Intelligent Archive Visionary Use Case: Precision Agriculture Scenario

September, 2003

Robert Harberts, L. Roelofs
Global Science & Technology, GST
Greenbelt, MD
Harberts@gst.com

H.K. Ramapriyan, G. McConaughy, C. Lynnes, K. McDonald, S. Kempler
NASA/GSFC
Greenbelt, MD 20771

ABSTRACT

A NASA sponsored study team is actively developing new concepts for an intelligent archive (IA). The conceptualized IA possesses innovative capabilities for managing data, making them easily available, and for improving utilizing unprecedented volumes of data in coming decades. The study addresses the problem of getting the most societal benefit from data volumes that NASA expects to continue to accumulate in increasing rates from future missions. Petabyte-era data volumes pose a variety of challenges to existing archival and distribution practices. The IA study team has adopted use case scenarios as an approach to better understand ways to address these challenges as part of its IA architecture study. Use case scenarios represent an effective tool for exploring many of these challenges from different application and operational perspectives. Describing a particular use scenario affords a means to study the capabilities and functions for future systems with which to derive challenging new requirements applicable to visionary systems. Study of use cases also help indicate opportunities for applying intelligent systems research and where systems intelligence can be integrated into an archive.

We describe a use case scenario based on future precision agriculture operations scoped to the size of a single farm. The scenario consists of an exploration of data, information, and cyber-system capabilities that would be utilized in future precision agricultural practices. This affords an end-to-end context for studying roles for intelligent systems and from which to test IA concepts with challenging new requirements. We introduce concepts for a “visionary” digital farm that provides powerful information services to a farmer in support of crop planning, cultivating, and harvesting. The digital farm is conceived as a comprehensive information service for the farmer that is highly responsive and relevant to his daily, near-term, and long-term operational information needs. An IA supported digital farm provides farmers with the right information while simultaneously hiding the complexities associated with producing them. Given the advantages of the digital farm we next examine the implications of the precision agriculture use case on intelligent archives and associated data volumes. We recognize that the value of a virtual farm application to a grower is largely realized in how it acts as an intelligent interface between the information needs of the application domain and the resources provided by intelligent archives through an evolved cyber infrastructure. Overall this is a use case for greater data utilization.

TABLE OF CONTENTS

INTRODUCTION

NASA expects to accumulate very large volumes of earth and space science data in the next decades as a consequence of advances in observational instrumentation capabilities. More powerful space-based and terrestrial sensors (ranging large to nano-scale) including new strategies for deploying distributed and networked sensors will be responsible for a host of observations rendered as digital data. Managing these data while simultaneously extracting the most societal value from them pose significant challenges for designers of future archive systems. Studies that look deeply into the next generation of data archiving systems are yielding new concepts to address these challenges. A study team funded under the NASA Ames CITC Intelligent Systems program has been developing the concept of an Intelligent Archive (IA). The IA team studies the role of an IA within the context of an end-to-end scientific enterprise. This context embraces a chain of systems extending from those involved in data acquisition to those that derive information and knowledge from those observations and measurements for use by various societal enterprises. Implications of terabyte and petabyte-era data volumes on systems and infrastructures that efficiently support science-based applications stimulated research into new concepts for intelligent archives in the years 2012-2020.

Beyond the obvious dependencies on improvements in hardware technologies to meet requirements for storing and supporting efficient access to future distributed data volumes, advances are needed in concepts and tools to enable intelligent data management and utilization. While the growth rate and quantity of data increases rapidly, new challenges are posed to basic data management and utilization. Key challenges that characterize future data volumes and use include [4.]:

- 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 IA team has developed a conceptual architecture with new intelligent system-based capabilities for future archives to address these challenges. Part of the process for developing IA concepts included the use of scenarios describing data usage in applications combined with projections of advances in relevant technologies. In this manner an abstracted architecture could be defined without regard to physical implementation. Architecture for future archives can therefore be considered from the point of view of the functions that need to exist in support of various usage scenarios and overall requirements. This points out that the functions of an intelligent archive are more stable than the physical architectures and technologies used to implement them. By 'discovering' and abstracting required components and processes from scenarios into functional elements, we were able to explore application strategies of technologies and system resources for future intelligent archives.

