The Monadics!

TWO 2σ SIGMA

Robbie Gladmon • Matt Kotsenas • Derek Mansen

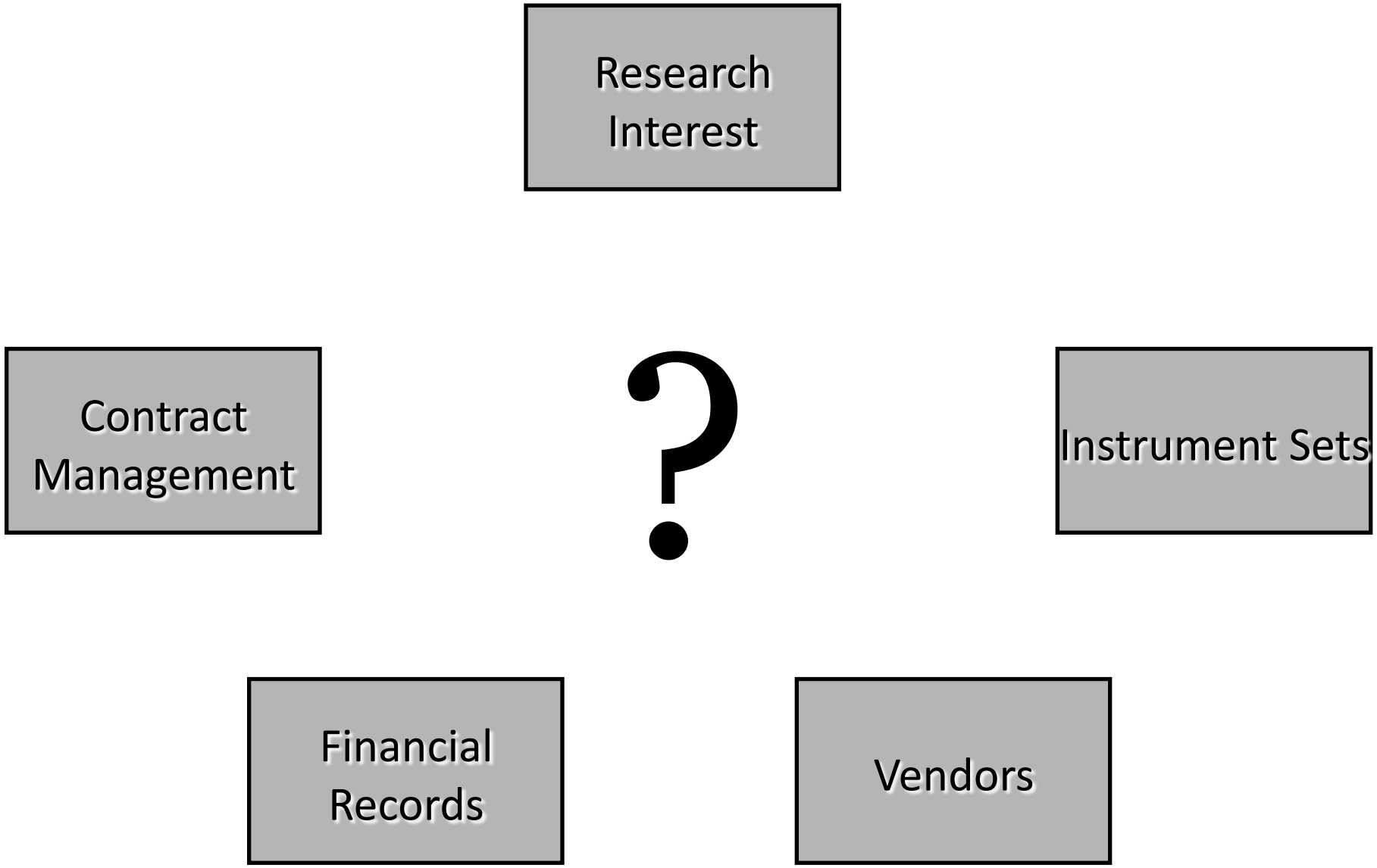
Tom Rudick • Marc Weil

Project Overview

Two Sigma's Domain

- Financial investment company
- Disparate Datasets

The Monadics! TWO SIGMA



The Monadics!

