting several basic research activities and conceptual studies.[30] This report is on one of the conceptual studies within IDU, namely, Intelligent Archives (IA).

STUDY MOTIVATION

The motivating factors for this study are:

- Data acquisition and accumulation rates tend to outpace the ability to access and analyze them. For example, the rate at which EOS data are accumulating in the archives today is about 3 TB/day.
- Beyond the obvious needs for more efficient storage and access to data that are met by improvements in hardware technologies, advances are needed in concepts and tools to enable intelligent data understanding and utilization.
- If NASA is to migrate its technologies to operational agencies' decision support systems, it is necessary to demonstrate the feasibility of near-real-time utilization of vast quantities of data and the derived information and knowledge.
- The variety of data implies a heterogeneous and distributed set of data providers that serve a diverse, distributed community of users.
- Unassisted human-based manipulation of vast quantities of archived data for discovery purposes is difficult and potentially costly.
- While there is no substitute for human intelligence, it is necessary to provide automated "intelligent assistants" that can supplement the abilities of researchers to transform them into information and knowledge.
- The types of data access and usage in future years are difficult to anticipate and will vary depending on the particular research or application environment, its supporting data sources, and its heritage system infrastructure.

Advances in data management, hardware, and software/information systems technologies over the next decade provide exciting possibilities for improved access and utilization of data as well as design and deployment of true "knowledge building systems" (KBSs) that move beyond the basic capabilities of "Data and Information Systems." A KBS in the context of Earth sciences can be viewed as an end-to-end system starting with sensors (space-borne, airborne or Earth-bound) and ending with users who derive knowledge through scientific research and/or exploit the knowledge in real-life applications. Knowledge building involves a dynamic interplay between people and technology that transforms observations into data, data into information, and information into knowledge. An IA fits within this end-to-end context and will support this knowledge building enterprise with new capabilities and facilities.

The purpose of this report is to introduce the IA within its end-to-end context and show its architecture at a conceptual level. Section 2, The Conceptual Architecture, covers the definition of basic terms and a discussion of the architecture. Section 3 discusses intelligent archive use case scenarios to help illustrate operational concepts for an IA within its end-to-end context, in order to highlight new capabilities required by the IA. Section 4 describes technology supporting the intelligent archive and examines the future of key technologies that will influence the evolution of distributed intelligent systems. Section 5 concludes the report with a summary and recommendations for future work.

2. THE CONCEPTUAL ARCHITECTURE

INTRODUCTION

This section presents a context and conceptual architecture for Intelligent Archives (IAs). The architecture represents concepts and models for future intelligent systems that have new capabilities for advanced data utilization. High-level elements, objects, definitions, models, and viewpoints of an intelligent archive system are described. This section also serves as a technical reference for other report sections and for “drill-down” detailed white paper topics.

Advances in scientific knowledge are increasingly enhanced by progressive integration of computing, data management, communications, and sensor technological capabilities into the enterprise of science. This is the hallmark of digital era science; nevertheless, Earth and space science are enterprises that build knowledge. The Intelligent Archive will support this knowledge building enterprise with new capabilities and facilities. We define the terms data, information, and knowledge as follows:

- Data: an assemblage of measurements and observations, particularly from sensors or instruments, with little or no interpretation applied (e.g., measurements from scientific instruments or market’s past performance)

- Information: a summarization, abstraction or transformation of data into a more readily interpretable form (e.g., results after performing transformations by data mining, segmentation, classification, etc., such as a Landsat scene spatially indexed based on content, assigned a “class” value, fused with other data types, and subset for an application, for example a GIS)

- Knowledge: a summarization, abstraction, or transformation of information that increases our understanding of the physical world (e.g., predictions from model forward runs, published papers, output of heuristics, or other techniques applied to information to answer a “what if” question such as “What will the accident rate be if an ice storm hits the Washington D.C. Beltway between Chevy Chase and the Potomac crossing at 7 a.m.?”)

HIGH LEVEL ARCHITECTURE CONCEPT

The end-to-end system context for the IA begins with sensors and ends with knowledge and understanding. By considering a data transformation and utilization process we defined the context to include elements for scientific observations on one end and those elements that transform them for ultimate use as data, information, and knowledge on the other end. Twenty-first century science enterprises will increasingly be supported by computing applications and systems that are, in turn, supported by vast powerful computing and communication infrastructures in this context of transformation and use (see Figure 2-1).

With this “end-to-end” context, we can organize the key elements of the knowledge building enterprise into an overall system viewpoint. In addition, the context provides an excellent framework for applying use case scenarios to discover new requirements, as we discuss in Section 3, page 19. One should notice immediately in Figure 2-1 that the beginning and end of the system context lie within a general distributed system. In fact, all but the smallest intelligent archives will be built upon a geographically wide integration of physical resources and information repositories.

The functions of the IA can be deployed over discrete computing facilities or over a network of facilities as an open, distributed resource. Small to large intelligent archives in the near future can be built and operated either collaboratively or privately because of evolving technologies, standards, and methods. Rapidly changing computing technology will permit the formation of new enterprise system capabilities built upon distributed interconnected infrastructures; a system of systems. With the inclusion of more self-directed automated processes for autonomous discovery, self-aware management, and self-healing systems, new advantages can be realized for both archive users and operators.

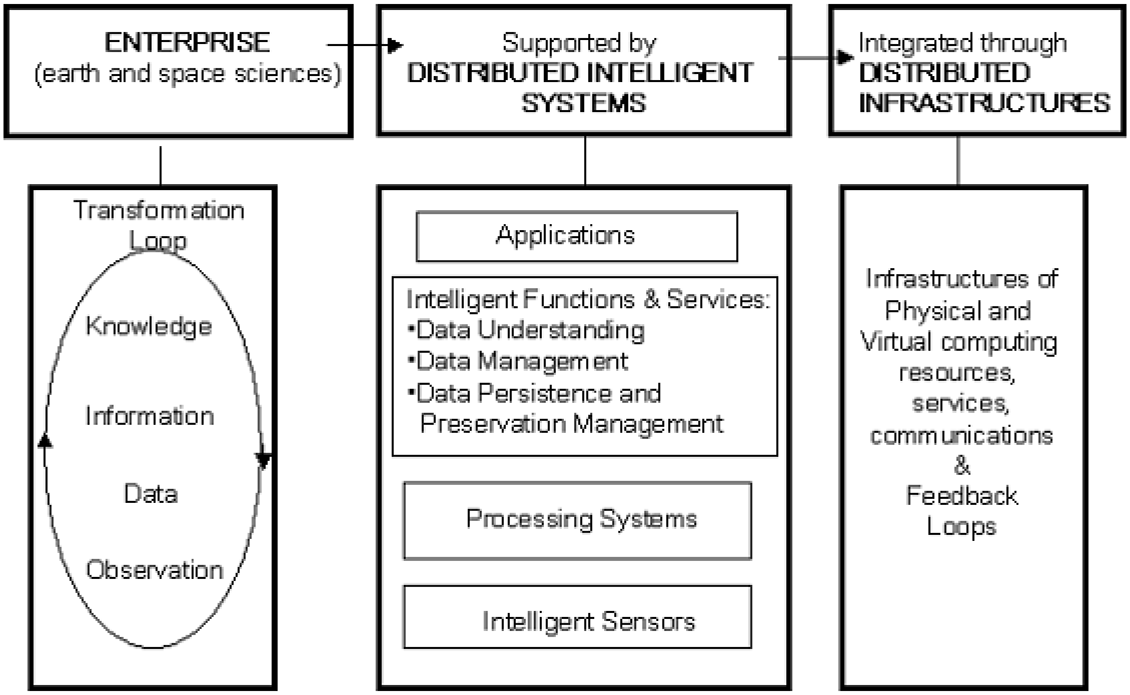
ENTERPRISE
(earth and space sciences)
Supported by
DISTRIBUTED INTELLIGENT SYSTEMS
Integrated through
DISTRIBUTED INFRASTRUCTURES
Transformation Loop
Knowledge
Information
Data
Observation
Applications
Intelligent Functions & Services:
•Data Understanding
•Data Management
•Data Persistence and Preservation Management
Processing Systems
Intelligent Sensors
Infrastructures of Physical and Virtual computing resources, services, communications & Feedback Loops

Figure 2-1: Context and Scope of the Intelligent Archive

This conceptual architecture is forward looking to anticipate how intelligent systems and changing technology can be configured to improve upon conventional process-chains of data acquisition, data processing, distribution, archiving, and use. A critical element of the future configuration will be feedback loops connecting intelligent algorithms, methods, and technologies within the archive. When implemented throughout an enterprise system, these automatic feedback loops create adaptive systems that respond to the need for greater data utilization and improved digital scientific services.

ARCHITECTURE OF THE INTELLIGENT ARCHIVE

ARCHITECTURE ELEMENTS

Architecture elements represent a class or aggregation of objects at the highest level of architecture abstraction. These abstractions can be decomposed into finer levels of detail when modeling a system and composing an architecture representation. Key elements for an IA include:

- Actors – anyone or any thing that interacts with the system; stakeholders (e.g., Scientist End-User, Application End-User, Interim Archive, Permanent Archive, Standard Product Producer, Value-Added Provider)

- Applications –software that interfaces between actors and systems or infrastructures

- Systems – components of an enterprise supporting kinds of processes and functions

- Computing Infrastructures - configurations of computing resources and services that enable and support systems and applications

- Observation Infrastructures – configurations of components aligned to make observations, collect data, and interface with other systems and infrastructures

- Object: a thing, either real world or conceptual, i.e., something concrete or a concept with definable boundaries and meaning.[31] Objects can be hierarchically modeled into different levels of abstraction and detail. This is important for capturing the components and relationships within and among elements that make up the intended system. We introduce a set of representative component objects for each key system element (actors are external) below. While this is not a comprehensive list they are starting points for developing more detailed conceptual descriptions of future intelligent science data management systems.

The rest of this subsection provides a brief discussion of Applications, Systems, Computing Infrastructures, and Observation Infrastructures. Following this, a more detailed discussion is provided of Objects – Required Components of the System, Architecture Object Model, and Intelligent Archive Functions. The subsection is then concluded with a brief discussion of an Operational Scenario.