An outgrowth of this approach is a definition of characteristics that describe future intelligent archives. An IA is different from contemporary archives because the scope of meaning for the term 'archive' is extended from a simple repository of data to one that supports and facilitates derivation of information and knowledge. It supports greater data utilization integral to knowledge building systems. For example, items managed by an IA include [1]:

- 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

With the application of intelligent algorithms and intelligent system components, an IA has greater ability to operate more autonomously than conventional archives. This characteristic reflects higher degree of automation and self-managed operations. Accordingly, an IA is envisioned to provide highly responsive and accurate services to users (e.g. as an intelligent assistant) with less operator intervention. By embodying new capabilities like these an IA can be distinguished from archives of today with regard to its ability for [1]:

- 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)

While internal capabilities of the IA could be characterized and described like those above we recognized that a broader contextual perspective was necessary to further evolve the concept. For instance, the role of an IA in an environment of data providers and data and information consumers remained unclear. Questions about how an IA interacts as an interoperable and cooperative component in an overall distributed system environment proved initially ambiguous. We elected to develop use case scenarios to study how different application users could benefit from IA capabilities end-to-end.

USE CASE SCENARIOS FOR IA CONCEPTS

Use case scenarios help illustrate, exercise, and clarify IA concepts. Scenario-based approaches to developing intelligent archive concepts also help derive clearer understanding of end-to-end interrelationships among data and information, consumers, data providers, value-added information services, data archives, and data acquisition missions that underlie an IA [2],[3]. Furthermore, use-scenarios provide better insights into limitations of existing solutions, and the nature of challenges, issues, and probable opportunities for innovation. Scenarios of conceptualized ways to use intelligent archive resources help uncover a range of requirements for services and capabilities that can be mapped to existing and future technology applications. Consequently, forward looking, tangible and imaginative scenarios lead to a flexible, more comprehensive architecture for future intelligent system applications.

By starting with an existing realistic set of requirements we explored ideas from which to extrapolate the possible and then develop more speculative and visionary scenarios. Each was studied with regard to advanced data/information requirements and the challenge of assimilating progressively massive amounts of available data. Studying a conceptual application domain contributes to an appreciation of both context-specific requirements and those for common or generalized features. We first considered the current domain for data archive systems. With a basic model of data archive systems operating in a scientific knowledge building domain where data acquisition, preservation, access, distribution, and utilization are involved, we related near-term, next-generation, and future requirements to use case scenarios. The generalized model of common features that emerged from this study consists of four enterprise elements or tiers:

- Sources: includes missions, instruments, and platforms for geo-spatial observations acquired by data receiving stations.

- Providers: includes primary data processing, secondary processing and product production, services, archiving, and access/distribution channels.

- Value Chains: includes secondary and other data product provider service groups as part of end-to-end value-added services.

- Consumers: includes data and information users possessing challenging requirements for geo-spatial data products/services.

Each tier represents a high level stakeholder in the enterprise that has its own evolving requirements for information technology capacities (e.g., throughput, storage volumes, and communication bandwidth), capabilities, inputs, outputs, and interfaces. Interdependencies and various patterns of relationships (e.g. functional, transactional, and operational) exist within and among the enterprise tiers. Evolving requirements among the enterprise elements affect the nature of these interrelationships and dependencies. By considering the implications requirements for future archives have on supporting technologies and, conversely, how technological advances might be exploited for interoperable elements of enterprise systems we were able to characterize some visionary use cases.

In the following sections we present a use case study based on precision agriculture. The first part introduces the use case context with illustrative scenarios. We also describe the concept of a digital farm that possesses advanced operational capabilities with which to explore requirements and roles for an IA. The next sections examine the use case's scope requirements for data, information, services, capabilities, and implications of these on stakeholders. Following these discussions we examine implications of the use case on the IA including a quantitative assessment of virtual farm data requirements.

PRECISION AGRICULTURE USE CASE STUDY

Precision agriculture is concerned with techniques and practices that maximize crop yield with minimum cost through improved use of data, information, and knowledge technologies. In this use case study we explore precision agriculture from the perspective of the consumer or "end-user" (e.g. fourth tier stakeholder). Special attention is given to data and information services useful in precision agriculture involving an individual farm. Additional stakeholder viewpoints on requirements include agribusiness, various government agencies, and research communities. By beginning with a single farm it is possible to study scaled up scenarios involving a grid of connected farms aggregated into regional, national or even global levels of cooperation. We start with a use case scenario based on the scope and parameters of a minimal agricultural production unit with regard to planning, cultivating, and harvesting crops.

