

What Is the Best Healthcare Data Warehouse Model? Comparing Enterprise Data Models, Independent Data Marts, and Late-Binding™ Solutions

By Steve Barlow

“The enterprise data model approach to data warehouse design is a top-down approach that most analytics vendors advocate today. In this approach, your goal is to model the perfect database from the outset—determining in advance everything you’d like to be able to analyze to improve outcomes, safety and patient satisfaction. And then you structure the database accordingly.”

Want to know the best [healthcare data warehouse](#) for your organization? You’ll need to start first by modeling the data, because the data model used to build your healthcare enterprise data warehouse (EDW) will have a significant effect on both the time-to-value and the adaptability of your system going forward.

I’d like to take the opportunity here to outline the strengths and weaknesses of the two most common relational data models and compare them to the [Late-Binding™](#) approach we use at Health Catalyst®.

But first, a quick note. When I talk about “binding” data, I’m referring to the process of mapping data aggregated from source systems to standardized vocabularies (e.g., SNOMED and RxNorm) and business rules (e.g., length-of-stay definitions, ADT rules) in the data warehouse—optimizing data from all these different sources so that it can be used together for analysis. Binding data in such a way is required in any relational database model.

Each of the models I describe below bind data at different times in the design process, some earlier, some later. As you’ll see, we believe that binding data later is better.

Enterprise Data Model Approach

The enterprise data model approach to [data warehouse](#) design is a top-down approach that most analytics vendors advocate today.

In this approach, your goal is to model the perfect database from the outset—determining in advance everything you’d like to be able to

Extracting data from existing systems and making it all play well together in a net-new system is like trying to transform an apple into a banana. With patience, the right skills, and a bit of magic, it's possible—but it is incredibly time-consuming and expensive.

analyze to improve outcomes, safety and patient satisfaction. And then you structure the database accordingly.

In theory, if you're building a new system in a vacuum from the ground up, this is the way to go. But in the reality of healthcare, you're not building a net-new system when you implement an EDW. You're building a secondary system that receives data from systems already deployed. Extracting data from existing systems and making it all play well together in a net-new system is like trying to transform an apple into a banana. With patience, the right skills, and a bit of magic, it's possible—but it is incredibly time-consuming and expensive.

In all my years in the healthcare analytics space, I've never seen a project using this approach bear much fruit until well after two years of effort. This delayed time-to-value is a significant downside of this model. Binding the data and defining every possible business rule in advance takes a lot of time.

