



Architecture for Person Matching and De-duplication

Noam H. Arzt, PhD

American Public Health Association

Annual Meeting

Session 5052.1

October 29, 2008

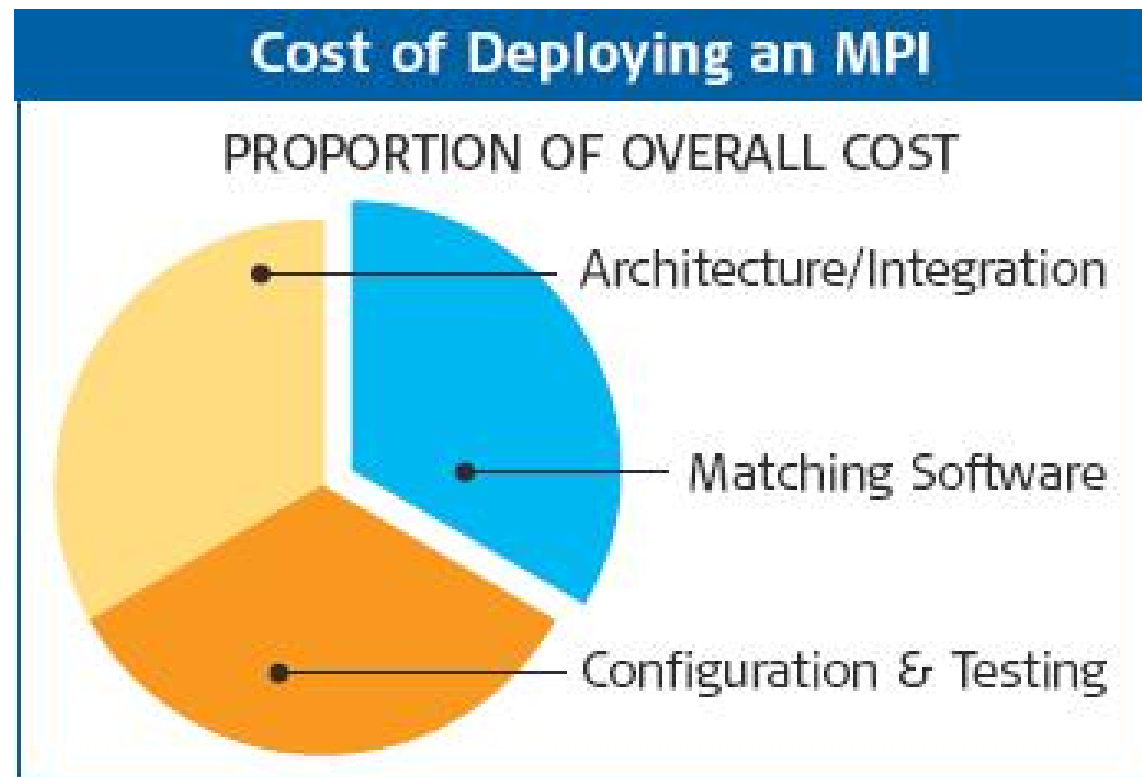


Presenter Disclosures

(1) The following personal financial relationships with commercial interests relevant to this presentation existed during the past 12 months:

No relationships to disclose

De-duplication Strategies





Assumptions

- Functionality under any of these options may differ if different matching/de-duplication products are used.
- Matching/de-duplication products likely will require custom programming to integrate the software into existing systems, or to provide interfaces (batch, queuing, or interactive) to necessary users and processes.

