# A Strategic Approach to Engineering Information Portals

*The business-driven, information-centric, architecture-enabled Information Engineering methodology described in this paper provides an effective, productive, common sense approach to developing strategic data (Information Portals).*

## Table of Contents

## What Is an Information Portal?

> *An "Information Portal" is a collection of computer-based information that is critical to successful execution of enterprise initiatives.*

An Information Portal is more than an archive for corporate data and more than a new way of accessing corporate data. An Information Portal is a subject-oriented repository designed with enterprise-wide access in mind. It provides tools to satisfy the information needs of enterprise managers at all organizational levels — not just for complex data queries, but as a general facility for getting quick, accurate, and often insightful information. An Information Portal is designed so that its users can recognize the information they want and access that information using simple tools.

One of the principal reasons for developing an Information Portal is to integrate operational data from various sources into a single and consistent architecture that supports analysis and decision-making within the enterprise. Operational (legacy) systems create, update and delete production data that "feed" the Information Portal.

an Information Portal is analogous to a physical warehouse. Operational systems create data "parts" that are loaded into the warehouse. Some of those parts are summarized into information "components" and stored in the warehouse. Information Portal users make requests and are delivered information "products" that are created from the components and parts stored in the warehouse.

an Information Portal is typically a blending of technologies, including relational and multidimensional databases, client/server architecture, extraction/transformation programs, graphical user interfaces, and more.

Data warehousing is one of the hottest industry trends — for good reason. A well-defined and properly implemented Information Portal can be a valuable competitive tool.
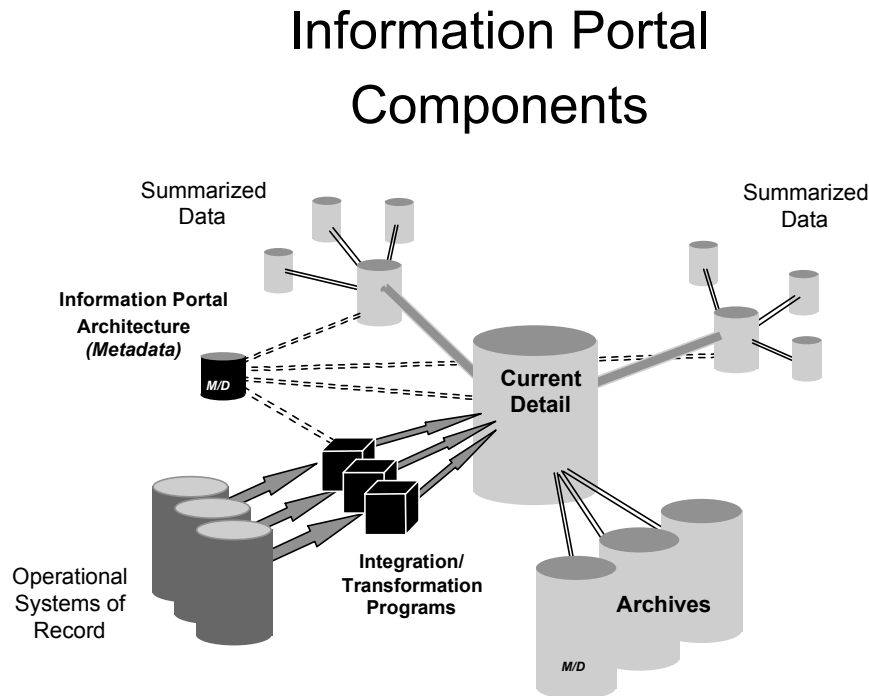
## Information Portal Benefits

Implementing an Information Portal provides significant benefits -- many tangible, some intangible.

- **More cost-effective decision making**. an Information Portal allows reduction of staff and computer resources required to support queries and reports against operational and production databases. This typically offers significant savings. Having an Information Portal also eliminates the resource drain on production systems when executing long-running, complex queries and reports.

- **Better enterprise intelligence**. Increased quality and flexibility of enterprise analysis arises from the multi-tiered data structures of an Information Portal that support data ranging from detailed transactional level to high-level summary information. Guaranteed data accuracy and reliability result from ensuring that an Information Portal contains only "trusted" data.