USE SCENARIO: GROWER AND FARM

The farm is characterized as a relatively small spatial area (considered in acres) for agricultural products suited to regional ecological, weather, and growing constraints. Examples of farms vary regionally such as a Midwestern 360 acre farm for corn and soybeans; a South Dakota 2000+ acre wheat field; an average California Central Valley 800 acre vegetable and fruits farm; and a Montana 10000+ acre cattle ranch. Geo-spatial information relevant to these farms puts a premium on high-resolution data both spatially and temporally. We will use the scenario to generate challenging requirements with which to exercise the conceptual IA architecture and better understand the dynamic interactions among its components and functions

PLANNING, CULTIVATING, HARVESTING

Quality data, information, and knowledge are highly important for conducting precision farming activities like crop planning, cultivation, and harvesting. Information-intensive support services required for each activity include monitoring current conditions, collecting histories and time series studies, running trends/risks analysis, making predictions, forecasts, "what-if" investigations, and harvest outcome comparisons. Detailed information about land, weather, water, agriculture markets, prior yields, agricultural-chemical options, seeds, etc. is required for planning crop selection and planting. High-resolution information is required to monitor crops, assess risks, and make decisions about appropriate interventions to maintain crop health. Similarly, to maximize yields, decisions about timing harvests require information about current and future conditions. Remotely sensed information about farm assets including information collected from the farm about outcomes of plans, cultivation techniques, and harvests, is integrated within a farm's database for perpetual use.

VISION CONCEPT: DIGITAL FARM INTERFACE

Humans perpetually invent ways to reduce work, or at least to make it more manageable. While mechanical machines alleviate physical work, digital machines address information-age burdens. The future of farming exploits advantages from both types of machines. Farming represents a dynamic interplay among machines, information, and human knowledge-based techniques employed in the critical pursuit of food production. Precision farming research is one such frontier where applying information technologies promises to improve operational efficiency while simultaneously maximizing crop yields.

The concept of a digital farm merges applications of information technologies, digital information, GIS, GPS, archives, value added-providers, sensors, and human-machine interfaces (see Figure 1). Quality high-resolution geo-spatial data are integral to precision agriculture and the decision-making processes practiced within it. The following speculative scenario describes how future growers might access and use specialized information selected from vast data stores that are constantly updated with hourly observations. The grower's information resources represent the consequence of interoperating services, value-added processes, automation, and filtering of data of specific relevancy to the farmer. The following scenario illustrates how a digital farm might operate.