APPLICATIONS

Analysis Tools – software tools supporting the transformation of data into information and knowledge

Exploration Tools – human-machine interactive portal to data, information, services, and system capabilities for knowledge building activities

Application Model – a complex application-specific model of a subject composed of relevant data, information, algorithms, and structures to view that subject in a virtual context (e.g., a virtual farm)

Visualization Tools – software operating on complex data and information structures for allowing researchers to explore the data content and subject matter visually

SYSTEMS

Science Models – scientific systems representing understandings of natural systems; such computer modeling systems support continued model development, operation, visualization, and knowledge building

Data Production – system processes for production of data and metadata products

Archive – system of capabilities including management, persistence, access, and distribution of data, information, and knowledge products; archives range from small, single-data-type management to institutional multi-product long-term data management systems; archives can be distributed, local, and/or virtualized

Data Assimilation – systems that blend independent observations for comparison with model forecasts and calculate new initial states for next forecasts and alignments with future observations

COMPUTING INFRASTRUCTURES

Virtual Infrastructure – the configurable interface between physically distributed resources & services and application system software (a Grid is a current example described in Section 4, page 30)

Physical Infrastructure – the actual computational, storage, and communication resources and services

Dedicated Infrastructure – privately controlled and restricted computational resources and services for specific use (may not participate or be exposed to an openly shared or collaborative environment)

OBSERVATION INFRASTRUCTURES

Platforms – structures supporting sensors and interactions with other systems

Sensors – instruments for making observations and generating data

Sensor Webs – arrays of deployable, configurable, and interactive sensors and spacecraft

REQUIRED COMPONENTS OF THE SYSTEM

The requirements for the IA are driven by the need to accurately and automatically transform tremendous volumes of data into information and ultimately knowledge. The IA satisfies these requirements by incorporating system intelligence into support systems and services. The conceptual architecture shows how system intelligence can be integrated with systems, with supporting infrastructures, and even with the processes that manage the evolving configuration of host machines and software applications.

The application of system intelligence to knowledge building support systems and services is the key to the accurate and reliable automation of the processes transforming enormous data volumes into information, and ultimately knowledge. Thus, the architecture stipulates management requirements for the intelligent system to ensure the system functions in the broader enterprise of end-to-end data transformation. For example, system intelligence would be incorporated when existing applications are interconnected or integrated with an automated scheduling system. Many of the system elements that partially satisfy these requirements are illustrated in the IA model shown in Figure 2-2.

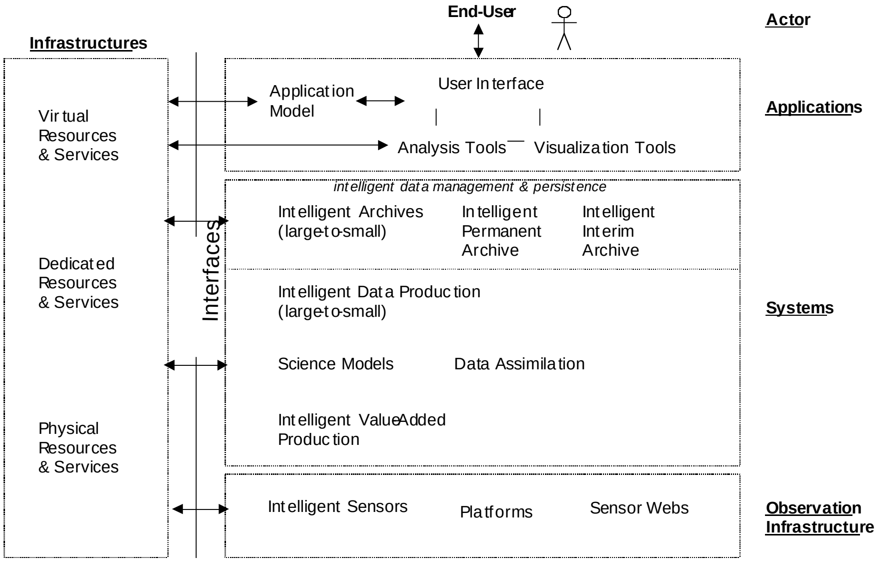
End-User
Actor
Infrastructures
Applications
Virtual Resources & Services
Application Model
User Interface
Analysis Tools
Visualization Tools
intelligent data management & persistence
Intelligent Archives (large-to-small)
Intelligent Permanent Archive
Intelligent Interim Archive
Dedicated Resources & Services
Interfaces
Intelligent Data Production (large-to-small)
Systems
Science Models
Data Assimilation
Physical Resources & Services
Intelligent Value-Added Production
Intelligent Sensors
Platforms
Sensor Webs
Observation Infrastructure

Figure 2-2: Model for the Knowledge Building System

An archive of the future should perform core functions with increased automation, efficiency, and flexibility in a dynamic knowledge building enterprise. Archive-related functions requiring system intelligence and data understanding include:

- Data management – store/retrieve, control, and distribute data, information, and knowledge (i.e., data products, models, simulation outputs, extended complex metadata, product-algorithms)
- Value-added products and services – data characterization, mining, fusion, transformations, re-sampling, re-projections, feature detections, and quality control
- Operations – high reliability, self-monitoring, self-healing, self-adjusting, autonomic "lights-out" internal operations
- Interfaces with resource infrastructures – dynamic sensing and recruitment of connected computing capabilities, communication assets, and persistent storage.
- Interfaces with cooperating systems – expose available services and interfaces for other systems to use including external applications

System intelligence also resides in interface functions between systems as intelligent system-to-system interfaces become virtualization layers. A virtualized system offers abstract, functional system interfaces that can be registered as a service and then used by other systems to support interoperations among applications. Virtualized system resources (see left vertical column in

Figure 2-2), such as Grid-like plug-ins, permit interactions between archive-related systems and physical resource infrastructures (see Section 4, page 30).

ARCHITECTURE OBJECT MODEL

An understanding of concepts for intelligent archives emerges from a model of enterprise system objects. The object model describes the types of objects in the system and various relationships that exist among them. Expanded models that show more details about attributes, operations, and associations of each class are beyond the scope of this report. They will be the topic of a "drill-down" white paper. The knowledge building system introduced in Figure 2-2 contains four subclasses that aggregate next level systems ranging from IAs to intelligent sensors (see Figure 2-3). Three of these are further decomposed into member object classes ("…" denotes more classes exist).

Some examples of associations between lower level classes and higher level-classes represent logical relationships. For example, both interim and permanent archives are related operationally or functionally with many data production systems. In other words an archive can participate with data production systems in a dual role as suppliers of source data for processing and as a repository for finished data products. On the other hand sensors are associated with production systems and an interim archive in more complex ways.

In one scenario, many sensors are directly associated with one interim archive. Conversely, several archives may support data from one kind of sensor. And multiple sensors can be associated with multiple archives. To illustrate this, we can imagine a time when a spaceborne sensor may also have a local archive as part of its configuration of technology. Furthermore, this local sensor-based archive can interface with a ground-based archive to form a dynamic virtualized archive supporting new kinds of on-demand data access closer to the source of observation in time. Persistence of data will vary based on the amount of distribution from the root source, replication generations, and policies of each archive organization participating in managing these data.

Additional scenarios can be derived from logical associations among the different object classes for intelligent archives. In one view, sensors and archives and production systems are logically interrelated as separate objects. But in an end-to-end context, the architecture affords future system developers flexibility with regard to implementations and configurations of component objects, such as the spaceborne sensor's archive above. A logical archive, either sensor-hosted or ground-based, could also be seamlessly bundled with a data production capability. Such an archive component with a nested data production component could include a self-improving capability for generating content-based metadata products. Other kinds of intelligent archive functions similarly can be configured resulting in dedicated and/or virtualized system capabilities.

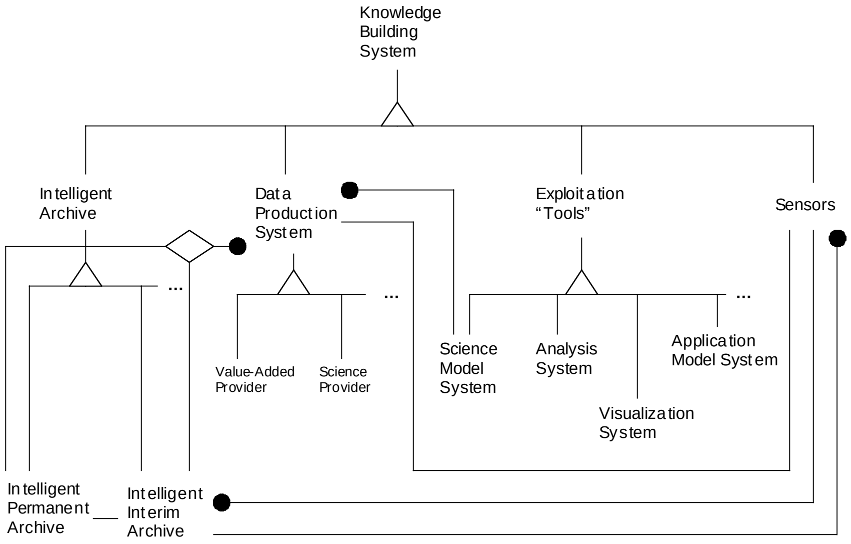

Figure 2-3: End-to-End Context of Knowledge System Object Model

INTELLIGENT ARCHIVE FUNCTIONS

An IA can have a broad range of functions and purposes. Differences in these characteristics contribute to a taxonomic classification of archives. For example, there can be interim and permanent archives. Permanent archives preserve science data, information, and knowledge assets indefinitely, whereas interim archives establish policies on the duration of data retention. While these have much in common, the former has significant needs for intelligence in the difficult job of data migration, the latter in its decisions on retention. Some archives will perform functions required for science missions, local science team activities, value-added production services, and the knowledge building enterprise as a whole. Each will have a mix of common and unique functional intelligent capabilities.

Figure 2-4 introduces an overview of the core functional capabilities. The figure also illustrates how functions (i.e., software applications and libraries) are shared between subclasses of archives and how archives cooperate with each other and external users.

Both interim and permanent archive systems draw upon at least five core functions. The rounded rectangular boxes represent core system functions with corresponding sub-functions listed in dashed boxes. Intelligent systems and data understanding technology underlie many of the core archive functions. Future functions for operations will be characterized by intelligent capabilities that increase automated self-regulation, recovery and performance tuning, and management of interface control information for seamless cooperation with other systems. Intelligent data management governs such functions as semantic queries, retrieval strategies, cataloging of assets for human and machine level access, and machine learning applied to access patterns to improve storage and retrieval strategies.