- **Enhanced customer service**. An enterprise can maintain better customer relationships by correlating all customer data via a single Information Portal architecture.

- **Business reengineering**. Allowing unlimited analysis of enterprise information often provides insights into enterprise processes that may yield breakthrough ideas for reengineering those processes. Just defining the requirements for an Information Portal results in better enterprise goals and measures. Knowing what information is important to an enterprise will provide direction and priority for reengineering efforts.

- **Information system reengineering**. an Information Portal that is based upon enterprise-wide data requirements provides a cost-effective means of establishing both data standardization and operational system interoperability. Information Portal development can be an effective first step in reengineering the enterprise's legacy systems.

## *Information Portal Components*

The following primer describes each of the components of an Information Portal (see figure). This description is based upon the work of W. H. Inmon, credited as the father of the Information Portal c

# Information Portal
# Components

Summarized Data

Information Portal Architecture *(Metadata)*

M/D

Summarized Data

**Current Detail**

Operational Systems of Record

**Integration/ Transformation Programs**

**Archives**

*M/D*

## Summarized Data

*Lightly summarized data* are the hallmark of an Information Portal. All enterprise elements (department, region, function, etc.) do not have the same information requirements, so effective Information Portal design provides for customized, lightly summarized data for every enterprise element (see Data Mart, below). An enterprise element may have access to both detailed and summarized data, but there will be much less than the total stored in current detail.

*Highly summarized data* are primarily for enterprise executives. Highly summarized data can come from either the lightly summarized data used by enterprise elements or from current detail. Data volume at this level is much less than other levels and represents an eclectic collection supporting a wide variety of needs and interests. In addition to access to highly summarized data, executives also have the capability of accessing increasing levels of detail through a "drill down" process.

## Current Detail

The heart of an Information Portal is its *current detail*, where the bulk of data resides. Current detail comes directly from operational systems and may be stored as raw data or as aggregations of raw data. Current detail, organized by subject area, represents the entire enterprise, rather than a given application.

Current detail is the lowest level of data granularity in the Information Portal. Every data entity in current detail is a snapshot, at a moment in time, representing the instance when the data are

accurate. Current detail is typically two to five years old. Current detail refreshment occurs as frequently as necessary to support enterprise requirements.

## System of Record

A *system of record* is the source of the data that feed the Information Portal. Data in an Information Portal differ from operational systems data in that they can only be read, not modified. Thus, it is necessary that an Information Portal be populated with the highest quality data available, i.e., data that are most timely, complete, accurate, and have the best structural conformance to the Information Portal. Often these data are closest to the source of entry into the production environment. In other cases, a system of record may be one containing already summarized data.

## Integration and Transformation Programs

Even the highest quality operational data cannot usually be copied, as is, into an Information Portal. Raw operational data are virtually unintelligible to most information users. Additionally, operational data seldom conform to the logical, subject-oriented structure of an Information Portal. Further, different operational systems represent data differently, use different codes for the same thing, squeeze multiple pieces of information into one field, and more. Operational systems in most enterprises have been developed independently (no common enterprise data architecture), which means that most operational data are stored and managed redundantly and may even reside in many different physical sources: old mainframe files, non-relational databases, indexed flat files, even proprietary tape and card-based systems. These operational data must be cleaned up, edited, and reformatted before being loaded into an Information Portal.

As operational data items pass from their systems of record to an Information Portal, the following scripts need to be developed:

- Reformatting, recalculating, or modifying key structures

- Adding time elements

- Identifying default values

- Supplying logic to choose between multiple data sources

- Summarizing, tallying, and merging data from multiple sources

When either operational or Information Portal environments change, these scripts will need to be modified to reflect that change.

## Archives

Information Portal archives contain old data (normally over two years old) of significant, continuing interest and value to the enterprise. There is usually a massive amount of data stored in the Information Portal archives, with a low incidence of access. Archive data are most often used for forecasting and trend analysis. Although archive data may be stored with the same level of granularity as current detail, it is more likely that archive data are aggregated as they are archived. Archives include not only old data (in raw or summarized form); they also include the metadata that describes the old data's characteristics.

