

Paraphrase of Twitter Stream from “Sentient Enterprise” Talk by Oliver Ratzesberger

The following is a paraphrase of the #TD3PI twitter stream during the talk by Oliver Ratzesberger on Sentient Enterprise at the Teradata Third-Party Influencer briefing on June 5, 2014 in San Diego. The intent is to smooth the flow of the audience’s interpretation of the talk and avoid adding content. – R. Hackathorn

Sentient Enterprise

Over the past eight months, the Sentient Enterprise was developed by [@Ratzesberger](#) and [@MohanSawhney](#). Sentience is the ability to feel, perceive or experience subjectivity, so the intent is to enable enterprises to possess the same qualities. As the ultimate goal, a sentient enterprise is to analytics, what the [Six Sigma doctrine](#) is to quality.

Data anarchy is plaguing our industry. Most companies are sinking into data anarchy today. It is characterized by expenditures of time and money to sift through data manually, amid the chaos of constantly putting out fires.

It is a mistake for a company to create a data warehouse when it does not match the business needs. Data warehouses are often created backwards. The company builds it and then asks how they will use the data in the business.

The Five Stages

The [Strategic Analytics Maturity Model](#) for the Sentient Enterprise consists of five stages, as shown at right.

Stage 1 of the Sentient Enterprise is **Agile Data Warehousing**, which evolves from data marts to virtual data marts (VDM) and then to data labs.



Marketing comes with an idea on the back of a napkin, and IT responds with a 50-page document. Business cannot wait. We must avoid monstrous development cycle that takes forever.

We have created data marts to be agile, but data accuracy and consistency often suffer. A data mart cannot be cheap enough to justify its existence. We need standard processes, governance, training, managed platforms, or data anarchy will result. For a company to be agile is not the Wild West. Data agile is very difficult to do properly.

A working prototype is 100x better than a requirements document.

eBay went from databases with 100x data duplication to an integrated data warehouse in 20 days. We then unplugged the old systems! This is like the old days when explorers burned their ships dashing hopes of returning home.

Anyone can provision their own virtual data mart (sandbox or data lab) and start using it within 5 minutes. The problem is not the creation of data marts; it is the implications of ungoverned data later. ‘Timebox’ it! And then turn it off! Eliminate unneeded duplicate data that

accumulates during re-architecting efforts. To combat data drift, eBay went to 'timeboxing' projects to 30, 60, 90 days. Sandboxes should expire, at which time users get an email warning of its termination (with an option to review).

Insight by @Steve_Dine, "As companies move off of traditional databases to analytical platforms, they often miss the opportunity to rationalize their data. The transition is often just a fork-lift of old to new without improving the data."

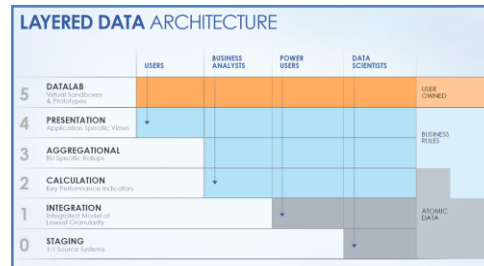
If you cannot be agile with structured data, you certainly cannot be agile with big data because it is so much harder to manage.

Insight by @ColinJWhite, "VDM are good but data must get into EDW. Separate investigative analytic platform is still required."

Layered Data Architecture

[Layered Data Architecture](#) (LDA) is: Staging, Integration, Calculation, Aggregation, Presentation and Data Lab.

Only design a project as comprehensive as necessary. Then add what works into the larger architecture. Do not stop when 'done'.



Experimental analytics do not always use enterprise data.

The problem with the enterprise data model is that you always get new data from the outside, which forces you to model the whole freaking world. We cannot model the whole world. So, do not try. Integrate at lowest granularity.

Insight by @Merv, "Focus on the golden attributes: agree, succeed, then extend."

Insight by @Tony Baer, "Summary of Oliver's message - First, governance; second, agile. They are not contradictory."

At the top of LDA are data labs along with virtual sandboxes and prototypes. Within data labs, we leverage analytics to determine what people are doing with data.

eBay build a data warehouse of DW usage, containing query logs to understand how VDM were being used. Within data labs, we leverage analytics to know what people are doing with the data. We also had to change the basic organizational model by embedding developers with end users, which is part of Stage 1.

Stage 2 is **Behavioral Data Platform** used to design/monitor behavioral patterns of DW users and in usage of subject-area data. Describe the "behavioral" data patterns by DW users and by subject area data. When semi-structured data is ingest, note the tags used and then model them. The value comes from behaviors, not from transactions. If the company cannot get behavioral data right, the company should not continue because the result will have a mess.

Metadata today doesn't come from some library but from the sum of interactions among us. Pattern detectors for these interactions are part of an 'instrumented data fabric' that responds/adapts to its own usage patterns.

Semi-structured data modeling in the Integration Layer of LDA is essential to managing it. NOSQL has become popular because you can add a field and just use it.

Stage 3 is the **Collaborative Ideation Platform** consisting of the data hub, which does not store data but enables the crowd-sourcing of data. The data hub was used 60K per month with storytelling as blog articles that linked to live Tableau dashboards. For example, Kaggle is the crowd-sourcing of data analytic competitions.

To survive metadata management with no-schema big data, you must crowd-source it by capturing and sharing it like LinkedIn crowd-sources metadata about professional's experiences, skills, and associates. Social interaction graphs about data usage.

Crowd-sourcing metadata may even involve gamification. At eBay, people solving a data problem got a prize. Instrument your own systems and turn on all that collaboration! Drive people's behavior to change by using search, social interactions, blogs and gaming to enhance collaboration.

Stage 4 is the **Analytic Application Platform**, which begins with deployment at zero cost; then the hard work begins. Lineage of analytic applications is as important as data lineage. Unfortunately, analytics is usually created after the fact. Turn that upside down. Do data listening that pushes relevant data to the company.

Has your company operationalized customer segments cross functionally? Most have not. The first customer segmentation app generated over a million segments, of which about a 1,000 were eventually used.

Data listening is critical to the future of data in enterprises. An example is [Disney's Magic Band](#) to enhance customer experience at their theme parks. All programming tools should push data to the Listening Framework in one line of code, using REST/JSON architectures.



Stage 5 is **Autonomous Decisioning Platform**. Avoid spending 90% of your time on finding data, instead your time should be spent on decision making.

Company should act as a single organism.

References

The following persons contributed tweets to the above paraphrase: @bevelson, @big_analytics, @charleshoman, @claudia_imhoff, @colinjwhite, @constellationrg, @czds, @dangraham_, @danwoodsearly, @datagenius, @davidpwu, @dbstodder, @dego963, @drnatalie, @dutchlight360, @ggheorghiu, @hackathorn, @jackwmson, @jadp, @ken_rosen, @kyield, @lindy_ryan, @maikgroenewegen, @marksmithvr, @matwg, @merv, @neilraden, @nrouda, @orzota, @prussom, @redjane, @richardwinter, @rmclane, @robpaller, @rosejason, @rwang0, @scott_gnau, @shawnrog, @steve_dine, @tcs_sap, @teradata, @tobyhudsonb2b, @tonybaer, @tonycosentinovr, @tracykowal, @valerie_fink, @williammcknight, @yvesmulkers.

The XLS twitter data is available [here](#), along with a PDF transcript [here](#).