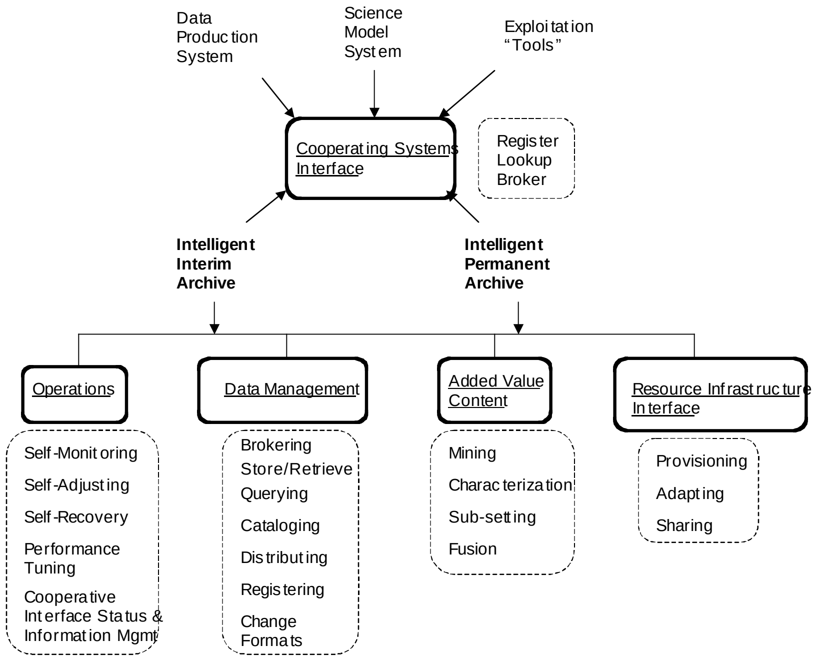

Figure 2-4: A Model of the Archive Focused on Objects and Functions

Intelligent data understanding techniques will support functions like data characterization and data mining, improving overall archive metadata generation. Dynamic complexities of computing resource allocation and management will also be enabled by intelligent functions for provisioning, adapting and sharing infrastructure resources. Similarly, functions that support cooperation with external systems will have intelligent interfaces to dynamically register, locate, and broker interoperable distributed services for on-demand seamless functioning. This sample of intelligent archive functionality sets the stage for introducing emergent system behaviors.

OPERATIONAL SCENARIO

Embedded and distributed system intelligence permeates archives and the other architecture component objects in the end-to-end system context. Given the abstracted architecture object and functional model, we can explore the behavior of this knowledge building system using a scenario. Figure 2-5 describes a sample operational flow.

Transformations of observations into utilized data and information by the enterprise system flow horizontally among the high-level system component objects in Figure 2-5. Interactions below each object show how the component objects collaborate for a sample scenario. The behavior of the collaboration follows two aspects of an illustrative operational scenario; routine processes and on-demand use.

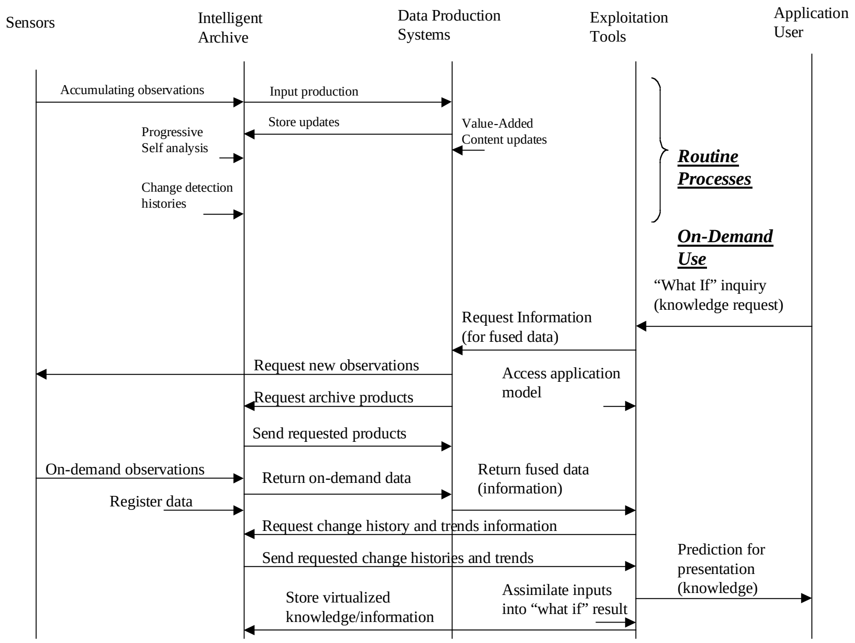
Sensors
Intelligent Archive
Data Production Systems
Exploitation Tools
Application User
Accumulating observations
Input production
Store updates
Value-Added Content updates
Progressive Self analysis
Change detection histories
Routine Processes
On-Demand Use
"What If" inquiry (knowledge request)
Request Information (for fused data)
Request new observations
Access application model
Request archive products
Send requested products
On-demand observations
Return on-demand data
Return fused data (information)
Register data
Request change history and trends information
Send requested change histories and trends
Prediction for presentation (knowledge)
Store virtualized knowledge/information
Assimilate inputs into "what if" result

Figure 2-5: Interaction Diagram Describing Example Operational Flow

Observations accumulate routinely in an archive. The archive provides data to production systems that in turn, store products in an archive. These flows are conducted seamlessly through intelligent functions for coordinated interoperable processing. The archive also automatically analyzes its contents to build and update archive metadata. Metadata are enriched by goal-oriented discovery and data characterization algorithms. These algorithms constantly detect features and trends in the data over time and update archive metadata accordingly.

Application users have flexible, on-demand access to data and information concurrently with routine operations. For example, in one simplified scenario, a scientist developing a high resolution model of a weather event needs to perform a "what if" inquiry to test predictive accuracy of the model. The scientist uses an application tool interfaced with the knowledge building system to formulate an inquiry. The application tool automatically selects required services and issues requests against real or virtual archive holdings. Requests for virtual holdings trigger a cascade of requests to obtain, for example, fused data, or subsets of observations, or the output from another predictive model or even new observations. Though not shown in the figure, that cascade of actions across these cooperating nodes would require the IA to be a repository of protocols of the permissible rules for collaboration, and might even require storing an image of the dynamic "state" of its collaborating nodes to assure efficient collaboration. The role of machine intelligence here is not so much to solve the scientific problem but rather to acquire specific data and computational resources that answer recognizable steps in the solution specified by the scientist. The IA complements the scientist who cannot be expected to track the changing holdings and capabilities of a vast computational enterprise.

Continuing with the sequence scenario, requests are issued and addressed through collaboration among production, archive, and sensor systems. As the on-demand observation is returned it is registered with the archive and returned to the production system. Segmentation, re-sampling, rescaling, subsetting, and data fusion services invoked at the production system are performed and the results sent to the exploitation tool. Upon high-speed analysis the exploitation tool issues an additional request for long-term trend information from the archive that is returned and assimilated into a "what-if" visualization result that is simultaneously stored in the archive and presented to the scientist. (See the scenario subsection, page 19, for further operational discussion and illustrations.)

ARCHITECTURE SUMMARY

Here, we have described a conceptual architecture that will evolve along with changing technology but one with a persistent viewpoint for addressing improved data management and use. The architectural viewpoint encompasses a context that includes applications, systems, infrastructures, and interfaces yielding a coherent distributed knowledge system. The architecture envisions new functional features with the ability to:

- Store low level data and produce products on demand (virtual products made possible by powerful but inexpensive processing resources)

- Distribute data products to many nodes to allow data mining to be performed concurrently where each node is expected to have its own storage and processing resources

- Permit the storage of data to transcend initial mission/project life

- Flexibly store data close to the sensor allowing the sensing system to perform data maintenance (calibration, correlation, product generation)

- Enable sensors and science model systems to become IA system "users"

- Facilitate data backup by replicating data on distributed system nodes (the number of replications and versioning will be determined by data usage, value and storage cost)

- Support system-wide collaboration and adaptive load management with low level planning/scheduling functions combined with self monitoring capabilities (these constantly assess the health of system elements, user demands, data inputs, and long-term science demands formulating alternatives that will enable the system to adapt to changing circumstances as they occur)

Intelligent archives offer new capabilities distinguishable from archives of today. Highlights of these differences include:

- Manages and stores full representations of information, and knowledge in addition to data

- Builds intelligence about building data, information, knowledge, and accompanying services involved in a science enterprise

- Performs self-analysis to enrich metadata that adds value to an archive's holdings

- Performs analyses to detect long-term trends and features

- Interacts as a cooperative node in a “web” of other systems to perform knowledge building (where knowledge building involves the transformations from data to information to knowledge) instead of just data pipelining

- Is aware of other nodes in the knowledge building system (participates in open systems interfaces and protocols for virtualization, and collaborative interoperability

Many operational scenario variations and permutations are possible given the architecture model of elements, component objects, and associations for an intelligent knowledge building system. The conceptual architecture establishes a framework for how future intelligent archives operate in a context of cooperating systems and infrastructures. Furthermore, the architecture is a necessary guide where opportunities for integrating intelligent systems, data understanding, and machine learning can be identified and mapped to goals for automation that increases effective data utilization.

3. INTELLIGENT ARCHIVE USE CASE SCENARIOS

PERSPECTIVES THROUGH USE CASE SCENARIOS

This section describes scenario-based perspectives with which to develop visionary IA capabilities. Scenario-based approaches drive clear pictures of end-to-end interrelationships among data and information, consumers, data providers, value-added information services, archives, and data acquisition missions[32,33]. Exploration of conceptual architectures through use-scenarios provides insights into limitations of existing solutions, challenges, issues, and probable opportunities for innovation. We have explored scenarios for precision agriculture, advanced weather prediction, and virtual astronomy observatories to uncover requirements for intelligent system services and capabilities. The consideration of these three scenarios has yielded additional perspectives on how intelligent archives and intelligent system applications can lead to better data utilization. A brief discussion of these scenarios follows.

PRECISION AGRICULTURE SCENARIO

INTRODUCTION

Precision agriculture is concerned with refining agricultural practices to maximize crop yield with minimum cost through improved use of data, information, and knowledge technologies. Demands for data and information services are explored through a scenario involving a grower (farmer) for an individual farm. Such a system could form part of a larger grid, with precision farms aggregated for example to the regional level. Use cases for this example consist of planning, cultivating, and harvesting.