## Metadata

One of the most important parts of an Information Portal is its metadata — or data about data. Also called Information Portal architecture, metadata is integral to all levels of the Information Portal, but exists and functions in a different dimension from other warehouse data. Metadata that is used by Information Portal developers to manage and control Information Portal creation and maintenance resides outside the Information Portal. Metadata for Information Portal users is part of the Data

Warehouse itself and controls access and analysis of the Information Portal contents.  To an Information Portal user, metadata is like a "card catalog" to the subjects available.

### *Information Portal Structure*

an Information Portal may have any of several structures:

# Information Portal
# Structure
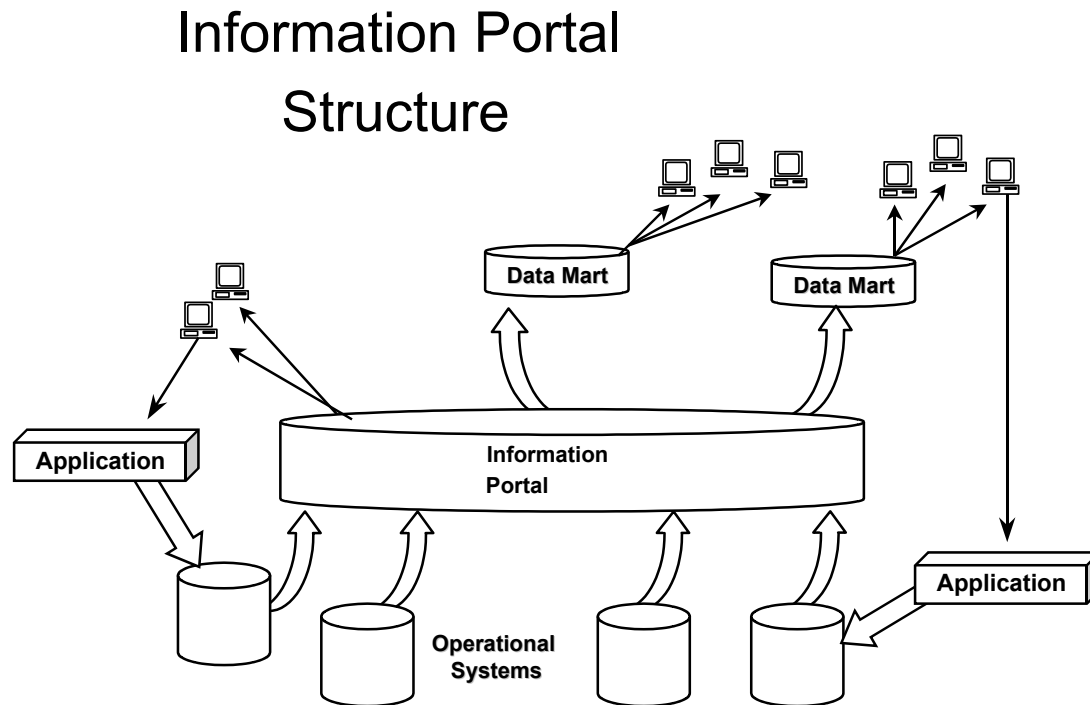


- **Physical Information Portal** - physical database in which all the data for the Information Portal are stored, along with metadata and processing logic for scrubbing, organizing, packaging and  processing the detail data.

- **Logical Information Portal** - also contains metadata, including enterprise rules and processing logic for scrubbing, organizing, packaging and processing the data, but does not contain actual data. Instead, it contains the information necessary to access the data wherever they reside. This structure is effective only when there is a single source for the data and they are known to be  accurate and timely.

- **Data Mart** - subset of an enterprise-wide Information Portal, which typically supports an enterprise element (department, region, function, etc.).  As part of an iterative Information Portal development process, an enterprise builds a series of physical (or logical) data marts over time and links them via an enterprise-wide logical Information Portal or feeds them from a single physical warehouse.

### *Information Portal Engineering*

There are three popular "approaches" for information warehousing.  Unfortunately, two of them are quick-fix solutions that ultimately waste resources and do not fully meet enterprise information needs.

