Mike Kaiser
Sr. Manager

Agile Coach
Columbus Data and Analytics Lead
COHAA Board Member

Agile is Hard to Do With Data Warehousing (and Still Worth It)

Twitter: kaisermike
LinkedIn: mikekaiser
Who are you?

• Who here is currently doing (or has done) an agile data project?

• Team Roles
Agile is Hard to Do With Data Warehousing (and Still Worth It)

- Why agile for data warehouse
- Why data warehouse efforts can be harder to do in an agile way
- Some ways to increase agility on data warehouse efforts
CHALLENGES

EXECUTIVE COMMITMENT
“WE’RE NOT SEEING VALUE. THIS DATA IS NOT USEFUL.”

BI PROGRAM COST
BECAUSE USERS WERE NOT PART OF IMPLEMENTATION, CONTINUOUS REFACTORIZATION IS COMMONPLACE.

REALIZED VALUE
NEVER FULLY ACHIEVED BECAUSE THE BUSINESS WAS NOT ON THE JOURNEY.

COMMON CHALLENGES
• Large up-front effort to marshal executive willpower
• Users unable to articulate reporting requirements in advance.
• Heavy business time commitment
• IT labors independent of business input
• Unable to show value (for a long time)
• Focus on large amounts of extraneous or un-enriched data.
• Huge program with significant risk.
• Rarely achieve the value promised.

www.centricconsulting.com
A NEW APPROACH

- Executives can build confidence and commitment over time.
- Users only need to articulate how analysis can improve their capability.
- 4-6 hrs/week stakeholder commitment.
- IT is collaborates with business to create true value.
- New value delivered every 2-4 weeks ... or more frequently
- Focus only on data enrichment with known ROI.
- Low risk. Program grows with value delivered.
- Often over-delivers on anticipated value.
What if we delivered analytics every 2 weeks (...or less)?
<table>
<thead>
<tr>
<th></th>
<th>APPLICATION DEVELOPMENT</th>
<th>BUSINESS INTELLIGENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements</td>
<td>Requirements are tied to business process or have been in practice (manually) before.</td>
<td>Consumers of BI have difficulty visualizing what they would like to see. Most users have not had the capability of an advanced analytics and reporting tool, much less data mining and alerts.</td>
</tr>
<tr>
<td>Return on</td>
<td>ROI can be estimated from operational efficiency and easier identification of know issues.</td>
<td>ROI is generally difficult to quantify. How do you quantify the value of something you don’t know?</td>
</tr>
<tr>
<td>Investment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tooling</td>
<td>Developers have languages and development tools which are the product of years of maximizing toward agility (object oriented, service oriented, continuous integration...)</td>
<td>Developers have exciting new advancements with fast in-memory processing, multidimensional cubes, and cloud scaling. However, languages and tools lag behind those in application development with respect to agility.</td>
</tr>
<tr>
<td>Integration</td>
<td>Developers specify an input format or Web Service interface.</td>
<td>Master records must be cross-referenced and consolidated using complex ETL transformation.</td>
</tr>
<tr>
<td>Data Quality</td>
<td>Developers fully control what data comes in through the UI.</td>
<td>Analysts and developers must anticipate undefined data quality issues, resolve missing referential integrity and translate into domain values.</td>
</tr>
<tr>
<td>Visibility of</td>
<td>Issues are often hidden because users interact with one record at a time.</td>
<td>Users are highly aware of invalid data, seeing them on reports. In particular, missing referential integrity and invalid domain values make reporting awkward for users.</td>
</tr>
<tr>
<td>Issues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td>Interface is highly controlled and well-defined. Discrete behaviors may be tested.</td>
<td>There are no controls on data generated in source systems. This may cause a working ETL component to break in unexpected ways. Testers must be able to work with large data sets and not interfaces.</td>
</tr>
<tr>
<td>Utilization</td>
<td>Often users have no choice; they must use the operational system.</td>
<td>Users have developed their own backdoors and process for obtaining information. Business Managers are slow to change something that is working (even if ineffective)</td>
</tr>
<tr>
<td>Organizational</td>
<td>Often focused on a particular business area.</td>
<td>Often generates insights by spanning business areas.</td>
</tr>
<tr>
<td>Scope</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Got to get certain things right!
USER STORIES
What Are Epics

What They Are...
- Business Goals
- Statements of ROI
- Analysis Description
- Comprehensible to Business
- Ready for Prioritization

What They Are Not...
- Functional
- Report Requirements
- Delivery Estimates
- Limited to Available Data
- Detailed Specifications

The perspective of the individual using the analysis.