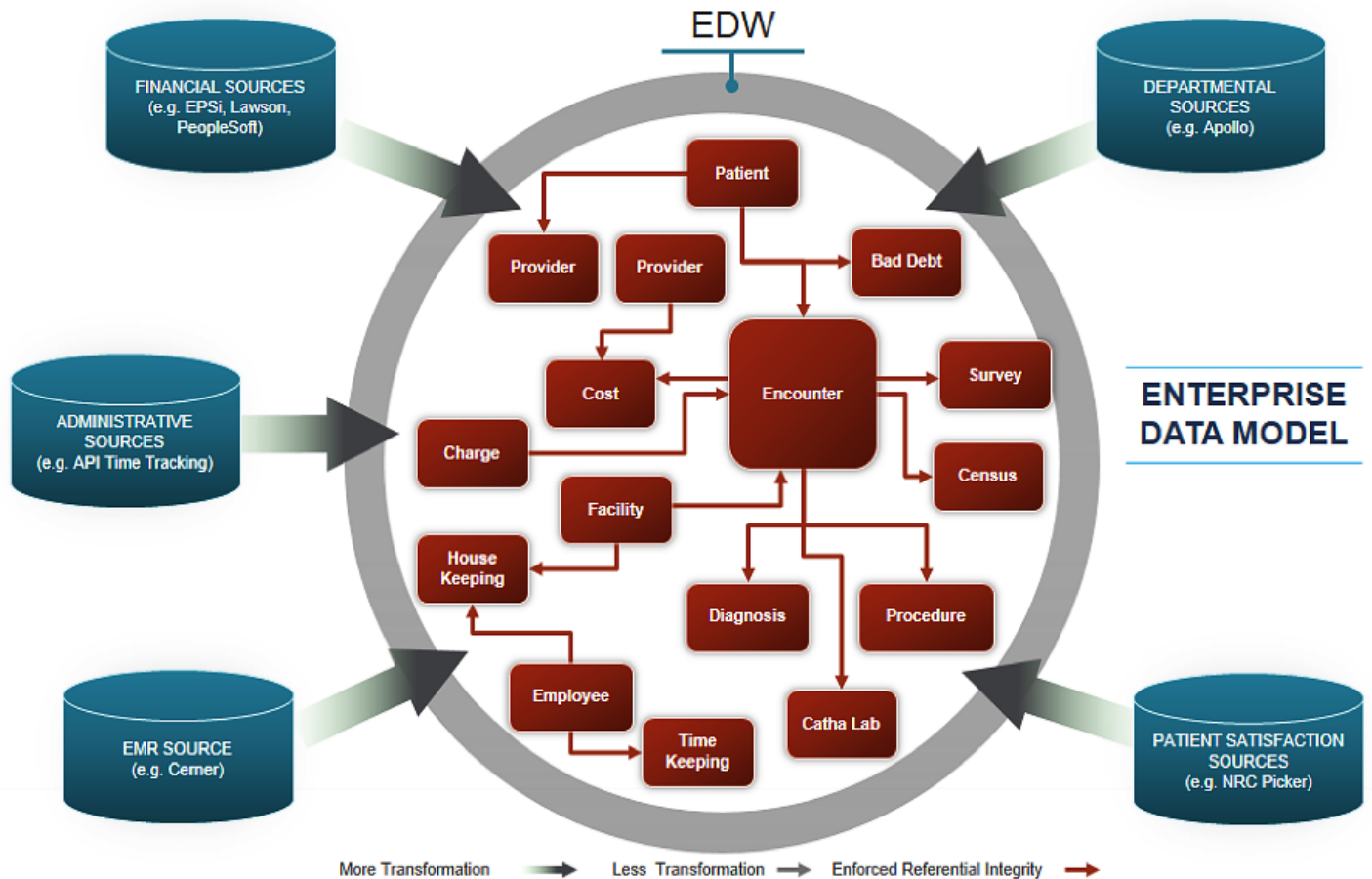
Two further drawbacks of the approach are:

- This model binds data very early, and once data is bound, it becomes very difficult and time-consuming to make changes. In healthcare, business rules, use cases, and vocabularies change rapidly. By the time you've spent two years turning your apple into a banana... you may find that what you really need now is an orange. But because your data was bound to rules and vocabularies from the outset, you're stuck with the banana.
- This model tends to disregard the realities of the data your organization actually has available. In an ideal world, you may want to measure cost per case or diabetes care. But do you currently capture the data that can give you those answers? The better, more realistic approach is to build your EDW to the data you already have, moving toward your ideal incrementally. The enterprise data model does not allow for such an incremental approach.

Here is the enterprise data model in graphic form:



Enterprise Data Model



Independent Data Mart Approach

The independent data mart approach to data warehouse design is a bottoms-up approach in which you start small, building individual data marts as you need them. If you want to analyze revenue cycle or oncology, you build a separate data mart for each, just bringing in data from the handful of source systems that apply to that area.

The benefit of this approach is that you can start implementing and measuring much more quickly—a big difference from the two- to five-year lifecycle of the enterprise data model approach.

At Health Catalyst, we advocate a late-binding approach to data modeling that overcomes the challenges inherent in the architectural models I just described. It is an adaptive, pragmatic approach designed to handle the rapidly changing business rules and vocabularies that characterize the healthcare environment.

However, here are three major drawbacks of this model:

- With all of these isolated data marts in place, you don't have an atomic-level data warehouse from which to build additional data marts in the future. Typically, data marts do not contain data at the lowest level of granularity. Data transformed in a data mart is usually summarized up a level or two. This means that the data mart may present you with information that a certain metric is below your benchmark, but it doesn't contain the granular data that enables you to dig down and determine why that metric is low.
- This model bombards source systems repeatedly and unnecessarily. You must build redundant feeds from each source system to feed these data marts. Imagine building a new feed into your EHR for every data mart you deploy: heart failure, pregnancy, asthma, diabetes, oncology ... and the list could go on and on.
- Like the previous model, this approach binds data quite early in the process. As data is brought into each independent data mart, it is mapped into the predefined data model—inhibiting the adaptability of the analytics solution.