*Data Dump*  -- all enterprise data are replicated or made available with no attempt to "scrub" or even categorize the data.  This is like dumping all the contents of a physical warehouse in the middle of the floor – new stuff, old stuff, and broken stuff -- and asking your customers to pick out what they need from the pile.

*Magic Window* -- to the data wherever they exist in the enterprise, again without ensuring data quality.  This is like a big sack in which there are rubies and emeralds and gold nuggets and broken glass and rat droppings and poisonous snakes.  Sometimes you can "mine" a gem, but at some point you will quit putting your hand in the bag.

*Strategic Information Portal* -- results in enterprise information based upon business requirements and common data architecture!

"Information Portal" and "Information Portal Architecture" are relatively new terms that describe the methods and concepts that have used for almost twenty years to develop and implement Executive Information Systems (EIS), Decision Support Systems (DSS), and Management Information Systems (MIS).  These years of practical experience have been distilled as best practices that make up the approach to Information Portal development described in this paper (illustrated in the figure below).

The "*Critical Success Factors for Information Portal Engineering*" are described in another of the author's white papers.

One of the keys to success in scaleable Information Portal development is using an iterative approach that includes active participation of potential Information Portal users.

Like any other large information systems project, Information Portal development can get bogged down if the scope is too broad, and the number of people involved is too large.  A clear purpose and scope is necessary to manage the application of information systems resources, as well as the expectations of potential Information Portal users.  The simplest way to limit scope is to build an enterprise Information Portal one data mart at a time.  Each data mart supports a single organizational element, enterprise function or business object (e.g., customer, product, account, etc.), and the scope of development  is  limited by the data mart requirements.
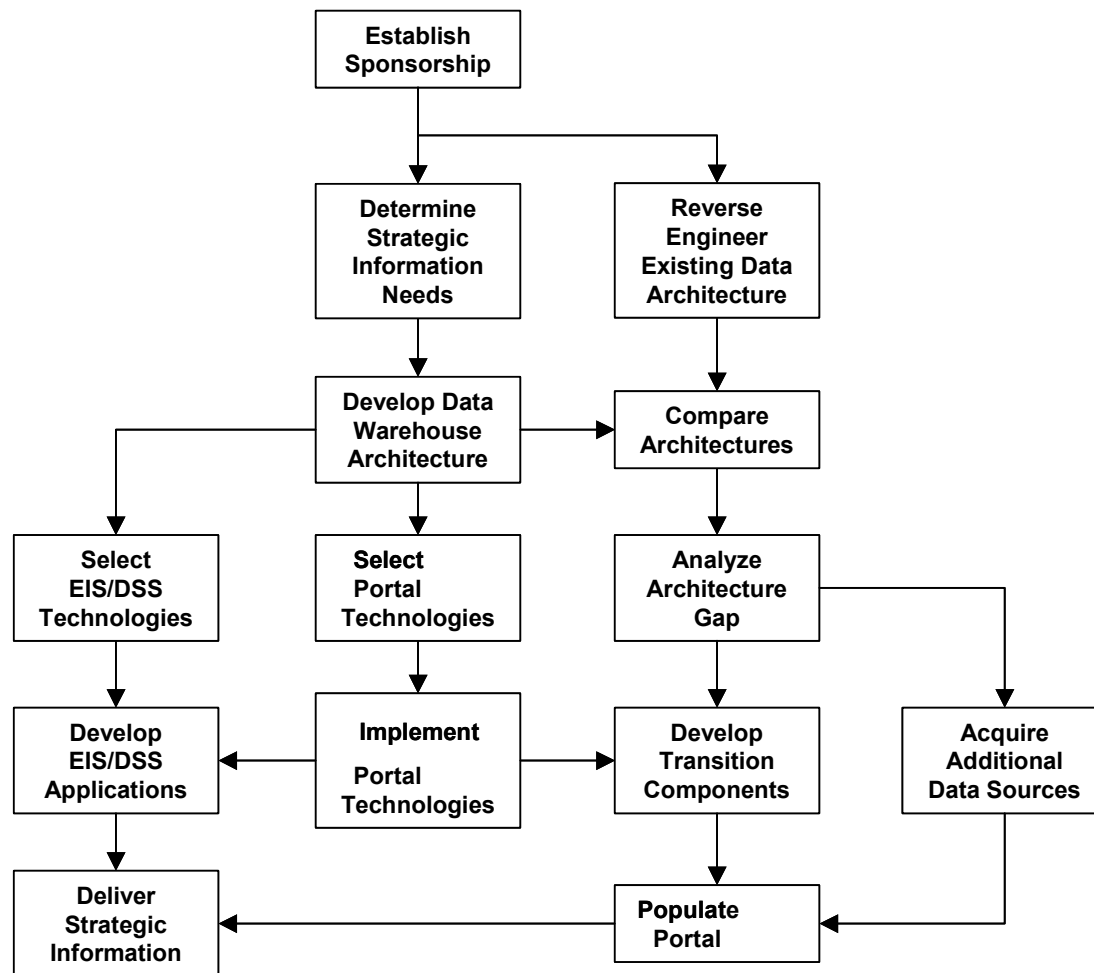
For the initial data mart, which usually provides the Information Portal proof-of-concept, the scope must be sufficient to provide real, immediate, and high profile benefits.  After the first data mart is developed and implemented, additional data marts can be developed and integrated over time as enterprise needs  dictate and as resources are available.