DESCRIPTION

Precision agriculture in this scenario is concerned with the scope and parameters of a minimal agricultural production unit. The farm is characterized as a relatively small spatial area (considered in acres) for agricultural products suited to regional ecological, weather, and growing constraints. Geo-spatial information relevant to these farms puts a premium on high-resolution data both spatially and temporally. A digital farm is a sophisticated intelligent data and information pool. It serves as an interface between specific interests of the farm and vast external data sources. It also serves to integrate and administer different kinds of extracted and filtered information.

Daily data, information, and knowledge support the principal farm activities of crop planning, cultivation, and harvesting. Information-intensive support services for each activity include current conditions monitoring, histories and time series studies, trends/risks analysis, prediction and forecasts, “what-if” investigations, and outcome comparisons. Detailed information about land, weather, water, agriculture markets, prior yields, agri-chemical options, seeds, etc. help in planning crop selection and planting. High-resolution information is required to monitor, assess risks, and make decisions about appropriate interventions to maintain crop health. Similarly, to maximize yields, decisions about harvest timing benefit from information about current and future conditions. The digital farm maintains a model for perpetual use that integrates remotely sensed information about farm assets and information collected from the farm about outcomes of plans, cultivation techniques, and harvests.

OPERATIONS

Summary and detailed information about every aspect of the farm, including past, present, and future, is available through an interactive digital farm application. The digital farm works on behalf of the grower. Digital farm services are available from multiple portals (workstations, mobile devices) from within the house, farm buildings, vehicles, or even a combine. Interaction with the digital assistant can be conducted by natural language either via voice or keyboard.

A digital farm can interpret, broker, and fulfill requests for information dynamically and autonomously. The digital farm fulfills this role by brokering and invoking services to create and maintain encyclopedic farm-relevant information that is ontologically, spatially, and temporally organized. The digital farm must preserve information about soil, crop, weather, and moisture conditions. Moreover, it constantly updates its information with farm-specific in situ sensor inputs and from external sources of data. Indeed the farm’s interfaces with the outside are crucial to pooling farm-relevant data from primary archives and agricultural services.

Virtual farm services produce a multidimensional model of the entire farm that the grower can inspect from his or her office or combine cab. The virtual farm serves as an interactive reference of farm-specific assets integrated with historical, current, and modeling information. Views of the farm can be summoned to within a square meter with variable time series. Types of information range from historical to actual current conditions to what-if scenarios cast into the future.

ROLES FOR SYSTEM INTELLIGENCE

The virtual farm is predicated on applications of intelligent data understanding. Therefore a key theme is filtering and distilling specific farm-relevant information from an avalanche of data involving a chain of intelligent data reduction and information/knowledge extraction services. Intelligent capabilities must derive and transform information from data and models scaled to the interests of the specific farm (to aid in “what-if” predictions). Functional components that facilitate the distillation of information relevant for a farm on a twenty-four hour basis include:

- Instrument/sensors for resolving farm-scale features and assets

- Hyperspectral data mining and extraction techniques
- Machine learning for automating the evolution of system intelligence
- Modeling and prediction algorithms
- Visualization of preprocessed and dynamically processed data

These component technologies must be designed and implemented to interoperate throughout data production, quality assurance, cataloging, metadata extraction, data content characterization, data product archiving, and custom services. Data understanding is crucial for automating the distillation, characterization, detection, and extraction of such things as objects, phenomena, features, conditions, and other contributors to knowledge.

TECHNOLOGY CHALLENGES AND OPPORTUNITIES

Precision agriculture will benefit from intelligent systems technologies that enable virtual farms to broker requests for appropriate internal/external information services. Technologies will be needed to support dynamic and semantic interactions between information querying interfaces and intelligent archive systems. Precision agriculture is predicated on accurate, temporal, spatial, and multi-parameter data tailored to the concerns of the consumer. Serving required data and information on-demand requires embedded intelligent technologies throughout the enterprise system. A major challenge for cooperating systems used in precision agriculture will be to rapidly and accurately transform observation data into specific data products usable by each agricultural unit. Functional transformation capabilities must be coupled with considerable throughput capacities. For example, future annual precision agriculture data consumption scaled to the requirements of a California Central Valley agricultural zone is estimated at 1.5 TB per 1000 acres. This quantity becomes a significant challenge when considering there are approximately 940,000,000 acres of farmland in the US. Extrapolating further nearly 1,411PB of data per year will need to be processed into local views often in near-real time. Opportunities for embedded intelligent systems operating in high-performance infrastructures to help crop growers access, manage, and use valuable data daily include:

- Sensors, networks, and interfaces to interoperable processing infrastructures
- Services environments for producing, managing, distilling, and filtering precision agriculture data and information products (i.e., historical, current, what-if predictions)
- Algorithms for hyperspectral data understanding and mining
- Automated data access (subscription), query, and data fusion
- Intelligent interfaces for machine-to-machine and human-machine interactions.

ADVANCED WEATHER FORECASTING SKILL SCENARIO

INTRODUCTION

Earth science models provide the primary tool for weather forecasting. The basic concept involves applying a suite of equations (models) to measurements of heat, cloudiness, humidity, and other integral parameters to project how those factors change over time – thus forecasting the behavior of the atmosphere. Because weather forecasting science attempts to accurately simulate the real world, it must rely on observations, modeling, and theory. Determining correct initial conditions for starting models needs to be addressed when attempting to minimize errors and prevent models from straying into uselessness. If science can get the weather models right with more detailed observations, then we might see nearly perfect three-day forecasts and even reliable ones beyond ten days.[34]

Models are complex computer codes that divide the atmosphere, land, and oceans into hundreds of interacting grids. Until recently, the resolution of the grid boxes has been as large as 185 miles on a side. This meant that many of the Earth's features that affect changes in weather were smaller than the boxes and had to be approximated. New models are now approaching 20 miles on a side and can account for more of these details (but still not all of them). This refined resolution provides more precision in the model and hence, more accuracy in its predictive power. Next generation systems envisioned to support and improve this process tightly couple modeling systems with observation systems.

DESCRIPTION

The context for a future weather forecasting skill scenario interrelates observation, science, information services, and information consumers. The scenario is based on conceptual elements of the ESTO Weather Prediction Technology Investment Study.[35] This study identified science applications and technology required to enable skilled weather forecasts of ten to fourteen days by 2025. We introduce the concepts for this scenario here and reserve full descriptions for a related white paper.

Overall the future weather forecasting systems includes weather science teams; weather remote sensing missions; data processing, modeling, dissemination, and archiving groups; weather prediction information services; and weather information consumers. Atmospheric and climate sciences collect and assimilate data into research models. Building and improving models is an integral part of this knowledge system.

Predicting or forecasting future weather conditions over a particular region requires accurate data and knowledge about atmospheric forces, physical parameters, and the interrelatedness of atmosphere to the whole Earth system. However, the accuracy of weather predictions tends to fade rapidly as a function of time due to a lack of precise initial conditions and the non-linear complexities of weather.

OPERATIONS

While the theoretical upper limit of weather forecasts is two weeks, the precision of forecasts from an operational weather system is limited by the methods for gathering essential data on the initial conditions (actual state) of the atmosphere at any model starting time. In this advanced scenario, precision can be improved within fixed limitations on measurement sensors by organizing two-way interactions between the observing systems, data assimilation, and computer models. This new organization can autonomously tailor observing strategies based on knowledge of current and future states of the atmosphere obtained by modeling. By means of these interactions, a weather system following the advanced scenario can confirm, predict, and validate model parameterizations or make measurements in specific locations to improve models and forecasts. The system can also anticipate support of reconfigured prediction models based on autonomously obtained observation inputs. Overall complexities of coordinating and commanding interaction among the interdependent parts of the overall system will require system-wide embedded intelligence, ensuring seamless consistent orchestration and control.

The observing system provides comprehensive observations and measurements in real time. It must be flexible to provide special observations on demand. And, optimally, the observation system combines all sources of sensor assets for surface, atmospheric, and space–based data. The envisioned sensor web will be reconfigurable and tasked in response to evolving model requirements or events of interest. The modeling and data assimilation system couples terrestrial and space observations to a continuously running mesoscale model. In this manner, the global mesoscale model's sophisticated parameterization can be updated quickly and regularly with actual observational inputs. This kind of periodic infusion of actual observational inputs into the model overcomes model "drift." A significant challenge for this vision system is being able to rapidly transform acquired data for use in real-time.

ROLE FOR INTELLIGENCE SYSTEMS

Several opportunities for intelligent system applications can be identified for architectural segments of the weather forecasting system. The observing system segment requires intelligence for:

- Observation management (overall arrays and individual members)
- Scheduling, commanding, onboard operational autonomy management
- Distributed processing coordination
- Automated calibration and quality control, communications integrity
- Translating dynamic observation requests into optimal and timely results
- Implementing flexibility within and across platforms
- Tailoring observing strategies based on knowledge of current states
- Detecting, recognizing, and capturing specific events, states, and phenomena
- Registering and tagging observation data with appropriate information compliant with standards for product development and use by other archives, systems, disciplines, and campaigns
- Adapting to new configurations autonomously or by human command

Opportunities for applying intelligent systems to the modeling/data assimilation system segment include:

- Data ingest and assimilation management
- Brokering observational data requests and observation targeting decisions
- Interpreting data for recalibration, transformation, re-projections, and conversions
- Facilitating model coupling
- Data mining/data understanding, automated analysis

- Determining where data collection is needed
- Automatically detecting gaps or quality issues
- Management of model results for interoperability
- Creation of information and knowledge products

In addition to system-to-system and within-segment processing opportunities for machine intelligence, intelligent applications are required for facilitating human-in-the-loop interactions. These areas include monitoring, control, recovery, tuning, operation, and modification functions for the creation of user-specified and standard weather products.

Data management issues underlie the end-to-end system. Because of the real-time and near real-time characteristics of the visionary weather forecasting system, a heavy premium is placed on dynamic data management. This is reinforced by the distributed data processing that will also occur in various parts of the overall system (e.g., at the sensor source, in a data processing facility, at an archive, a modeling center, or with a science analysis team).

Sensor webs and data assimilation/modeling centers will generate huge volumes of data and information necessitating new intelligent data management strategies to cope with the dynamic complexities associated with acquisition, processing, use, and persistence. Intelligent data management and access to these massive data stores on a dynamic basis indicates a critical need for an IA solution.