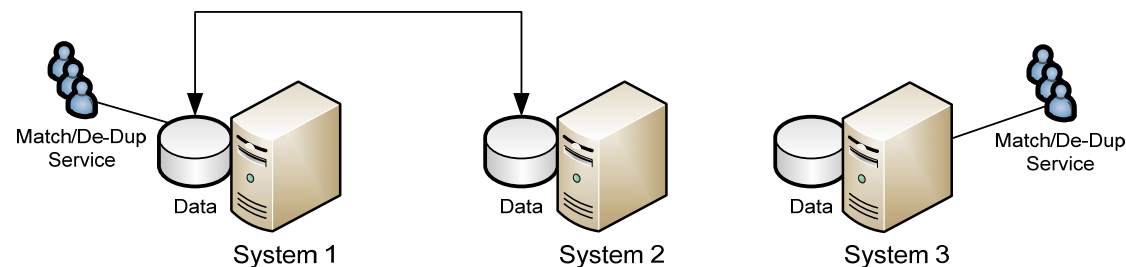


Four Models

1. Centralized
2. Cooperative
3. Distributed
4. Data Warehouse

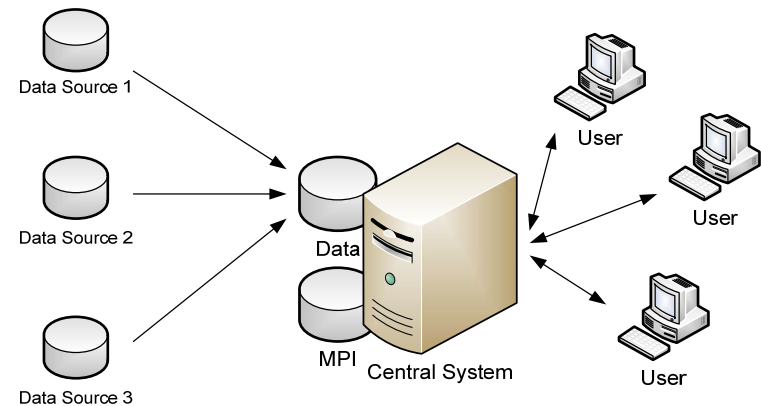
Distributed Model

- *De facto* approach by most agencies
- Limited leadership and commitment
- Individual programs negotiate and implement interoperability



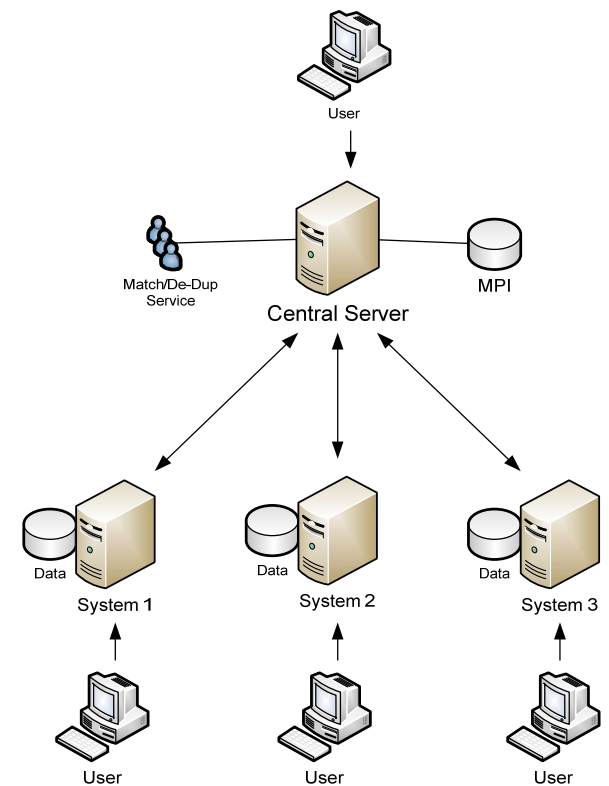
Centralized Model

- Best under strong leadership, clear requirements, solid funding
- Usually centers around a single application or suite (e.g., ERP)
- Organization-wide or a subset
- Can support operations and query