Health Catalyst Late-Binding™ Approach

At Health Catalyst, we advocate a late-binding approach to data modeling that overcomes the challenges inherent in the architectural models I just described. It is an adaptive, pragmatic approach designed to handle the rapidly changing business rules and vocabularies that characterize the healthcare environment.

Here is a high-level description of the late-binding process:

- We take data in its atomic form from source systems and bring it into source marts within the EDW. We do not impose transformation on it yet; rather, we try to keep the data as raw as possible in the source marts. We do perform minimal data conformance at this stage. For example, we make sure that the "patient name" field in one source mart is structured the same as "patient name" in another source mart. But we don't bind the data to any volatile business rules or vocabularies at this point. We minimize remodeling data in the data warehouse until the

“There is an enormous shortage of experienced data and software engineers in healthcare — in fact, across all industries. There simply aren’t enough skilled people available in the labor market to sufficiently lower the risk of building your own EDW.”

analytic use case requires it and rely instead on the natural data models of the source systems as much as possible.

- Drawing from these source marts, we create a data mart. At this point, we perform some transformation of the data, but we only bind the data when necessary—when a specific business driver or use case calls for it. When we determine that we want to analyze a specific use case—for example, identifying and reducing the number of elective deliveries prior to 39 weeks’ gestation—we bind the data more tightly in that specific data mart.
- This incremental process features ideal binding points for data rules and vocabulary throughout. We’ve identified approximately 20 core data elements that are fundamental to almost all analytic use cases in the healthcare industry. Because these data elements don’t change often, we bind to them early. We bind other, more volatile rules and vocabularies as late as possible.

Our approach is something like just-in-time data binding. Rather than trying to hammer out a data model up front when you can only guess at what all the use cases for the data will be, you bind the data late in the process to solve an actual clinical or business problem. You don’t have to make lasting decisions about your data model up front when you can’t see what’s coming down the road in two, three or five years.

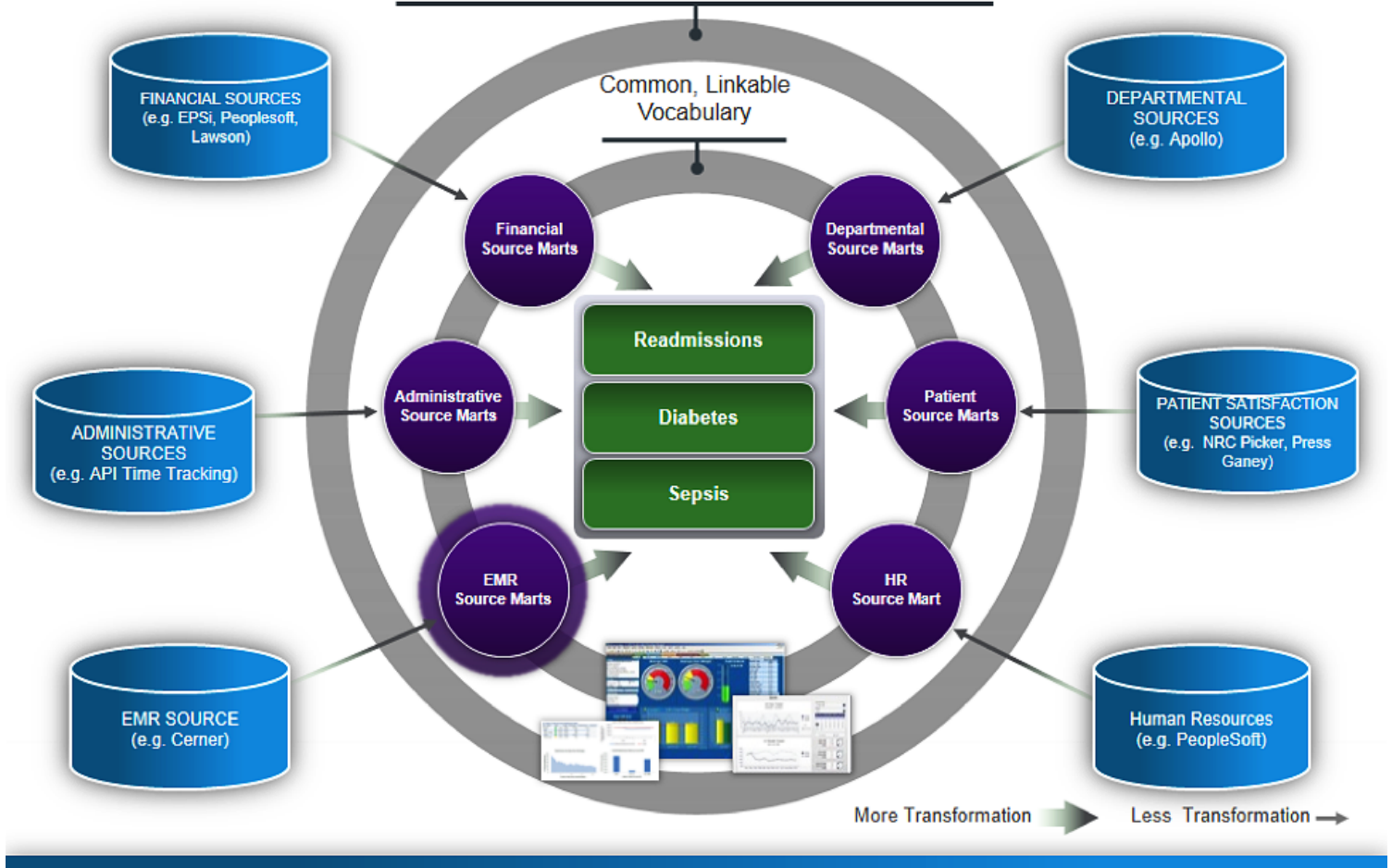
The late-binding approach gives you maximum flexibility for using your data to tackle a wide variety of use cases as the need arises. And it prevents you from wasting a lot of resources.

