# The New Alphabet Soup: Models of Data Integration, Part 2

*Noam H. Arzt, PhD*

In my last column, I focused on the first three models of data integration: smart card (Model 1), peer-to-peer (Model 2, with several variations), and information broker (Model 3, sometimes referred to as a federated model). This column will focus on the two remaining models, which represent points along a continuum from least centralized to most centralized implementation.

### Continuing Along the Continuum

While the information broker model only contains patient demographics in its central hub, Model 4, the partitioned warehouse, introduces a central database operated by the RHIO that assembles complete, consolidated records of people and their medical data (see Figure 1). The actual data is contained in segmented data "vaults" that isolate the medical data supplied by each participating institution and function as surrogates for the local data systems of the participating organizations.

Similar to Model 3, the central database contains a master patient index that tags each patient's record

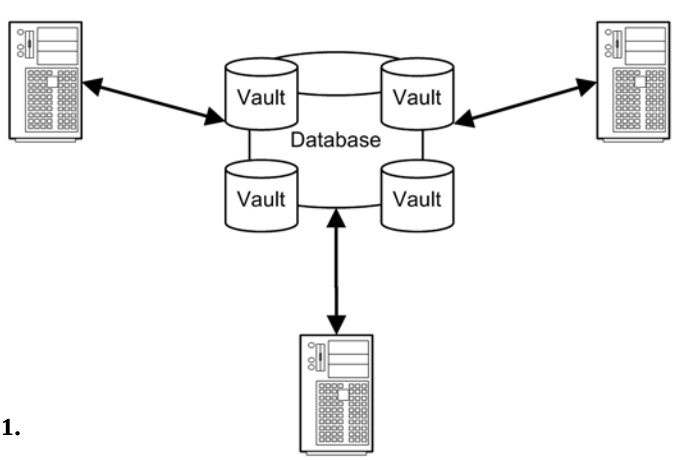


**Figure 1.**

JOURNAL OF HEALTHCARE INFORMATION MANAGEMENT

with the participating systems that contain data about that patient. But unlike Model 3, the central index need go no further than its local data "vaults" to fulfill a valid query about a patient. All the medical data is available and assembled on the fly based on the needs of a particular query.

Participating systems are required to periodically update data in the central database cluster. As before, standards for communication exist for data formats and message types (such as HL7), vocabulary (such as SNOMED), and communications techniques (such as Web Services and SOAP). This model can support real-time or batch communications.

This model has less real-time dependence on other participating systems because all the data to satisfy a query about a patient is located centrally. It implements a strict "need-to-know" policy for data access because clinical data is held in segmented "vaults" and only released as required.

The model facilitates community-wide data analysis because data is easily consolidated centrally. It scales well as long as appropriate investments are made in central resources. Economies of scale are introduced through the use of large-scale central resources. This model probably will provide better expertise in managing central resources because of the scale and class of products used. This model should be able to support existing systems well, because data only needs to be extracted and sent to the central repository.

On the other hand, strong central coordination is required because the central database cluster needs to be carefully managed and maintained for this system to work. There is a dependence on a large central database for inter-system queries.

> *"The final model, the central warehouse, builds on Model 4, but rather than keeping clinical data isolated in vaults, all demographic and clinical data is consolidated into a single central database, or data warehouse."*

Queries still require a central system to consolidate data on the fly and therefore may take longer than desired to fulfill. There is a data timeliness issue—data submission from participating systems to the central database cluster may lag, and as a result queries may yield inaccurately consolidated records.

This model may require a large effort to keep not only demographic records but also clinical records free from duplication, because these records will be collected from numerous disparate sources. This model is harder to implement incrementally, because it requires a larger upfront investment in central resources. It is unclear how to economically implement a large number of isolated vaults for small data providers as the project matures and smaller data providers join. Finally, this is likely a fairly expensive option to implement, not only technically but organizationally.

**Building a Central Warehouse**

The final model, the central warehouse, builds on Model 4, but rather than keeping clinical data isolated in vaults, all demographic and clinical data is consolidated into a single central database, or data warehouse (see Figure 2).

Various methods can be used to bring data together or relate data from multiple sources together in the same database. All the medical data is available immediately to fulfill the needs of a particular query. Participating systems still are required to supply data periodically to the central system. Standards for communication exist for data formats and message types (such as HL7), vocabulary (such as

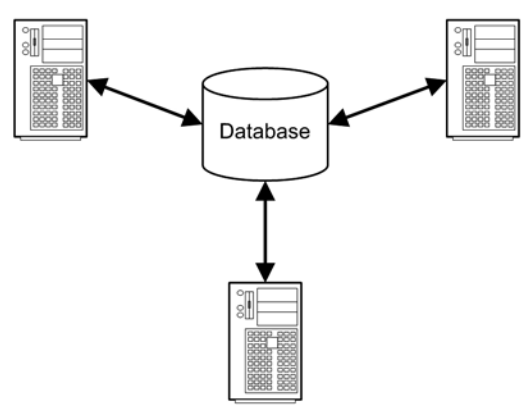


**Figure 2.**

SNOMED) and communications techniques (such as Web Services and SOAP). This model can support either real-time or batch communications.

In this model, the querying system's response to a data request is quicker than other models because all the data is already centrally maintained and consolidated. There is less real-time dependence on other participating systems for the same reason. This model facilitates community-wide data analysis because data is available centrally, and it scales well as long as appropriate investments are made in central resources.

Once again, economies of scale are introduced through the use of large-scale central resources. This model should be able to support existing systems well, because data needs only to be extracted and sent to the central repository.

The limitations of this model are similar to the partitioned warehouse (Model 4), although because all data is already consolidated, queries should execute more quickly than with a partitioned warehouse.

**Consider the Alternatives**

These last two models move us to the far end of the integration continuum with options that are fairly centralized to implement. These more centralized approaches are becoming popular with a new breed of application service provider-based electronic medical records applications that are being positioned as "out-of-the-box" RHIO solutions.

All of these models present viable alternatives for RHIO development and deployment. The specific requirements will dictate a best fit for a particular project. The discussion of the strengths and weaknesses of each model provides some depth to assist in considering each one. Ultimately, a decision comes down to making various tradeoffs between different attributes of the models.

In the next column, I will discuss models for application integration.

**About the Author**

Noam H. Arzt, PhD, is president and founder of HLN Consulting, LLC, San Diego, which consults in healthcare systems integration, especially in public health. He can be reached at arzt@hln.com.