TECHNOLOGY CHALLENGES AND OPPORTUNITIES

Achieving future skilled weather forecasting goals requires innovative space-based, airborne, and terrestrial sensor systems producing weather data of various resolutions, rates, bands, parameters, and volumes. Improvements to existing forecasting system capabilities combined with evolving infrastructures and innovative research technologies can enable skilled weather forecasts of ten to fourteen days by 2025 (current forecast predictive skill is five to seven days). Challenging areas for innovative technology and applications include:

- Quality, mixed-resolution observations and data acquisition systems
- High-speed communication and processing of observations
- Rapid complex data assimilation strategies
- Predictive modeling strategies and algorithms
- Powerful archiving, distribution, and interactive visualization technology infrastructures

Advanced weather model building and analytical tools will generate additional types of data and information requiring special archive services. An assessment of expected optimized global data volumes covering required parameters, temporal, horizontal, and vertical resolutions, and vertical measurement layers could reach 7.5 TB/day by 2025.[36] Technologies supporting very rapid operational processing and data management services will be required for especially for data assimilation, interchange, modeling, archiving, and visualization uses.

VIRTUAL OBSERVATORY

INTRODUCTION

Virtual observatories aim to handle a proliferation of large astronomy datasets with built-in software tools for scientists to query and mine data across archives.[37] Over the next decades astronomers expect to practice “precision cosmology” for example, leading to characterizations of the size, structure, and evolution of the universe. Achieving goals such as these will require the collection and integration of petabytes of data from space and ground surveys.[38] Already several projects worldwide are developing virtual observatory mechanisms to federate collections of data and information for an entire scientific discipline (i.e., National Virtual Observatory, Astrophysical Virtual Observatory, Astrogrid, Astrovirtel). The goal is to knit these projects together so that ground and space-based astronomy archives are linked and accessible to all. Intelligent archives are well suited to the functional task of integrating services that a virtual observatory requires to manage and make accessible astronomy datasets.

DESCRIPTION

The data avalanche in astronomy is attributable to increasing use of Charge Coupled Device (CCD) cameras for telescopes. Advances in electronics permit the deployment of gigapixel instruments used to conduct multi-spectral surveys of the sky. Indeed the volume of pixels from a typical astronomical CCD detector doubles every two years with this rate increasing in the near future.[39] Like genomics and other cutting-edge fields of science, astrophysics must cope with the growing problem of how to manage and make use of enormous distributed amounts of data generated by digital-based instruments and experiments.

A virtual observatory is a science-driven effort that emphasizes bottom-up sharing through common integrated services for both observational data and toolkits of analysis software. Many existing astronomical data archives are associated with a specific instrument. Each specific archive has historical reasons to archive data in a particular way. But systematic studies and surveys of the celestial objects depend on combining multi-spectral and multiple instrument data from multiple archives. One solution is advocated by the archive institutions participating in the National Virtual Observatory. In this solution, each institution maintains control over individual data holdings but institutions can share data by conforming to extensible metadata standards and interchange protocols. Users in the community will also be able to share data and analysis tools through properly designed virtual observatory interfaces. A core set of standards, interoperability, and management services will be necessary to support distributed virtual observatory operations.

OPERATIONS

Different telescopes acquire different observational data in different formats managed in geographically distributed archives. Many archives also link data from multiple experiments. A virtual observatory of collective archives will operate based on a common system approach for data pipelining, archiving, and retrieval. It also ensures easy access to data in these archives by a diverse community of users. The system will enable distributed development of a suite of commonly usable new software tools for querying, correlation, visualization, and statistical comparisons of shared data.

Virtual archives utilize high-speed network connectivity among participating archives and terascale computing facilities.[40] Because bandwidth is a limiting factor computation needs to be performed close to the data. But Grid technology infrastructures will allow remote access to both data and computing/analysis facilities facilitated by an operational virtual observatory.

ROLES FOR SYSTEM INTELLIGENCE

Opportunities for intelligent systems in virtual observatories include: automated data corrections, reprocessing, and recalibrations (e.g., data "curation"); cross-dataset and cross spectrum data mining with automated phenomena "discovery"; open-ended resource discovery; enriched cross-archive metadata generation for uniform query and browse, data fusion and combination management. Roles for intelligent archive capabilities are aligned with the goals for virtual observatories to closely link and coordinate data curation, archive data management, uniform access, and data mining services. Overall intelligent archive and Grid functions will allow astronomers to more easily find, access, analyze, and manipulate data as if it were local to their workstations.

TECHNOLOGY CHALLENGES AND OPPORTUNITIES

Concepts for multi-wavelength synoptic virtual observatories have already inspired active research in Grid technology. Grid infrastructures will enable large-scale interoperable sharing of data and computing resources necessary to support next generation astrophysical surveys, analysis, and research. New algorithms and tools will also be needed for data mining, statistical analysis, visualization, data fusion, data filtering, and virtual interactive interfaces. New protocols and standards for transparent access to multiple distributed heterogeneous systems will also be needed to support data discovery, metadata and query interchange, code-shipping, and data product delivery. Intelligent interfaces will also be required for storing and managing both data and metadata throughout a virtual observatory. Opportunities for intelligent archives are well suited for this emerging but visionary space science system environment.

4. TECHNOLOGY SUPPORTING THE INTELLIGENT ARCHIVE

PURPOSE

It is useful to understand the effect new and emerging technologies will have on intelligent archive concepts at the time of postulated prototyped and full system implementation (2012 through 2025). Based on a review of relevant literature we believe that one can expect to see dramatic changes in both functionality and performance in information system technologies in the time period of interest. However, because of the difficulty of accurately projecting into the future, this report looks at supporting technology in the near- to mid-term, and then provides a short glimpse of the longer term.

NEAR- TO MID- TERM SUPPORTING TECHNOLOGY

"Evolutionary" technology changes are more likely to affect our ability to prototype and begin implementation of an IA. Some of these changes are a little closer to our vision at this point and are discussed in this section.

The algorithms developed for the Intelligent Data Understanding area in the Intelligent Systems Project demonstrate what is possible when intelligent systems are applied to scientific research. Will it be feasible to deploy an intelligent system in an operational environment characterized by prodigious data volumes and an expectation for near instantaneous recognition of phenomena? Intelligent archives depend on robust intelligent algorithms to meet the challenging performance requirements of data intensive earth science and space science research. When this will be feasible also depends on the availability of sufficient capacities and types of computing, storage, and communication technologies.

In this report we introduce a survey of technology areas that have a bearing on the future of intelligent archives and the architectures in which they will be configured. Each of these technology areas should continue to be monitored and updated for new developments. The perspective of NASA's Intelligent Systems program underscores the importance of conducting broad technology reviews when developing conceptual architectures and intelligent archives:

> *"The IS Program is developing critical system intelligence capabilities that will not otherwise grow out of current NASA information technology R&D activities, and that cannot realistically be expected to emerge from other sectors."*[41]

In order to focus attention on what must be done in any future IA development, it will be necessary to track what will likely be done by others and avoid duplication. In this section we introduce a survey of technology areas that have a bearing on the future of intelligent archives and the architectures in which they will be configured. Each of these technology areas should continue to be carefully monitored and updated for new developments.

It is certainly true that the IA must be affordable as well as feasible. Matured estimates of costs for developing intelligent archives balanced with cost savings and the value to science have yet to be explored.

DATA PROCESSING

This report has pointed out the challenges of data volume and processing load. Indeed, it may never be possible to perform all desired processing on the sensor data, particularly for hyperspectral sensors. The technology survey has focused on how improvements in computing hardware might mitigate the problem.

For over thirty years, the power of computers has doubled roughly every 18 months, an empirical result dubbed Moore's Law. For that reason, software that was once confined to a few special purpose machines now runs everywhere, e.g., computer graphics. While Moore's Law has held over three decades in silicon-based semiconductor technology, it is now approaching quantum mechanical limits. Several alternatives (e.g., quantum computing, asynchronous computing, 3-dimensional chips) are in research and development in an attempt to surmount the quantum barrier and continue uninterrupted progress. An interruption of many years is possible, however, alternative plans are required. If Moore's Law ends or passes through a period of stagnation, a second, lesser known industry trend will become important. Technology that falls behind Moore's Law of performance does not disappear immediately; but continues to sell in high volume at a lower price due to improvement in production techniques and amortization of capital expenses. With lower prices, it would be affordable to build extremely large collections of semiconductors in a single unit. If the parts can be made to cooperate effectively, the resulting computer will set a new standard for performance/price ratio. The barriers to this approach are the software and hardware interconnection technology.

The limitations imposed by interconnections are most noticed in highly-parallel processor configurations. The gold standard for a highly-parallel supercomputer is to produce N*R operations per second where N is the number of processors capable of R operations per second. To reach that goal, specialized interconnection technology must provide a fast connection between any two processors as well as between any processor and any memory element. The required “switching-fabric” is more complicated and expensive than the technology of the Internet and many different designs have been tried. The optimal design has not yet been found.

The previous paragraphs address parallel computing elements that are closely coupled and executing in “fine-grained” parallel processing mode. There is also a class of problems that allow the work to be subdivided into large parcels and executed in “coarse-grain” mode. For these problems, one might consider a configuration that divides the work over widely separated machines interconnected by a network, i.e., “grid computing,” which is addressed later in the report.

For space applications it is important to mention that Moore’s Law applies for flight-qualified hardware, but the law shows a 5 to 7 year delay relative to ground-based components. One reason for the delay is that each generation of space-qualified components has started out as a ground-based, commercial design. After that follows an engineering program to radiation-harden the design. The demand for space-qualified components is too small to justify the investment in an entirely new design. On the other hand, continual progress on the ground almost guarantees that a significant amount of processing can be eventually collocated with the sensor in space.

DATA STORAGE

Data storage capacity has improved exponentially over time at a rate similar to the Moore’s Law (doubling capacity every 18 months) improvement of processing speed. However, over the last decade, science data volumes (see Figure 4-1) have risen commensurately with computing power and storage capacity for a given cost as demand keeps pace with capacity. The reason is a complex interplay between science requirements and what can be afforded. Thus, we cannot assume that further advances in storage or computing technology will solve the technological problem posed by future growth in data volumes. Furthermore, growth in gigabytes/$ in storage capacity does not tell the whole story. While this figure has been increasing rapidly in both disk and tape technology areas, the ability to access large quantities of data (i.e., seek time and throughput) has not been keeping pace with storage capacity. Random access to stored data still incurs latency due to the mechanical motion of access mechanisms, an effect magnified when slower mechanisms such as tape or removable media are used to reduce the cost/byte of storage in an archive. In tape archives, fundamental limitations are eventually reached at the macroscopic level, that is, the speeds of such elements as the robotic arms or tape transports.[42] In disk systems, the cost of connecting and managing large disk arrays eventually dominates the cost of the raw disks.[43] It is thus uncertain when or even if disk storage will replace tape storage for data volumes at the highest end.