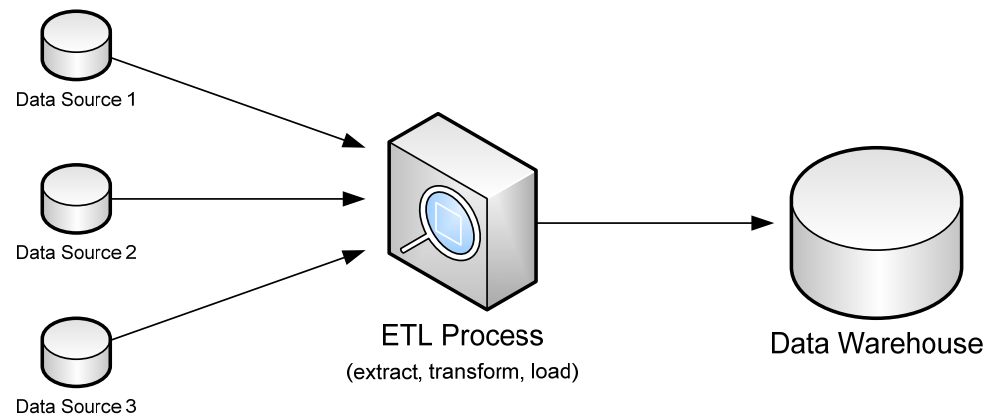
Cooperative Model

- More autonomy to individual departments or programs
- Requires strong standards and a desire to interoperate
- Can be phased in over time
- May require significant modifications to participating systems
- Usually supports operations, but sometimes also query



Warehouse Model

- Primarily supports data analysis rather than operations
- Usually exists in parallel to operational systems
- Can be built incrementally
- COTS or home-grown products to support ETL



Model Comparison: Organizational Factors

Factor	Centralized/ Warehouse	Cooperative	Distributed
Political Sponsorship	Strong	Conceptual support Less commitment	Passive at best
IT Leadership	Strong CIO and mandate	Strong CIO but less mandate over apps	Central IT primarily concerned with infrastructure
IT Staff	Strong centralized staff	Centralized and distributed staff	Centralized staff primarily concerned with infrastructure
Formal PMO	Essential	Formal methodology essential; PMO less so	PMO has little authority if there at all
Strategy	Centralized, planned	Mixed, but agency- wide coordination	Little central coordination
Data Sharing Laws and Privacy	Must support data consolidation	Must support selective consolidation at least	Model can tolerate less permissive laws

Model Comparison: System Features

Factor	Centralized/ Warehouse	Cooperative	Distributed
Use of MPI	Strong; COTS or Custom-developed	Not mandated but may be present	Not usually a strong feature
Record De-duplication Strategy	Embedded in MPI and its services	Embedded in MPI and its services or offered separately	At best offered in individual participating systems
Security	Easier to maintain given centralization of services	More challenging given loosely coupled nature, > interfaces	In the hands of the individual system managers
System Acquisition Style	COTS or custom-developed	Interfaces tend to be custom-developed	Interfaces tend to be custom-developed
Support for Analysis	Centralized, planned	Mixed, but agency-wide coordination	Little central coordination

Model Comparison: Process-related Attributes

Factor	Centralized/ Warehouse	Cooperative	Distributed
Service Delivery	More centralized delivery model	Either centralized or distributed	More distributed delivery model
Technical Standards Enforcement	Easier, as more effort is central	Moderate, as compliance enables interoperability	Harder, as systems are largely stand-alone
System Requirements	More stable and clear	Moderately stable and clear	Less stable and often unclear
System Development Coordination	Very coordinated	More independent but can be coordinated	Largely independent and uncoordinated
Technical Innovation	Less interested	Moderately interested	Not very interested