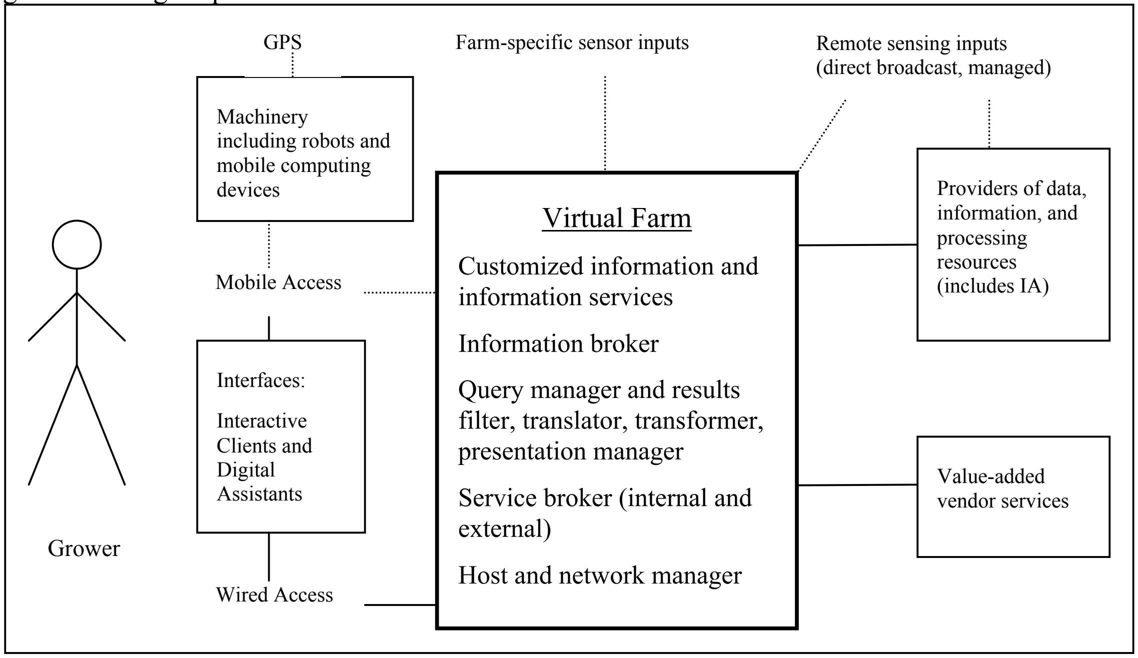

Figure 1: Conceptual Virtual Farm Context

DIGITAL FARM VISION SCENARIO

When a South Dakota wheat grower rises early each morning, part of his or her morning routine involves perusing world news, local news, and a "news" review of the entire wheat farm. Summary and detailed information about every aspect of the farm, including past, present, and future, is made easily available through interactive high-resolution panels. A digital assistant or "agent" works on behalf of the grower to collect and provide valued information services. In the background mode this artificial assistant actively monitors sources of data and information for updates to harvest and constantly update in the digital farm database. An interactive mode is also available for the grower to speedily access ad hoc information. The digital assistant is accessible from any interface (e.g. workstations and mobile devices) within the house, farm buildings, vehicles, or even the combine. Interaction with the digital assistant can be conducted by natural language via voice or touch screens and keyboards.

"Intelligent" digital assistant software can interpret, broker, and fulfill requests for information dynamically and autonomously. This feature of a digital farm skillfully performs its role by brokering and invoking services from a digital virtual wheat farm database that contains encyclopedic farm-relevant information ontologically, spatially, and temporally organized. The digital farm automatically ensures that its information resources about soil, crop, weather, and moisture conditions are constantly updated. To do so digital farm software has interfaces with external data and information sources for input as well as with farm-specific sensor inputs (see figure 1 Virtual Farm links). These functional interfaces are crucial to pooling farm-relevant data from multiple data sources such as primary archives and agricultural "value-added" services.

The digital assistant can also produce different views of this information by summoning an array of digital functional services. These virtual farm services produce virtual 4-D model (i.e. topographic with temporal perspectives) of the entire farm that the grower can inspect from his or her office or combine cab monitor for example. 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 sequences. Types of information range from historical to actual current conditions to what-if scenarios cast into future periods of interest. Because the grower's digital farm can "learn" from his or her queries and interests, the content and services it provides adapt with change and specificity over time.

Some of the farm machinery also interacts with the digital farm information services (see figure 1). Autonomous and semi-autonomous machines that plant, cultivate, and harvest crops are precisely controlled with a combination of GPS, distributed functions, and data from the digital farm. Optimal applications of seeds, fertilizers, and chemicals can be controlled and recorded via wireless digital farm services. Similarly, data taken from the field during cultivation and harvesting can be relayed to the digital farm as input for archiving and further use.

Overall the digital farm concept represents a sophisticated intelligent data and information pool. It serves as an interface between specific interests about the farm and vast external data sources. It also serves to integrate and administer different kinds of extracted (i.e. mined) and filtered information. Digital farms also provide the basis for intelligent interfaces for machine-to machine (e.g., smart farm machinery, digital assistants) and human-machine interactions.

DIGITAL FARM: CONSEQUENCES

We speculate that within fifteen years, digital information services and systems will be integrated into many automated agricultural functions. Farm functions requiring improved accuracy, risk mitigation, production optimization, safety, environmental, and regulatory compliance will be managed better with higher levels of integration with information services. The agricultural production unit could also be integrated into a collaborative and adaptive system that efficiently shares information for the benefit of all agricultural enterprise stakeholders. In the next decades both i*n situ* and remote sensing observations/measurements should be smoothly interfaced with data assimilation services that support a variety of private, public, and cooperative concerns. Demonstrated improvements in production and quality of agricultural products continue to be realized, thus reinforcing further evolution of the agri-information and agri-business enterprise.

Another consequence of digital farms includes how they usher in an era characterized as highly integrated agriculture – digital farms containing local precision agricultural data and histories could be subsumed into a connected grid of many farms aggregated into county, state, regional, national, and perhaps even global farming sectors. An agricultural grid of services and interchangeable information could support many levels of functional cooperation involving members with individual operational parameters to those of large collective organizations. Furthermore, it is conceivable that such a grid of agricultural stakeholders could itself be interfaced with economic grids that link markets to enterprises in new ways. It may be common practice in this future time for interconnected agri-economies, businesses, and world markets to transact business based on clearer pictures of actual crop production status summarized at various levels of aggregation down to the individual production unit.

In the next section we turn our attention to implications of the digital farm scenario on requirements for data, information, services, and functional capabilities to support precision agriculture. This aspect of the use case scenario study characterizes future data utilization by one application domain (i.e. precision agriculture) with which to explore implications for the intelligent archive concept.