This area of storage performance is a matter of particular concern for the IA because the number of random access queries against the data store will rise dramatically when the demand from autonomously operated intelligent algorithms is added to the current demand from scientist-led investigations. Issues concerning increasing numbers of available granules will also affect database and disk efficiency, seek times, and throughput. However, it is possible that a breakthrough in storage technology could solve the data access problem. For example, the IBM "Millipede" nanoscale probe storage technology uses electron force microscopy technology to support parallel reads and writes, with projections of very high storage densities and low production costs.[44] In planning for future data volume growth, it will be important to track not just the storage capacity of raw media, but the overall cost for a given technology to store and distribute high data volumes. Also, with the replacement of tape by disk somewhat uncertain, caching and archive organization strategies may continue to have a significant role in an intelligent archive.

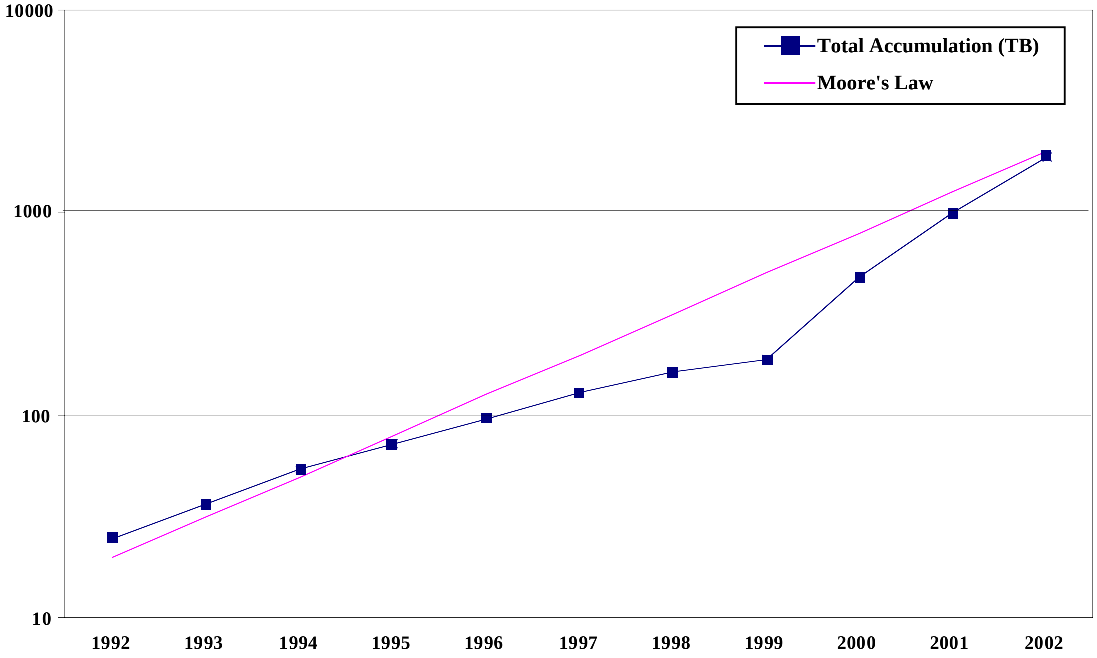

Figure 4-1: Comparison of Earth Science Data Archive Volume Growth to Moore's Law Governing Computer Power

COMMUNICATIONS

For ground-based networks, optical fiber communication is improving faster than any other current technology. Nevertheless, bottlenecks remain in connectivity, resulting from a number of factors such as switch speed and economic viability. Within the next several years, however, optical switching elements should begin appearing on the market. The transition to optical petabit bandwidth and switching capacities are underway. Dense Wave Division Multiplexing (DWDM) for example, multiplies bandwidth capacity by using different color wavelengths to carry data in parallel. Next generation DWDM capabilities will have 300-500 colors each supporting 40Gbps.[45] Microelectro mechanical switching (MEMS) is currently poised to leapfrog current switching technology and render petabit capacities in the very near future. As switch technology advances, and assuming sufficient economic incentives remain, the Internet may evolve to the point where a task might transparently draw on data or processing power anywhere. Moreover, a network of ground stations might select an optimal station for each data transfer with a spacecraft enabling free-space optical communications without concern for occasional interference from the weather.

For communication between spacecraft, the defense agencies have taken a lead in demonstrating optical communication links including those using passive transponders. These passive units could be used in low-power micro satellites in a sensor web.

INFRASTRUCTURES AND THE GRID

There is widespread interest in developing architectures that support science and engineering using distributed computing, as manifested in a number of “Grid” architectures currently underway.[46] The Grid architecture research focuses on large-scale resource sharing, innovative applications, and, in some cases, high-performance storage and computing. Examples of Grids that are presently being built include NSF’s National Technology Grid, NASA’s Information Power Grid, DOE’s DISCOM, GriPhyN, NEESgrid, Particle Physics Data Grid, and the European Data Grid[47]). Some of these architectures (e.g., the Information Power Grid) show promise in providing the infrastructure for an intelligent data system, e.g., the Intelligent Archive.

If Grid technology matures and remains available in the time frame of the Intelligent Archive, then the IA could be constructed as a diverse set of intelligent components, such as archives, sensors, models and applications, all set within, and linked together by, the Grid infrastructure. The Grid infrastructure itself might also embody intelligence, for example, in making decisions about whether to move data to algorithms, algorithms to data, or both to a high-speed computing facility. However, it is important to note that Grid research and development is targeted toward the interaction of the various software components in a distributed data system. This leaves much research to be done on the intelligent components themselves.

The whole infrastructure for systems and applications in a knowledge enterprise rests on core services provided by the physical assets, computers, storage, communications, etc. External systems can interface with the infrastructure as dedicated (private/local) or shared (virtual) resources. Two kinds of interfaces exist: one between the infrastructure and the systems/applications and one that integrates physical resources into a collective virtual resource. A key objective of the Grid architecture is the transparent, adaptive, dynamic sharing of resources in a highly distributed environment. Grid technologies provide a generic approach to resource sharing that is applicable to many applications.[48] Figure 4-2 relates Grid architecture to an overall infrastructure for IA and cooperative systems. Systems such as Intelligent Archives interact with computing infrastructures through a sophisticated interface. Grid protocols, services, and APIs including other open standards for Internet and Web interfacing become this high level interface. A closer inspection of Grid architecture helps illustrate how physical computing resources used by an IA for example, are organized, virtualized, and made accessible for shared use.

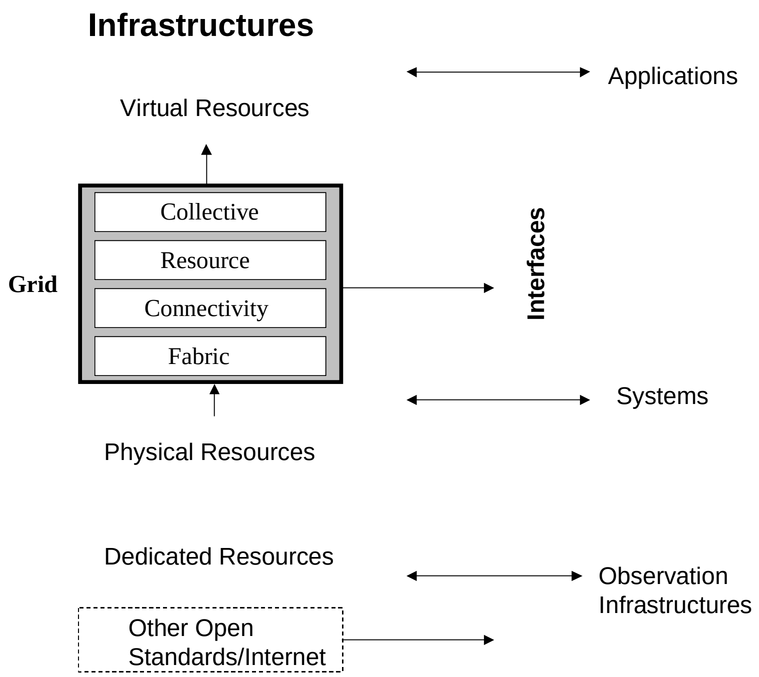

Figure 4-2: Resource Infrastructure and a Layered Grid Architecture

In a Grid architecture the *Fabric* layer provides the resources for which applications have shared access mediated by Grid protocols operating through the *Connectivity* layer.[49] The *Connectivity* layer defines core communication and authentication protocols required for Grid-specific network transactions. The *Resource* layer builds on Connectivity layer to define protocols for the secure negotiation, initiation, monitoring, control, accounting, and payment of sharing operations for individual resources. The *Collective* Layer, contains protocols and services that are not associated with any one specific resource but rather are global in nature and deal with interactions across collections of resources.[50] This contributes to the virtualization of resources. The important feature of the Collective layer is that it can deal with a wide variety of sharing behaviors without placing new requirements on the resources being shared. It is at this layer where there is the potential for adaptive management and control.

The evolution of Grid architectures is likely to be driven in part by the evolution of its current and potential software components as those components add new and more intelligent services. Thus, IA concepts and design strategies could help drive Grid functionality to be able to:

- Adapt to expected revolutionary changes in information technology
- Utilize intelligent processes at all layers so that the enterprise is able to operate in a semi autonomous fashion to the point that it is self maintaining and self correcting
- Include sensors and simulation models as members of the virtual organizations
- Support virtualization of services and functions seamlessly across the entire enterprise

LONG-TERM SUPPORTING TECHNOLOGY

Examining longer term supporting technology becomes substantially more difficult, but by estimating the functionality and performance of technologies based on historical information and possible future science and engineering achievements one can see interesting possibilities. The actual growth in performance expected from future systems may well be highly non-linear, may involve discontinuities, and may not accurately be predictable. Forecasting accurately beyond a few years into the future appears to be extremely risky. Jack Worlton[51] shows that even when using the "Delphi" technique (a group of experts answers questions and circulates their responses, iterating the process until opinions converge), forecasting accuracy fell exponentially with time, decreasing to 0.5 in 4 years. Extrapolation the analysis data indicates that accuracy falls to 0.3 in 6 years, 0.2 in 8 years, and that 10-year forecasts would be accurate in only one event in seven.