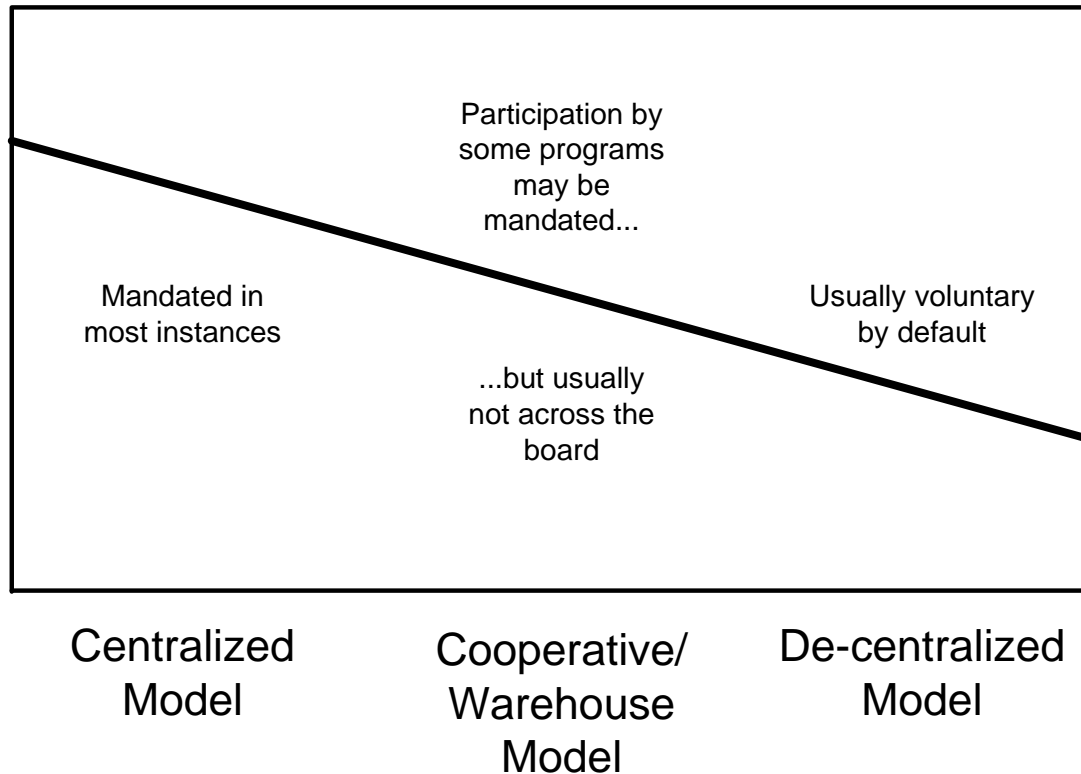
Model Comparison: Process-related Attributes

Factor	Centralized/ Warehouse	Cooperative	Distributed
Technical Risk	Fairly high, but so is potential gain	Moderate	Fairly low, but so are potential gains
Deployment Timetable	Incremental, but requires critical mass to activate	Incremental, but coordinated	Incremental, little coordination
System Deployment Style	Tightly-coupled	Loosely coupled	Uncoupled; usually replicated services
Cost	Higher up-front, though overall cost may be lower; software license cost may benefit from centralized approach	Moderate up-front cost, though overall cost usually higher	Cost widely distributed so difficult to track and understand; software costs can be higher though coordinated purchasing can help