What information will be analyzed; what conclusions will be reached?

What does this allow them to do? How does analysis support business process?

What is the broader impact to having this capability? What would move a CEO to provide funding?

Are there any other specific criteria for success of this analysis?

BONUS: What are the metrics and elements by which the metrics may be sliced?
WRITE BI EPICS

ROLE
As a purchasing manager...

ANALYSIS
I need to better predict the quantity of ingredients to purchase each day. This is based on daily specials, customer load, freshness of existing stock and expected order demand.

CAPABILITY
So that we can avoid running out of food (unhappy customers) or throwing away food (waste) at the end of the day.

STRATEGIC VALUE
This directly impacts the cash flow and, ultimately, the profitability of the restaurant. If we have a repeat of last year, we’ll need to take out a loan.

CRITERIA
This would need to take into account our menu order history rather than purchase history. We could manually subtract current stock from the purchase recommendation since this data is not maintained in a system.

METRICS & SLICERS
Purchase Quantity, Usage Quantity, Ingredient, Menu Item, Vendor, Day Of Week, Season Of Year, (Holiday?) Flag

WHAT ARE SOME FOLLOW-UP QUESTIONS WE COULD ASK?
• Once you have this information, how will it be applied in the purchasing process?
• Do you just purchase the recommended quantity minus current stock?
• Do you plan specials around ingredient availability or vice-versa?

WHY ASK ADDITIONAL QUESTIONS?
Uncover related analysis needs, metrics, slicers and success criteria.
BI USER STORIES
INFORMATION MODELING
TRADITIONAL BI CHALLENGES

• Each source system brings its own data model, workflow, business specialization, terminology.
• Data warehouses are often somewhat generalized, but are influenced by each source.
• This is a problem when the same subject matter comes from multiple sources.
DOCUMENT INFORMATION MODELS

A NEW APPROACH

• Completely ignore source systems.
• Design the structure based on the business alone.
• Business model and processes are very slow to change.

Business Leadership

Let’s ask these people.
DOCUMENT INFORMATION MODELS

BETTER OUTCOMES

- Requirements conversations become easier
- Drastic reduction of data query complexity
- Simpler ETL transformations
- Comprehensible and consistent terminology
- Rigor of data definition

(Same People)
AGILE DELIVERY
Agile Paradigm Shifts

- Business involvement throughout the project
- Empiricism and experimentation
- Maximize learning with short cycles
- Small team size
- Transparency
2-4 WEEKS
VALUE BASED
DELIVERY

START JOURNEY — HERE

Agile BI
“I need everything before the data warehouse is useful...”

A Subject Area (Or Source System) Approach Seems To Make Sense:

• Deal with one system at a time.
• Bring like data together (referential integrity)
• Make one trip. Get as many data elements as you can.
Not quite true. But our users usually need more than one subject area.

It just doesn’t work...

- Business priorities span systems
- Most data requires transformation
- Business Analyst and users can’t enrich requirements fast enough.
- It takes a long time to deliver.
- A 3rd of Data Elements are rarely, if ever, used.
Only bring in data that is directly related to a User Story in the Current Release. Anything extra is “gold plating”.

- Every data element adds risk to the project.
- Every piece of code must be tested.
- Each additional development task takes time and puts the release in jeopardy.
- Business analysts may not have had time to discover how a data element will be used.
FAST FEEDBACK
THE TECHNOLOGY DEMO

THE SOFTWARE ISN'T 100% COMPLETE.

IF IT HAD A USER INTERFACE YOU WOULD SEE SOMETHING HERE... HERE... AND SOMETIMES HERE.

AND THEN YOU'D BE SAYING, "I GOTTA GET ME SOME OF THAT."

ANY QUESTIONS?
INCREASE THE RATE AND SPEED OF FEEDBACK

Focus of BI Testing
- Transformation Testing (ETL-ing well?)
- Exploratory Testing
- User Business Value Testing
- Performance and Scalability Testing

Challenges
- Available automation test harnesses
- Native skillset of those on the team

Similarities With Agile Web Dev
- Automated testing critical to agility over time
- It is less expensive and lower stress when you know about bugs sooner
Questions / Comments

Twitter: kaisermike
LinkedIn: mikekaiser