The most successful way of forecasting technologies in 10 to 20+ year time frames is to observe historical technology changes and apply similar rates of change. Using such a method allows one to accommodate the effects of paradigm shifts (i.e., revolutions) in a technology (e.g.., going from tubes to transistors). What this means for the world of 2025 is, that based on the rate of change in the 1990s, one can expect to see a doubling of progress each decade [starting in 2000]. A diagram that presents a technology timeline of forecasted changes is shown in Figure 4-3 (see adjoining page).

A glimpse of the possible effects of "revolutionary technology advances" can be observed in Figure 4-3 and the "blending" of several key technologies including Nanotechnology, Biotechnology, Information Technology, and Cognitive Science (NBIC).[52] These co-evolving technologies are expected to have a profound impact on improving the performance and enabling new functionalities of future intelligent archive components and systems. It was speculated in a recent report on converging technologies by the NSF that:

> "In the early decades of the twenty-first century, concentrated efforts can unify science based on unity in nature, thereby advancing the combination of nanotechnology, biotechnology, information technology, and new humane technologies based in cognitive science. With proper attention to ethical issues and societal needs, converging technologies could determine a tremendous improvement in human abilities, societal outcomes, the nation's productivity, and the quality of life. This is a broad, cross cutting, emerging, and timely opportunity of interest to individuals, society and humanity in the long term."[53]

One of the key challenges of an IA will be its ability to host and support the intelligence-based algorithms, being developed by the Intelligent Data Understanding area of the Intelligent Systems program. These algorithms are being developed to perform functions that are currently in a research stage. Continued development of these research algorithms is essential for automating the means to manage and make use of rapidly increasing data acquisition rates, data complexities, and data volumes. In such an environment, it is unrealistic to expect success without greatly improved system performance. Improved performance will depend a great deal on the underlying hardware and software technologies of an IA. Figure 4-3 appears to assure us that the necessary underlying powerful technologies will indeed be there, but possibly in very unfamiliar forms.

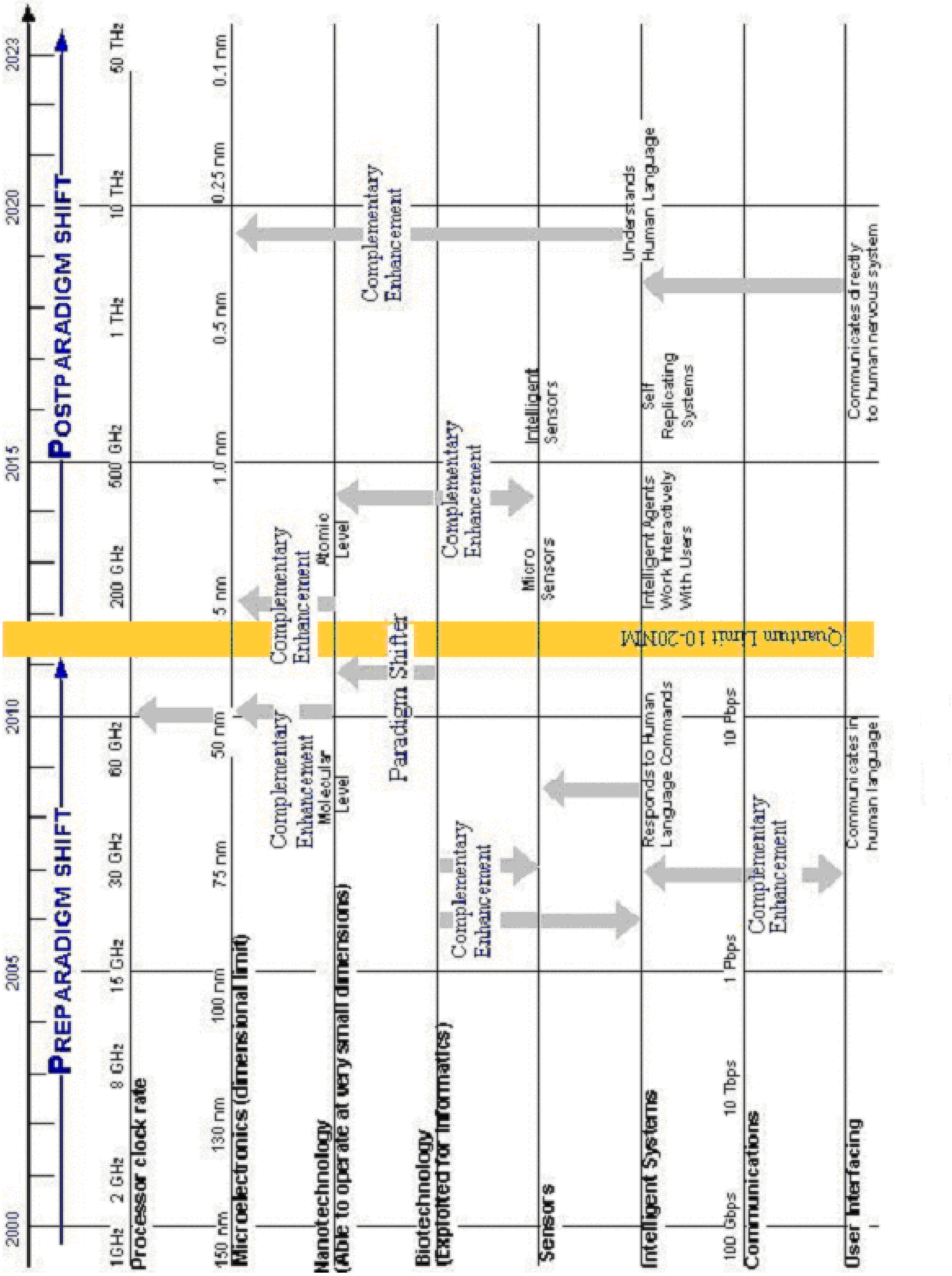
2000
2005
2010
2015
2020
2023
PREPARADIGM SHIFT
POSTPARADIGM SHIFT
1GHz
2 GHz
8 GHz
15 GHz
30 GHz
60 GHz
200 GHz
500 GHz
1 THz
10 THz
50 THz
Processor clock rate
150 nm
130 nm
100 nm
75 nm
50 nm
5 nm
1.0 nm
0.5 nm
0.25 nm
0.1 nm
Microelectronics (dimensional limit)
Nanotechnology (Able to operate at very small dimensions)
Complementary Enhancement
Molecular Level
Complementary Enhancement
Atomic Level
Paradigm Shifter
Complementary Enhancement
Biotechnology (Exploited for Informatics)
Complementary Enhancement
Complementary Enhancement
Sensors
Micro Sensors
Intelligent Sensors
Quantum Limit 10-20NM
Intelligent Systems
Responds to Human Language Commands
Intelligent Agents Work Interactively With Users
Self Replicating Systems
Understands Human Language
100 Gbps
10 Tbps
1 Pbps
10 Pbps
Communications
Complementary Enhancement
User Interfacing
Communicates in human language
Communicates directly to human nervous system

Figure 4-3: Technology Timeline

5. SUMMARY AND RECOMMENDATIONS

SUMMARY

This study has developed a conceptual architecture for future data archive systems that will efficiently manage large volumes of NASA data, support the extraction of information and knowledge by intelligent algorithms, facilitate scientific use of these data resources, and demonstrate feasibility of near-real-time utilization of the derived knowledge by NASA's operational partners. NASA's ESE combines NASA-unique capabilities for acquiring, storing, and distributing unprecedented amounts of data and information products to Earth science research communities. Forward-looking plans and science goals for improving predictive capabilities for weather, climate, and natural hazards continue to drive the trends toward producing and using ever increasing amounts of complex data. Astute exploitation of advancements in technology are integral to achieving these goals. The same is true for using technology in a coherent manner (e.g., architecture-oriented) to address the operational challenges of managing future distributed data intensive science.

The conceptual architecture presented in this report establishes a framework in which an *Intelligent Archive* generalizes the conventional archive from a simple repository of data to one that supports and facilitates the derivation of information and knowledge. An IA has greater ability to operate autonomously but participate cooperatively with other systems and applications than conventional archives. It also can provide new and better services to users (be they sensors, applications, or other production systems) with less operator intervention, more stability, and higher quality with less expensive human overhead.

The algorithms that automate the transformation of data to information and knowledge are currently being developed within the IDU area of the IS Program. The IA can be the system context for integrating and configuring many of these algorithms. It can also play a key support role in an end-to-end knowledge building system by:

- Storing and managing full representations of data, information, and knowledge
- Building intelligence about transformations on data, information, knowledge, and accompanying services involved in a scientific enterprise
- Performing self-analysis to enrich metadata that adds value to the archive's holdings
- Performing change detection to develop trending information
- Interacting as a cooperative node in a "web" of other systems to perform knowledge building (where knowledge building involves the transformations from data to information to knowledge) instead of just data pipelining
- Being aware of other nodes in the knowledge building system (participating in open systems interfaces and protocols for virtualization, and collaborative interoperability)

We have described the architecture of the IA in sufficient detail to distinguish its place and role in the end-to-end system. We have then investigated this role by developing three operational scenarios in some detail. Next, we have reviewed available technology to identify areas where additional progress would be required to implement an IA.

Our study of the IA has demonstrated the possibility of a new synthesis of ideas and technology. This new synthesis promises unique benefits for science, but it cannot be implemented without more research and effort. Consequently, we will close this report by recommending attention to certain key technical issues, continuing effort on IA studies, and initiation of actions outside the IA study.