Strategic Data Warehouse Engineering involves five key activities:


(1)  Establish sponsorship;
(2)  Identify enterprise needs;
(3)  Design Information Portal
      architecture;
(4)  Apply appropriate technology; and
(5)  Implement the Information Portal.

## Establish Sponsorship

The first step is to establish sponsorship for the Information Portal, if it does not already exist. Establishing the right sponsorship chain will help ensure successful development and implementation.  The sponsorship chain should include a data warehousing manager and two other key individuals.  At the top of the chain is an executive sponsor with resources to invest in information infrastructure improvement.  A project "driver" between the executive sponsor and the warehousing manager keeps the project moving and on schedule.

```
                    ┌──────────────┐
                    │  Establish   │
                    │ Sponsorship  │
                    └──────┬───────┘
              ┌────────────┴────────────┐
              ▼                         ▼
      ┌──────────────┐         ┌──────────────┐
      │  Determine   │         │   Reverse    │
      │  Strategic   │         │   Engineer   │
      │ Information  │         │ Existing Data│
      │    Needs     │         │ Architecture │
      └──────┬───────┘         └──────┬───────┘
             ▼                        ▼
      ┌──────────────┐         ┌──────────────┐
      │ Develop Data │────────▶│   Compare    │
      │  Warehouse   │         │ Architectures│
      │ Architecture │         └──────┬───────┘
      └──┬────────┬──┘                │
         │        ▼                   ▼
         │ ┌──────────────┐    ┌──────────────┐
         │ │    Select    │    │   Analyze    │
  ┌──────┘ │    Portal    │    │ Architecture │────────┐
  ▼        │ Technologies │    │     Gap      │        │
┌──────────────┐ └──────┬──┘   └──────┬───────┘        │
│    Select    │        ▼             ▼                ▼
│   EIS/DSS    │ ┌──────────────┐ ┌──────────────┐ ┌──────────────┐
│ Technologies │ │  Implement   │ │   Develop    │ │   Acquire    │
└──────┬───────┘ │              │ │  Transition  │ │  Additional  │
       ▼         │    Portal    │ │  Components  │ │ Data Sources │
┌──────────────┐ │ Technologies │ └──────┬───────┘ └──────┬───────┘
│   Develop    │◀│              │           │              │
│   EIS/DSS    │◀└──────────────┘           ▼              │
│ Applications │                    ┌──────────────┐       │
└──────┬───────┘                    │   Populate   │◀──────┘
       ▼                            │    Portal    │
┌──────────────┐                    └──────┬───────┘
│   Deliver    │◀──────────────────────────┘
│  Strategic   │
│ Information  │
└──────────────┘
```

An important aspect of establishing sponsorship is ensuring everyone in the enterprise understands the purpose of the Information Portal, its potential benefits, and the enterprise's plan for implementation.  an Information Portal engineering plan should be developed early in the Information Portal engineering cycle and should address all of the necessary activities.