Here is the Health Catalyst Late-Binding™, adaptive data model in graphic form:



Adaptive Data Warehouse

Metadata: EDW Atlas Security and Auditing



Resources

- Clinical Data Warehouse: Why You Really Need One <http://www.healthcatalyst.com/clinical-data-warehouse-why-you-need-one>
- Late-Binding™ Data Warehouse Platform <http://www.healthcatalyst.com/late-binding-data-warehouse-platform>
- Clinical Data Warehouse: Why You Really Need One <http://www.healthcatalyst.com/clinical-data-warehouse-why-you-need-one>



About the Author

Mr. Barlow is a co-founder of Health Catalyst and former CEO of the company. He oversees all development activities for Health Catalyst's suite of products and services. Mr. Barlow is a founding member and former chair of the Healthcare Data Warehousing Association. He began his career in healthcare over 18 years ago at Intermountain Healthcare, and acted as a member of the team that led Intermountain's nationally recognized improvements in quality of care delivery and reductions in cost. Mr. Barlow holds a BS from the University of Utah in health education and promotion.

ABOUT HEALTH CATALYST

Based in Salt Lake City, Health Catalyst delivers a proven, Late-Binding™ Data Warehouse platform and analytic applications that actually work in today's transforming healthcare environment. Health Catalyst data warehouse platforms aggregate and harness more than 3 trillion data points utilized in population health and ACO projects in support of over 22 million unique patients. Health Catalyst platform clients operate 96 hospitals and 1,095 clinics that account for over \$77 billion in care delivered annually. Health Catalyst maintains a current KLAS customer satisfaction score of 90/100, received the highest vendor rating in Chilmark's 2013 Clinical Analytics Market Trends Report, and was selected as a 2013 Gartner Cool Vendor. Health Catalyst was also recognized in 2013 as one of the best places to work by both Modern Healthcare magazine and Utah Business magazine.

Health Catalyst's platform and applications are being utilized at leading health systems including Allina Health, Indiana University Health, Memorial Hospital at Gulfport, MultiCare Health System, North Memorial Health Care, Providence Health & Services, Stanford Hospital & Clinics, and Texas Children's Hospital. Health Catalyst investors include CHV Capital (an Indiana University Health Company), HB Ventures, Kaiser Permanente Ventures, Norwest Venture Partners, Partners HealthCare, Sequoia Capital, and Sorenson Capital.

Visit www.healthcatalyst.com, and follow us on [Twitter](#), [LinkedIn](#), [Google+](#) and [Facebook](#).