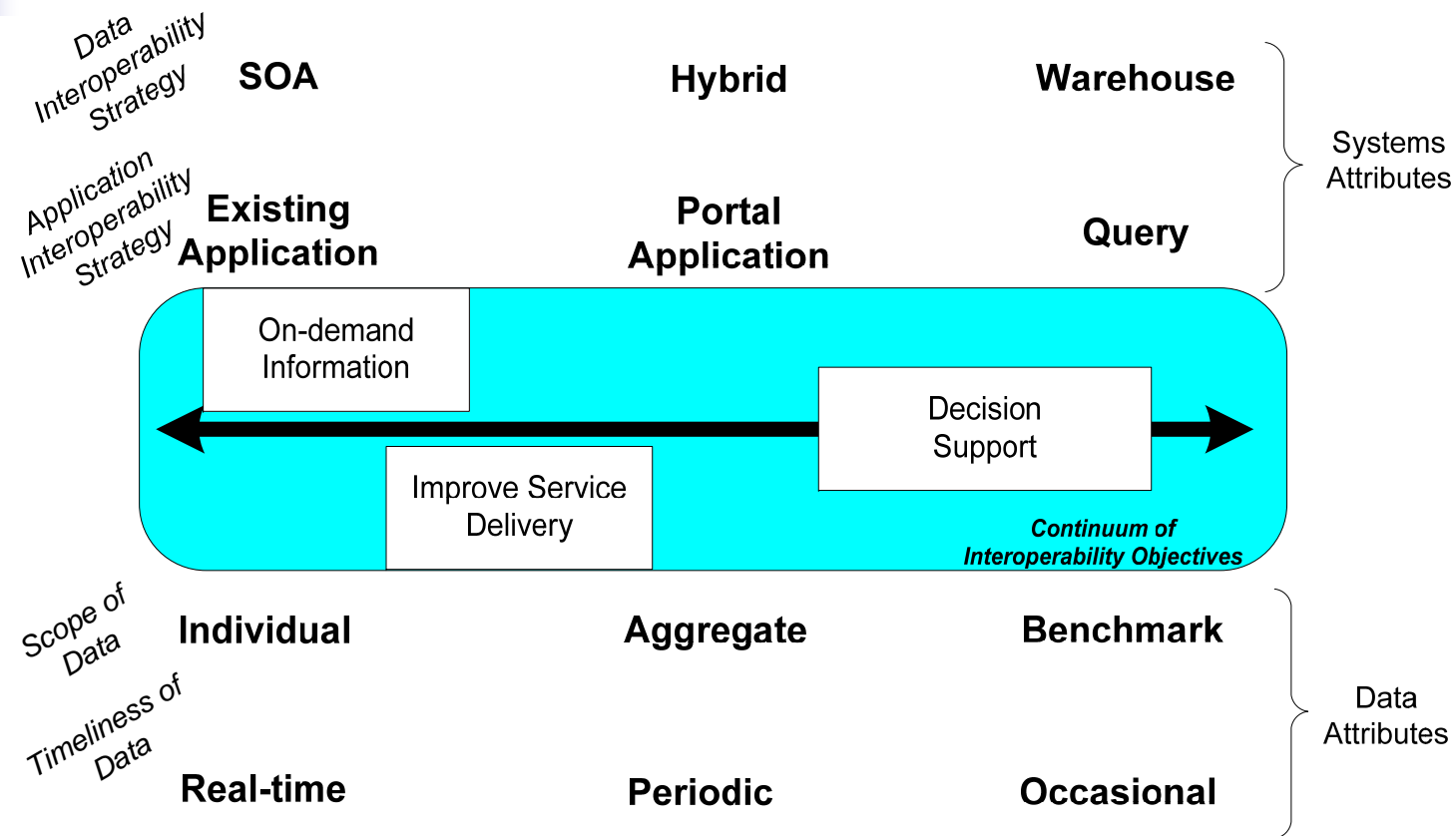
Model Comparison: Summary

Voluntary

Mandated



Model Comparison: Bringing Concepts Together



Additional Information



HLN's "Insights" at

<http://www.hln.com/resources/index.php>

**HLN's "Evolution of Public Health
Information Systems:
Enterprise-wide Approaches" at**

[http://www.hln.com/assets/pdf/
UT-White-Paper-Final.pdf](http://www.hln.com/assets/pdf/UT-White-Paper-Final.pdf)



PHII's "Unique Records Portfolio" at

<http://www.phii.org/pages/Portfolio-preOrder.html>

10/27/2008



Contact Information

Noam H. Arzt

President, HLN Consulting, LLC

858-538-2220 (Voice)

858-538-2209 (FAX)

arzt@hln.com

<http://www.hln.com/noam/>