## Identify Enterprise Needs

Identifying enterprise needs is a major component in the engineering life cycle for any information system, and it is crucial when engineering an Information Portal.  When developing operational  systems, there is often one single enterprise sponsor or one group of users with a clear view of what they need, what the system should look like, and how it should function.  When developing an Information Portal, there are always multiple potential users, each with a different idea of what an Information Portal is and what it should provide.  Because of this lack of a single focused direction, identifying precise enterprise needs is critical to the success of an Information Portal project.

Enterprise Information Portal needs should be expressed in terms of enterprise measures and critical success factors.  An enterprise's business plans typically provide the basis for defining enterprise strategic information needs.  Interviews with key enterprise managers and analysis other pertinent documentation are also techniques that should be used to determine these needs.

## Determine Measurement Cycles

Completely defining an enterprise measure includes describing the cycles or time periods used for the measure.  Are quarters, months, or hours appropriate for capturing useful measurement data?  How much historical data will be needed?  These vary greatly by enterprise.  The United States Federal Reserve Bank views enterprise measures in monthly, quarterly and annual increments and uses years of historical data to determine trends in the economy.  An insurance company requires decades of actuarial data for meaningful measures.  A telephone sales operation, on the other hand, uses hourly enterprise measures and may only keep a few weeks of information.

## Validate Measures

After identifying and defining enterprise needs, it is absolutely necessary to communicate them throughout the enterprise.  One of the best justifications for undertaking an Information Portal project is the synergy achieved through the process of defining and then communicating its critical success factors and measures.  Everyone becomes aware of precisely what defines success and how it is measured.  In addition, the measures undergo a "reality check" by people who were not involved in their development, but who may be measured by them and who will be involved in creating the raw data from which the measures will be derived.  Their feedback is used for refining the measures.

## Resolve Data Conflicts

As enterprise portals are developed, businesses frequently discover that very similar data entities and processes exist in various departments under different names.   If not corrected, homonyms, synonyms, and other data conflicts may be integrated into a corporate Information Portal as duplicative, inefficient data elements or processes that generate unnecessary and erroneous complexity.  One of the most commonly misused terms is "customer."

> To the Accounting Department, "customer" could mean the organization (or individual) that receives a bill.  "Customer" could also mean an individual receiving service or buying a product.  To the Sales Department, "customer" could mean the organizations on which the salesperson calls.  Providing any one of these interpretations as the enterprise definition of "customer" would not meet the needs of the enterprise and would doom its Information Portal effort to failure. Additionally, each department could use different names to describe the same data  entity  (Customer vs. Client vs. Prospect vs. Account…).

It is advisable to take great pains to resolve all data conflicts in the Information Portal strategic  information model before continuing with the next phase of the development cycle.

## Build an Enterprise Model

Before defining and designing the Information Portal architecture, it is first necessary to model the overall  enterprise  architecture.    First,  enterprise  measures  and  critical  success  factors should  be  documented.    Supporting  Information  Portal  data  entities  should  then  be documented  in  a  corresponding  data  model.  Information Portal data entities are those that, at any point in time, tell Information Portal users how well their enterprise is performing.  Providing a clear and unambiguous definition of every warehouse data entity, describing the way each is used,  as  well  as  defining  derivation  formulas,  aggregation  categories  and  time  periods,  are activities critical to capturing a clear understanding of an enterprise's measures.  The resulting enterprise  architecture  model,  which  links  enterprise   needs   with   Information   Portal   data entities  and  enterprise  rules,  becomes  both  requirements documentation and a source for communicating the contents of the Information Portal (its metadata) to its users.

## Design Information Portal Architecture

Only  after  defining  and  thoroughly  documenting  enterprise  needs  (measures  and  critical success factors), should an organization begin actual Information Portal architecture (metadata) design  (for

more information, see the author's white paper. "*Information Portal Architecture -- Blueprint for Success*"). This activity also involves active user participation in facilitated design sessions. There are two types of Information Portal metadata: structural and access.