SCOPE OF DIGITAL FARM REQUIREMENTS

Data from which to produce information used in the digital farm scenario are scoped to the physical parameters of an agricultural production unit. This means discrete agricultural production units are defined as relatively small geographic areas with specific boundaries nested into micro-ecological locales, regional habitats, and environments. Remotely sensed information must be relevant to the agricultural enterprise practiced by the grower and related to the agricultural assets of the farm. Precision agriculture relies on accurate, quality information keyed to geo-spatial resolutions of a meter or finer (to help differentiate among plants and distinguish plant conditions within the farm). Temporal resolutions often need to be hourly especially for weather information that is tailored for the locale and region of interest. Precision weather forecast requirements range between three hours ahead to months and perhaps even years. Examples of additional requirements for the content of a digital farm include:

- Land profiles that include soils, moisture, elevations, drainage patterns, and water sources.
- Digital Ortho-photo Quadrangles (DOQ) – digital images that coincide or integrate USGS topography maps – 1 meter resolutions
- Digital Elevation Models (DEM)
- Geo-rectified spatial data
- Ecological zone profiles including habitat, botanical profile, zoological profile, and supportable species including plants, insects, bacteria, viruses, animals, etc.
- Inventory of actual biological diversity
- Maps of contiguous environmental components and configurations
- Local calibrations including ground truth for valued and relevant agricultural assets with regular adjustments
- Local calibrations including ground truth for weeds, pests, plant-stress signatures, etc.

These requirements suggest that the content of a digital farm is largely composed of geo-registered land and weather data scaled to the physical dimensions of the farm and its immediate environment. Various services and functional capabilities are therefore required to acquire and transform data from various resources into information relevant to a particular farm. Part of this relevancy includes information that supports primary operations of the precision farming enterprise.

REQUIRED SERVICES AND CAPABILITIES

For a virtual farm to be an effective service to farmers it must possess a variety of functional capabilities that support precision agricultural operations. Key among those is the overall capability to broker modes of access to distributed sources of data and information. A virtual farm would be capable of invoking services sufficient to acquire, filter, and transform data into farm-relevant information. These include the capability to:

- Extract relevant information from data scaled to the interests of the specific farm
- Extract farm-specific information from models for forecasts and "what-ifs" projections
- Provide rapid turnaround for requested data/information ranging from complex to simple
- Broker information service requests with multiple service providers
- Provide easy to understand graphical presentations of information
- Communicate via wireless and wired connections

A variety of specialized services are also encompassed and provided by a digital farm. They play a role in supporting farm management activities with dynamic and interactive access to various data/information resources that can be chained into delivering customized information products. Examples of useful services identified in our study of the use case are listed below:

- Planting conditions indicators – predictive, actual, recommendations, interpretations
- Crop monitoring – daily observations that can be stored, interrogated, examined at various levels of detail, summarized, and shared
- Crop growth measurements – predictive, actual, interpretations
- Crop maturity – daily projections, actual
- Plant health indicators (stress, pests, infections) – actual, optimal, projected
- Identification and tracking of weeds and other infestations – type, growth, condition, spread, actual, projected
- Soil moisture indicators – historical, current and long-term, optimal vs. actual.
- Chemical dispersion impact – actual distribution correlated with other components

- Weather conditions – hourly monitoring of wind, temperature, precipitation, cloud cover
- Advance forecasts – daily, weekly, monthly, seasonal, annual, multi-year climate change
- Environmental alerts –concerns such as storm, hail, drought, fire, flood, freezes, snow
- Sunshine vs. cloudy predictors – correlation of projected weather with crop type options
- Microclimate zones and elevations – comprehensive survey of farmland environment
- Soil types and depths – mapping, survey, and correlation with ground cover and optimal crop types
- Plant census – distributions by type and location; current and historical
- Supportable biology – potential vs. actual; daily, seasonal, trends
- Stress signatures – causes and signs
- Interventions – chemical applications; actual, planned, historical, outcome analysis

From this list we ascertain that precision agriculture requires specific kinds of information for a variety of topics required throughout an operational year of a farm. Furthermore, such information in many cases needs to be extracted, derived, and processed to aid decision support. Timeliness, quality, and resolution of utilized information factors significantly into the quality of decision making expected when growers address issues confronting each aspects of sustaining farm productivity.

IMPLICATIONS OF OVERALL REQUIREMENTS ON COLLABORATIVE PRECISION AGRICULTURE STAKEHOLDERS

Growers equipped with a comprehensive agricultural GIS resource (i.e. digital farm) custom-oriented to the parameters, dimensions, and location of the farm, will depend on internal (farm sensors) and external sources for data and information services (e.g. NASA, non-NASA, and value-added providers). Principal among these sources would be intelligent archives and value-added providers of agriculture information services. Resource service providers of data and information would be required to make their service available twenty four hours a day by subscription and via dynamic requests (i.e. initiated by the grower, agents, and digital assistant).