KEY TECHNICAL ISSUES

The preceding report has mentioned a large number of issues that must be resolved before an IA can be built. Several key technical questions remain open and require further analysis:

- Projected throughput requirements of the IDU algorithms based on quantitative analysis in the context of an intelligent archive are currently unknown

- Integration of grid technologies with the IA's "knowledge building system" is not yet clear

- Alternatives regarding physical architectures that instantiate the functional architecture, starting from sensorwebs continuing through predictive models, and then delivery via future mobile technologies require more analysis

- Interfaces to emerging software architectures, such as e-commerce with science software, command-and-control vs. peer-to-peer negotiation paradigms, data- vs. software- mobile paradigms, collaborating micro-sensor data collection, and automated quality assessment, etc. require more analysis

CONTINUING STUDIES

At the present time, study efforts continue in the following areas:

- Analysis of what may be expected from commercially available enabling technologies in order to determine the gaps that NASA should address

- "Drill-down" on a few key technical issues that have surfaced in the preceding study, including:

    - Autonomous QA by and within an Intelligent Archive

    - Intelligent Resource Optimization

    - Automated Detection and Integration of New Sources of Data

    - Virtual Data Products using On-Demand Processing

    - Architectural Topologies

    - Data Transformations

    - Management dimensions of an intelligent archive (e.g., virtualization, self-awareness, autonomic concepts)

- Investigation of a potential partnership-based prototype to explore the concepts of the Intelligent Archive

RECOMMENDED ACTIONS

The following actions are outside the scope of this study, but are recommended to further the understanding and lead towards implementation of the IA:

- Characterize future data access, archiving, and utilization of scientific data within both scientific and applications enterprises. This characterization should attempt to specify "quantitative goals" that the computer science research community could solve for NASA

- Develop an implementation strategy that can guide IA prototypes developed by partnerships between NASA, industry, and universities into eventual operational implementations

REFERENCES

[1] NASA, 2000a, Understanding Earth System Change: NASA's Earth Science Enterprise Research Strategy for 2000-2010, http://www.earth.nasa.gov/visions/researchstrat/ResearchStrat.doc

[2] NASA, 2002a, Earth Science Technology Office (ESTO), http://esto.gsfc.nasa.gov

[3] NASA, 2002b, Advanced Information System Technologies (AIST) Capability and Needs, 2002 Version, http://esto.gsfc.nasa.gov/programs/aist/

[4] NASA, 2002c, Earth Science Enterprise Applications Strategy for 2002-2012, January 2002.

[5] NASA, 2000b, Exploring Our Home Planet, Earth Science Enterprise Strategic Plan, November 2000, http://www.earth.nasa.gov

[6] Schnase, J. L., J. A. Smith, T. J. Stohlgren, S. Graves, C. Trees, 2002, Biological Invasions: A Challenge in Ecological Forecasting, IEEE IGARSS and the 24th Canadian Symposium on Remote Sensing, Toronto, Canada, June 24-28, 2002.

[7] Chao, B. F., T. Farr, J. LaBrecque, R. Binschadler, B. Douglas, E. Rignot, C. K. Shum, J. Wahr, 2002, Understanding Sea Level Changes, IEEE IGARSS and the 24th Canadian Symposium on Remote Sensing, Toronto, Canada, June 24-28, 2002.

[8] Raymond C. A., P. R. Lundgren, S. N. Madsen, J. B. Rundle, 2002, Understanding and Responding to Earthquake Hazards, IEEE IGARSS and the 24th Canadian Symposium on Remote Sensing, Toronto, Canada, June 24-28, 2002.

[9] Ramapriyan H. K. and G. R. McConaughy, 1991, "Version 0 EOSDIS - An Overview," Technical Papers, 1991 ACSM-ASPRS Annual Convention, Baltimore, MD, March 1991.

[10] Ramapriyan, H. K., 2002, Satellite Imagery in Earth Science Applications, Chapter 3 in Image Databases: Search and Retrieval of Digital Imagery, V. Castelli and L. D. Bergman, editors, John Wiley and Sons, NY, pp. 35-82.

[11] Moore, M. and D. Lowe, 2001, Leveraging Open Source Development in Large Scale Science Data Management Systems, Proceedings of SPIE, Earth Observing Systems VI, San Diego, CA, August 1-3, 2001, Vol. 4483, pp. 291-300.

[12] Moore, M. and D. Lowe, 2002, Providing Rapid Access to EOS Data via Data Pools, Proceedings of SPIE, Earth Observing Systems VII, Seattle, WA, July 7-10, 2002.

[13] Olsen, L.M., 2000, "Discovering and Using Global Databases," in Global Environmental Databases: Present Situation; Future Directions, R. Tateishi and D. Hastings, editors, International Society for Photogrammetry and Remote Sensing (ISPRS).

[14] Smith, S. G. and R. T. Northcutt, 2000, "Improving Multiple Site Services: The GCMD Proxy Server System Implementation," EOGEO 2000. Earth Observation (EO) & Geo-Spatial (GEO) Web and Internet Workshop, April 17-19, 2000. http://gcmd.nasa.gov/conferences/EOGEO2000/Smith_Northcutt.html

[15] Pfister, R., 2001, "The Information Management System of NASA's EOS Data and Information System," in Proceedings of the American Meteorological Society (AMS) 81st Annual Meeting, 17th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Albuquerque, NM, January 14-19, 2001, pp. 161-163.

[16] McDonald, K. R., R. Suresh, L. Di, 2001, The Data and Information Access Link (DIAL), in Proceedings of the American Meteorological Society (AMS) 81st Annual Meeting, 17th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Albuquerque, NM, Jan 14-19, 2001, pp. 166-166.

[17] Cornillon, P., 2000, Personal Communication, December 2000.

[18] Frew, J., M. Freeston, L. Hill, G. Janée, M. Larsgaard, Q. Zheng, 1999, "Generic Query Metadata for Geospatial Digital Libraries," Proceedings of the Third IEEE META-DATA Conference, http://www.computer.org/proceedings/meta/1999/papers/55/jfrew.htm

[19] Pfister, R., 2001, "The Information Management System of NASA's EOS Data and Information System," in Proceedings of the American Meteorological Society (AMS) 81st Annual Meeting, 17th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Albuquerque, NM, January 14-19, 2001, pp. 161-163.

[20] NASA, 1999, "1999 EOS Reference Handbook: A Guide to NASA's Earth Science Enterprise and the Earth Observing System," M. King and R. Greenstone, editors, NASA Goddard Space Flight Center, http://eos.nasa.gov/

[21] ESIPFED, 2002, http://www.esipfed.org/

[22] Smarr, L., 1998, The Emerging National Technology Grid: Using High-End Computational Science to Predict the Broad-Based Future, http://www.jacobsschool.ucsd.edu/~lsmarr/talks/San%20Diego/index.htm

[23] Thigpen, W., 2002, http://www.ipg.nasa.gov/

[24] DISCOM, 2002, Distance and Distributed Computing and Communication, http://www.cs.sandia.gov/discom/main.html

[25] Avery, P., and I. Foster, 2002, Grid Physics Network (GriPhyN) Project Description," On line paper available at www.griphyn.org

[26] NEESgrid, 2001, http://www.neesgrid.org/

[27] PPDG, 2002, http://www.ppdg.net/

[28] NASA, 2002b, Advanced Information System Technologies (AIST) Capability and Needs, 2002 Version, http://esto.gsfc.nasa.gov/programs/aist/

[29] NASA, 2002d, Computing, Information and Communications Technology Program, Intelligent Systems Project, http://is.arc.nasa.gov/

[30] IDU 2002, Home Web Page for the Intelligent Data Understanding Element of the Intelligent Systems Program, http://is.arc.nasa.gov/IDU/IDU.html

[31] Quatrani, T., and G. Booch, 1998, Visual Modeling with Rational Rose and UML, pub. Addison-Wesley.1998, pg. 43.

[32] ibid.

[33] Lockheed Martin Advanced Concepts Center and Rational Software Corporation, 1996, Succeeding with the Booch and OMT Methods: A Practical Approach, Addison-Wesley, 1996.

[34] Matthews, Robert, 2001, “Don’t Blame the Butterfly,” New Scientist, August 2001, pp. 25-27.

[35] Steiner, Mark, 2001, “ESTO Weather Prediction Technology Investment Study,” October 5, 2001.

[36] ibid.

[37] Fender, Toni, 2002, “Astronomers Envision Linking World Data Archives,” Physics Today, February, 2002, p. 20.

[38] National Research Council, 2001, Astronomy and Astrophysics in the New Millennium, National Academy Press, Washington, D.C.

[39] Szalay, A., and J. Gray, 2001, “The World-Wide Telescope”, Science, Vol. 293, September 14, 2001, p. 2037.

[40] Brunner R.J., S.G. Djorgovski and A. S. Szalay A.S (eds.), 2001, 2001 Virtual Observatories of the Future, Astronomical Society of the Pacific Conference Series,
San Francisco, California, p. 357.

[41] NASA, 2002e, IS 2002, Home Web Page for the Intelligent Systems Program, http://is.arc.nasa.gov/index.html

[42] Gniewek, J. and S. M. Vogel, 1995, Influence of Technology on Magnetic Tape Storage Device Characteristics, GSFC Mass Storage Systems and Technologies conference, March 1995, http://esdis-it.gsfc.nasa.gov/MSST/conf1995.html

[43] Schwarz, T, 2001, Magnetic Tape (Emerging Technologies Panel), Eighteenth IEEE Symposium on Mass Storage Systems,
http://storageconference.org/2001/EmergTechPanel/Schwarz.pdf

[44] Rotman, D., 2001, "Nanotech Goes To Work," Technology Review, January/February, 2001, pp 62-68.

[45] Cross, T., 2001, "Seeing the Light of the Optical Tunnel; The Future of Optical Fiber and Switches," Boardwatch Magazine, April, 2001,
http://program.intel.com/solutions/shared/en/resources/insight/techtrends/optical.htm

[46] Grid Computing - Grid 2001: Second International Workshop Denver, CO, November 12, 2001, Proceedings (Lecture Notes in Computer Science, 2242), Springer Verlag; ISBN: 3540429492; (February 2002).

[47] Waldrop, M., 2002, "Grid Computing, Could Put The Planet's Information-Processing Power On Tap," Technology Review Magazine May 2002, pp 31-37.

[48] ibid.

[49] Foster, I., Kesselman, C., Nick, J., Turecke, S., 2002, "Grid Services for Distributed System Integration, " IEEE Computer June 2002, Pages 37-46.

[50] Foster, I., Kesselman, C., Nick, J., Turecke, S., 2001, "The Anatomy of the Grid," International Journal of Supercomputer Applications.

[51] Worlton, J., 1998, "Some Patterns of Technological Change in High-Performance Computers, " Worlton & Associates, September, 1988.

[52] Roco, M. C. and W. S. Bainbridge (editors), 2002, National Science Foundation Converging Technologies for Improving Human Performance, National Science Foundation and The Department of Commerce, NSF/DOC Sponsored Report, June 2002.

[53] ibid