**Structural metadata** is used for creation and maintenance of the Information Portal. It fully describes Information Portal structure and content. The basic building block of structural metadata is a model that describes its data entities, their characteristics, and how they are related to one another. The way potential Information Portal users currently use, or intend to use, enterprise measures provides insight into how to best serve them from the Information Portal, i.e., what data entities to include and how to aggregate detailed data entities. an Information Portal data model provides a means of documenting and identifying both strategic and operational uses of enterprise measures. It also provides the capability to document multi-dimensional summarization of detail data.

> *Strategic thinkers* tend to look for "big picture" answers, and therefore need very few aggregation categories. The "roll-ups" for each strategic aggregation of data, however, can be quite complex.

> *Operational thinkers* have a tendency to want to dissect and review every measure by every category used in their part of the enterprise, and thus tend to require large numbers of less complex aggregation categories.

Structural metadata identifies the system of record for all Information Portal data entities. It also fully describes the integration and transformation logic for moving each Information Portal entity from its system of record to the Information Portal. In addition, structural metadata defines the refreshment schedule and archive requirements for every data entity.

When the Information Portal structure changes, its metadata is changed accordingly. Old versions of the structural metadata are kept to document the changing nature of the Information Portal and allow access to archive data. This will ensure that decision support activities, such as trend analysis and data mining, continue to provide quality results.

Structural metadata also includes performance metrics for programs and queries so that users and developers know how long programs and queries should run. Information Portal performance tuning also uses these metrics.

**Access metadata** is the dynamic link between the Information Portal and end-user applications. It generally contains the enterprise measures supported by the Information Portal and a dictionary of standard terms including user-defined custom names and aliases. Access metadata also includes the location and description of Information Portal servers, databases, tables, detailed data, and summaries along with descriptions of original data sources and transformations.

Access metadata provides rules for drill up, drill down and views across enterprise dimensions and subject hierarchies like products, markets, and customers. Access metadata also allows rules for user-defined custom calculations and queries. In addition, access metadata contains individual, work group, and enterprise security for viewing, changing, and distributing custom calculations, summaries, or other analyses.

A critical part of designing the Information Portal architecture is reverse engineering the existing operational systems of record as physical design models in the architecture repository. Every data element in the physical design models is associated with its logical counterpart in the logical strategic information model. This provides the basis for gap analysis and defining transformation requirements. The next step involves identifying systems of record.

> Clearly defining enterprise Information Portal architecture also involves identifying the correct source of raw operational data to populate the Information Portal. This effort also addresses possible integration and transformation logic. Identifying the systems of record for Information Portal data entities is one means of validating enterprise measures.

## Apply the Correct Technology Solution

Only after fully defining enterprise requirements and designing the Information Portal architecture should an enterprise begin to select the technology for the Information Portal. Key technology issues, in addition to determining the hardware/software platform for the Information Portal, include developing programs for loading information into the Information Portal, implementing access control (security) mechanisms and designing the UI.

### *Determine Hardware/Software Platforms*

The following are some important considerations for determining a **hardware platform**:

How much data will be in the Information Portal and how much can the platform accommodate economically? How scaleable is the platform? Is it optimized for Information Portal performance? Will the platform support the software selected for the Information Portal?

Concurrent with hardware selection is the selection of **system software** to support the Information Portal. Among the choices are operating systems, development software, and database management systems. The structure and size of the Information Portal will determine system software requirements. For example, an Information Portal that includes data marts will require not only relational technology, but also multidimensional access and a client/server architecture.

### *Develop Integration and Transformation Programs*

Integration and transformation programs are necessary to extract information from operational systems and databases for both initial load and subsequent updates of the Information Portal. Sometimes, it is possible to develop a single program for both initial load and periodic updates of the Information Portal, but often circumstances make this an unacceptable development option.

- A separate initial load program is necessary when the volume of initial data is so large that it cannot be transferred without adversely impacting other users of the operational systems. This is particularly true when initial load and update volumes are significantly different.

- Separate programs also should be considered for capturing historical data from the operational systems for loading into the Information Portal, because this is usually a one-time process.

- An additional reason for separate initial warehouse loading programs involves historical data maintained separately from the operational systems (many operational systems only maintain the most recent values for data). This situation usually requires retrieval of historical data from archive and backup files.