Data and information is the common denominator for stakeholders in the precision agriculture enterprise. The flow of Information will be characterized more like a federation of sources and services that represent innovative, basic, value-added, and competitive services within and from outside the agricultural enterprise. For example, intelligent data archives of government-maintained remote sensing data could supply data directly to a virtual farm or to value-added information providers. Distributed and collaborative intelligent archives could exist among government, commercial, institutional (university), and private organizations that support virtual farms with information services. Increasingly, stakeholders will recognize the value of remote sensing and in-situ data for precision agriculture. This will encourage political and economic forces to define the evolution of stakeholder services and interoperations among them. In the next section we consider how an IA supports an enterprise and performs as a stakeholder in achieving effective data utilization.

INTELLIGENT ARCHIVE ROLE

An IA is intended to address relatively large-scale operational and functional data management and utilization requirements (e.g. primary remote sensing data archives, mission or discipline-specific archives, or custom commercial archives). Analyses of requirements from the precision agriculture use case context suggest a virtual farm's value stems from tailoring and customizing information germane to a farm. Information used by a farm will depend on applications of intelligent data understanding to extract, transform, and present the information it needs. A significant role for an IA is in distilling specific farm-relevant information from an avalanche of data. These transformations may involve complex process chains of data reduction and information/knowledge extraction services. Functional components for an IA to facilitate the distillation of information for a farm 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 predictions
- Algorithms for data products
- Visualization of preprocessed and dynamically processed data

These component technologies must be designed and implemented to interoperate throughout end-to-end processes for data production, quality assurance, cataloging, metadata extraction, data content characterization, data product archiving, and custom utilization tools. Automated data understanding is crucial for distillation, characterization, detection, and extraction of such things as objects, phenomena, features, conditions, and other knowledge-based subjects.

Data understanding within an IA can be illustrated another way with regard to alerts. Intelligent data understanding provides the means to automatically detect, recognize, and assess conditions or phenomena of concern, automatically formulate an alert message (with visualization as appropriate), automatically determine who should receive it, and send it via the most expedient channel(s).

An IA will also provide an interface through which it interacts and collaborates with other intelligent systems. One example involves system intelligence to accept and process semantic queries or broker requests to appropriate internal/external services. A chain of semantic filters and information retrievers comprise an IA functional layer for transacting requests. This layer may also filter relevant information on behalf of people (e.g. grower and crew) and present unambiguous, relevant information through interactive interfaces tailored to the concerns of the consumer. Additional implications of virtual farm requirements for an IA to address, considering an end-to-end perspective, include:

- Visualization, mining, analysis, and discovery tools support
- Digital virtual interfaces
- Digital assistants
- Query/result agents
- Interfaces to the Semantic Web
- Interfaces to a knowledge grid
- Distributed knowledge, information and data
- Archive-to-archive interoperability
- Distributed computation
- Distributed services
- Data intensive computing
- Remote analysis services
- Massive data volumes and complexities in data
- Responsiveness versus costs to extract useful knowledge and information

Another perspective on the role an IA performs in conjunction with the virtual farm concept can be gained from an assessment of virtual farm raw data requirements. In the following section we assess the kinds and potential quantities of data needed to fulfill informational services that a virtual farm would provide. Estimates for these data are based on examples and extrapolations taken from contemporary image data visualization and modeling sources.

VIRTUAL FARM QUANTITATIVE DATA REQUIREMENTS ASSESSMENT

Farms vary by size and production orientation (e.g., wheat vs. cattle, fruit vs. soybeans for example). Data requirements are estimated for a 1000 acre area farm with which variable size farms can then be calculated (e.g., actual acres/1000 x data estimates in bytes). Estimation of data requirements is difficult because they depend on various factors such as the number of sensors, data sources used, parameters required, and frequency of sampling. Some assumptions were required. Assessments of data for virtual farms are derived from several categories of data and information resources influenced by frequency of availability and need. Some initial working assumptions are:

Farm size: 1000 acres {Acre = 4047 m^2 (63.8x63.8m) * 1000 acre farm = 4047000 m^2 area = ~4km^2 area (or 2 km x 2 km farm)}

Remote sensing spatial resolution requirements: 1m, 10m, wider-range

Temporal resolutions: hourly, daily, history, 1-3 hour and longer-range measurements and forecasts

In situ sensors: mobile mounted, static, configurable (e.g., variable densities)

Parameters: soil conditions, plant conditions, weather conditions, etc.

Information Contexts: specific farm, local context for farm, regional contexts (includes interdependencies among micro-zones, eco-zones, growing zones, weather patterns, etc.)

CATEGORIES OF VIRTUAL FARM INPUT RESOURCES

The envisioned virtual farm utilizes a variety of data and information resources. These can include whole data inputs (whole scene or granule), subset data, preprocessed data, modeling results, subsets of larger modeling results, visualization products (such as animations), or preprocessed data to support interactive visualizations and merging with other data types. Correspondingly, quantification of data requirements will be variable due to differences in farm size, frequency of data sampling, and kind of visualization results. The following categories of estimated data help quantify potential ranges of data volumes to be managed by a virtual farm:

Data Sources - Satellite, airborne, in situ:

Satellite Remote Sensing (Sample Sources to Estimate Data Volumes):

Whole data:

SPOT 60x60 km scene = 3-4 GB @ 1m multi-spectral 16 days (use as model for farm-specific application)

Subset data: 1000 acre farm from a 60x60km SPOT scene = .001 x 4 GB = .004 GB

SPOT4 60x60 km scene = 36MB @ 10m panchromatic 16 days (use for farm and locality applications)

Subset data: .1% x 36 MB = .0036 MB

ASTER 60x60 km scene = 118 MB @ 15m 16 days (use for large farm and local/regional information input)

MODIS 10x10 degree scene = 139-680 MB @ 250m daily (use for regional context perspective)

Airborne Remote Sensing:

Example taken from AISA (Airborne Imaging Spectrometer for Applications)

1000 ft altitude = 1m pixel, 28 channels, swath 286 pixels wide. Variable-size data products owing to flight path and duration

~2.8 GB coverage/scan

4000 ft altitude = 4m pixel, 28 channels, variable size data products

~.8 GB for coverage/scan

In Situ Sources:

Mobile mounted - planters, harvesters, applicators, etc. = ~1 MB/hour operation

Static / configurable (densities variable) - data loggers = 4 B/hour x (50/acre * 1000) = 200 KB/hour

Micro-electro-mechanical systems (MEMS) = 10 KB/reading x (50000-1000000) = 500 KB –10 MB/aggregate reading (MEMS could also be deployable as microclimate sensors)

Modeling Results, Visualization, and Simulation Sources:

Weather Information Sources:

Microclimate zones (farm-situated) - monitored with the *in situ* sensors described above. These would provide the highest resolution weather data.

Local weather information - extracted from weather/climate reporting services (i.e., local news services). These sources would involve lower resolutions appropriate for monitoring changing conditions affecting the locale for the farm (e.g., hourly temperature, precipitation, cloudiness, forecasts)

Regional weather and climate histories and forecasts - extracted from state, federal and institutional modeling services. These sources provide the big picture at lower resolutions but yield longer-range forecasts and also histories of trends and patterns affecting the farm's region.

Forecast data feeds - from weather service providers. These could be a subset or could be scaled for farm concerns from master output models and visualizations. Data files for these might be 100MB to 1GB for high-resolution 3-D interactive visualizations. Visualization files that are static viewable animations would be smaller (5 –30 MB MPEG for example).

An active zoom-able visualization "window" on weather would involve a fusion of *in situ* sensor data and local/regional weather models/visualizations. Downloaded modeling data or preprocessed visualization files would be transformed and scaled within the parameters of the virtual farm. These data would have value as temporary and persistent stored resources. For example, a three-day moving "window" on weather updated hourly from external sources can be maintained while selected aspects of this view can be stored into a history file. Near-term storage of weather visualization files for a virtual farm could involve 100-500 MB for a three-day window (past, present, future forecast), depending on whether they are 2-D or 3-D. Compression techniques could reduce the daily files for history files but accumulated visualization data require large storage capacities.

Other:

In addition to weather data, the virtual farm will need to manage the data files produced as a result of selected modeling, visualization, and simulations. These secondary data products (based on image data, metadata, or statistics) can be output from value-added or scientific modeling algorithms designed to support precision farming scaled to the dimensions of the farm. Data files like these constitute externally preprocessed data or inputs that are used in conjunction with virtual farm software functions. Examples include custom models of local drainage patterns, ecosystems and habitats, soils, ground water and aquifer, seed performance histories, agriculture techniques, and economic models.

Speculative estimates for these kinds of data management requirements are made with regard to interactive versus static categorization (see Table 1). The largest files will be for interactive preprocessed visualization data representing multi-dimensional temporal, spatial, and parametric information. Persistence and high-resolution requirements will swell the data volume over time starting in the gigabyte range. Stored preprocessed data files such as these support powerful interactive information visualizations, simulations for "what-if" investigations, and modeling/analysis programs to study histories and make projections regarding farm operations.