Usually one set of integration and transformation programs initially loads the Information Portal, and a second set periodically updates the Information Portal. Update programs are generally smaller and simpler than programs developed to load the Information Portal. Update programs often are built into operational systems to trap new occurrences of data as they are added. This works best for well-documented, in-house operational systems. Update programs that extract data from commercial off-the-shelf software or from older, poorly documented, legacy systems typically capture and transform just the changes made since the last update. Over time, update programs will be changed to reflect changes in both operational systems and other data sources.

### *Develop Security Policies and Procedures*

an Information Portal is a read-only source of enterprise information, therefore developers need not be concerned unduly with controlling create, update and delete capabilities. However, developers will need to address the trade off between protecting a valuable corporate asset against unauthorized access and making the data accessible to anyone within the enterprise who can put it to good use. The best solution is to allow everyone in the enterprise to have access to

the enterprise measure definitions and derivations, but only allow access to the underlying detailed data only on an approved, need-to-know basis.

In addition to access security, an enterprise must be concerned with physical security for its Information Portal.  Because its contents are an extremely valuable corporate resource, they must be protected against loss and damage.  This protection is available in many forms ranging from simple backup and off-site storage strategies to installation of no-break power and redundant disk storage and computer systems.

### *Design User Interfaces*

Information Portal users get useful information from the Information Portal through user interfaces. It is the quality of the user interfaces that has the most impact on how users will perceive the Information Portal. The following criteria should be used to design and/or select  an Information Portal user interface.

- Two criteria for selecting an effective user interface are ease of use and performance.  For ease of use, most enterprises turn to graphical user interfaces.  For performance, developers must ensure that the hardware/software platform fully supports and is optimized for every chosen user interface.

- Another important selection criteria for user interfaces are the information needs and the level of computer literacy of potential users.  A general rule is that users of highly summarized data need simple, extremely graphical interfaces, and detail data users need more complex, but less graphical tools.

- One more critical user interface criterion is that the interface supports the access metadata designed for the Information Portal.

If a user interface is easy to use, allows all potential users to get the information they need in the format they need, and does it in an acceptable amount of time, it is the right interface.  It is generally easier to select the correct user interface if the users are involved in every step of the development/deployment life cycle.

## Implement the Data Warehouse

Data Warehouse implementation includes loading the preliminary data, implementing transformation programs, designing a user interface "look and feel," developing standard queries and reports, and thoroughly training Information Portal users.

Implementation begins with integration and transformation programs for ensuring data quality.  Effective loading and maintaining of data in the Information Portal and associated data marts is critical for decision support activities and also providing information to users and stakeholders.  Also, in order to allow users to properly compare data based on different periods of time, it is important that the Information Portal databases allow for the building of history, and not just provide snapshots that are valid only at one point in time.

### *Summary*

Information is a valuable resource.  A well-defined Information Portal, properly implemented, can be a valuable tool for managing and using that resource.  It translates the vast volumes of detailed,  unorganized data an enterprise captures via its operational systems into useful feedback, predictors, and warnings that help Information Portal users at every organizational level make informed decisions.

*Using the strategic approach for Information Portal Engineering described in this paper will help you deliver effective strategic information that exactly meets the needs of your enterprise -- public or private, large or small -- to the right people, in the right place, at the right time, in the right format.*

For more information about this topic please contact:


**Visible Systems Corporation**
**711 Atlantic Avenue**

**Boston, MA**
**contact@visiblesystemscorp.com**

Alan Perkins has been a Systems Analyst on the White House staff, Director of the US Army Data Processing School in Germany, Vice President of R&D for a virtual corporation, Vice President of Consulting for Visible Systems Corporation and General Manager of a high-tech consulting firm.  He has provided information and enterprise management consulting to numerous companies, associations and government agencies.

Mr. Perkins specializes in Enterprise Architecture Engineering.  He helps clients quickly engineer enterprise architectures that are actionable and adaptable.  His approach results in architectures that enable and facilitate enterprise initiatives such as Corporate Portals, Enterprise Data Warehouses, Enterprise Application Integration, Software Component Engineering, etc.