CATEGORY	DAILY	MONTHLY	ANNUAL	SUBSET
Satellite *				
High 1m	16 day pass	8GB	96GB	344.496GB x .1 reduction = 34.45GB
Med. 10m	“”	72MB	.864GB	
Med. 15m	“”	236MB	2.832GB	
Low 250m	680MB	20.4GB	244.8GB	
Airborne				
High 1m	2.8GB	*Planting:* *2 x 2/mo.* = 14.4GB *Cultivating;* *4 x 4/mo.* = 57.6GB *Harvesting:* *2 x 2/mo.* = 14.4GB *Winter 4mo:* 0	*Total* 86.4GB	
Med 4m	.8GB			
In Situ				
Mobile	6MB	*Planting:* 84, 48, MB *Cultivating:* 400, 48 MB *Harvesting:* 200, 48MB	*Mobile & static Total* 684MB	
Static	4.8MB .			
Configurable MEMS	240MB	7.2GB	86.4GB	
Visualizations	3 day window	30 day history	cumulative	
*Weather***	100-500MB	5GB	60GB	3-day window + compressed annual = 67GB
Interactive	5-30MB	900MB	10.8GB	
Static			60GB	
*Other ****				
TOTALS			648.8GB	274.9GB

Table 1: Estimated Data Consumption for 1000 Acre Farm

(* Assumes one scene or granule includes the 1000 acre farm)

(** may not need detailed cumulative weather so reduced history stored)

(*** Guesstimates)

Integrated external and internal preprocessed data assets represent extensions of farm-centric information to support specific operations (e.g., land management, profit maps, nutrient profiles, soil maps, chemical application maps, plantings, crop growth health, performance, crop yields, disease profiles, pest damage, harvest profiles, regulation compliance). Preprocessed data generated externally and internally (within the virtual farm) represents a significant data management requirement, one that an IA can address. For instance, cumulative, persistent storage of these data could represent hundreds of terabytes that a virtual farm could manage with virtual storage services offered by an IA.

Extrapolating estimated annual data consumption for single 1000 acre farms to the 600,000 acres of the Central Valley agriculture zone in California yields 275 GB x 600 = 165,000 GB [4]. This represents a potential distribution of 165 TB managed data for just one agricultural region. Persistence requirements may decrease at the virtual farm provided the intelligent archive provides faithful historical data that can be requested on-demand. Hence requirements for data persistence will be high for archiving services and rolling deletions strategies may be employable at the consumer end (e.g. virtual farm) though the option to retain history data is always available to it. Further extrapolation of this model to the entire United States (and to international applications) suggests intelligent data management, data understanding, distribution strategies, and information transformation services for agriculture alone will be in high demand.

CONCLUSION

The precision agriculture use case provides insights into requirements and capabilities that an IA must possess to support a digital farm. The value of a virtual farm is largely in how it acts as an intelligent interface between the information needs of a farm and the data and computational resources available to serve agricultural enterprise stakeholders. Its ability to request, receive, filter, and extract information from data and manage these resources on behalf of a farmer depends on a cyber infrastructure with intelligent archive services. The digital farm scenario illustrates a sophisticated application operating in conjunction with an IA for acquiring and using remote sensing data from a variety of sources (e.g. archives, in situ sensors, models, value-added providers). Our approach of studying requirements to support example use cases has contributed deeper insights into how an IA can address the goal of getting the most value from all available data for various applications. By studying additional use cases [5], [6] we expect to exercise concepts for an IA and bring into better focus a notional architecture with which to relate intelligent systems technology for effective integration into overall services of knowledge building and knowledge exploiting enterprises.

REFERENCES

[1.] Ramapriyan, H.K., G. McConaughy, C. Lynnes, S. Kempler, K. McDonald, R. Harberts, L. Roelofs, P. Baker, 2002. “Conceptual Study of Intelligent Archives of the Future”, Report prepared for the Intelligent Data Understanding program, http://daac.gsfc.nasa.govIDA/IA_report_8-27-02_baseline.pdf.

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

[3.] Quatrani, Terry, Grady Booch, “Visual Modeling with Rational Rose and UML” 21-39, Addison-Wesley, 1998

[4.] Brozovic, Nicholas, J. Carey, D Sunding, “How Big are Farms in the Westlands?”, July 2001, http://are.berkeley.edu/~brozovic/econ/westland/westsize.pdf

[5.] Harberts, R., et al, “Intelligent Archive visionary Use Case: Advance Weather Prediction”, September, 2003, http://daac.gsfc.nasa.gov/IDA/

[6.] Harberts, R., et al, “Intelligent Archive Visionary Use Case: Virtual Observatories”, 2003, http://daac.gsfc.nasa.gov/IDA/