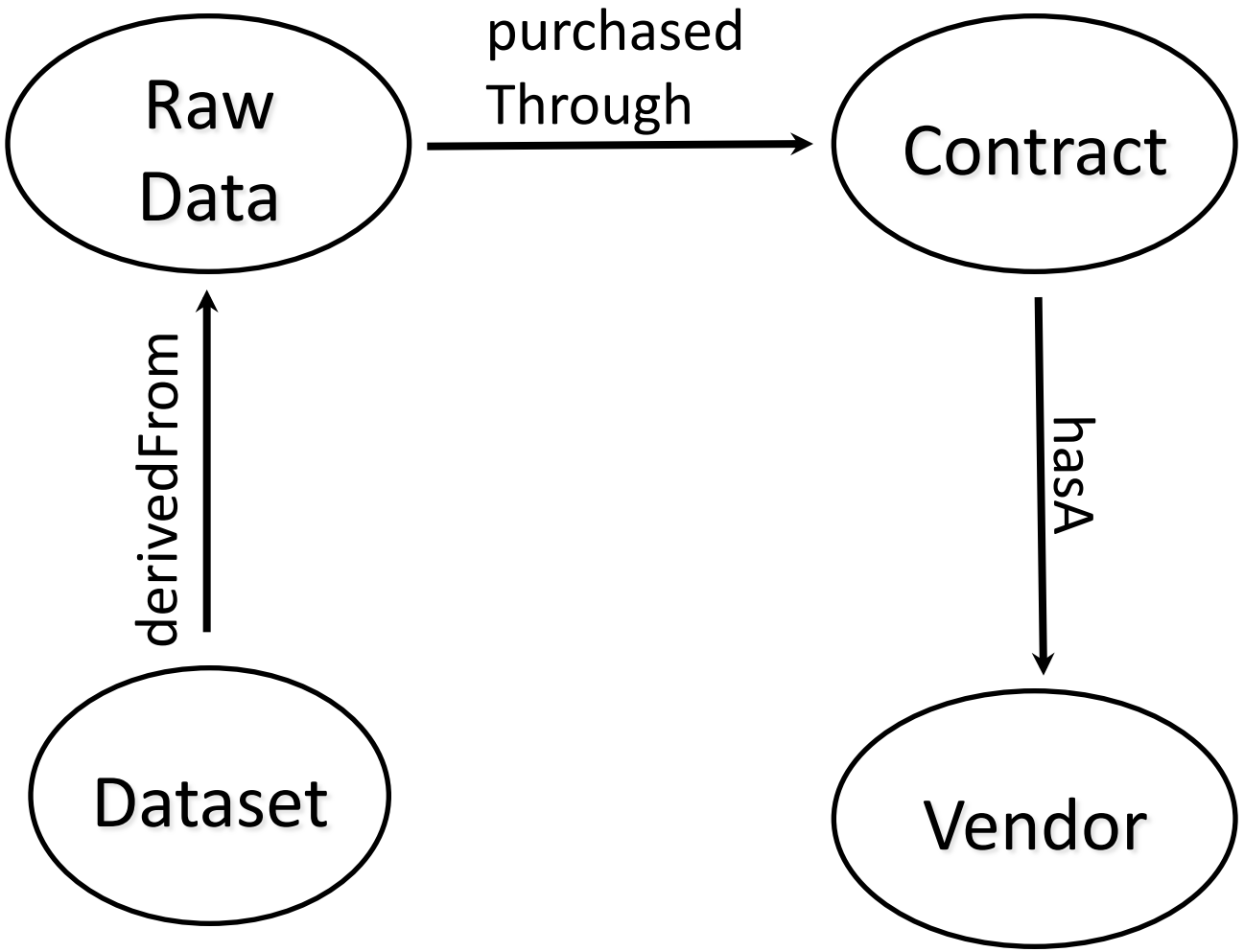
TWO SIGMA

Technologies

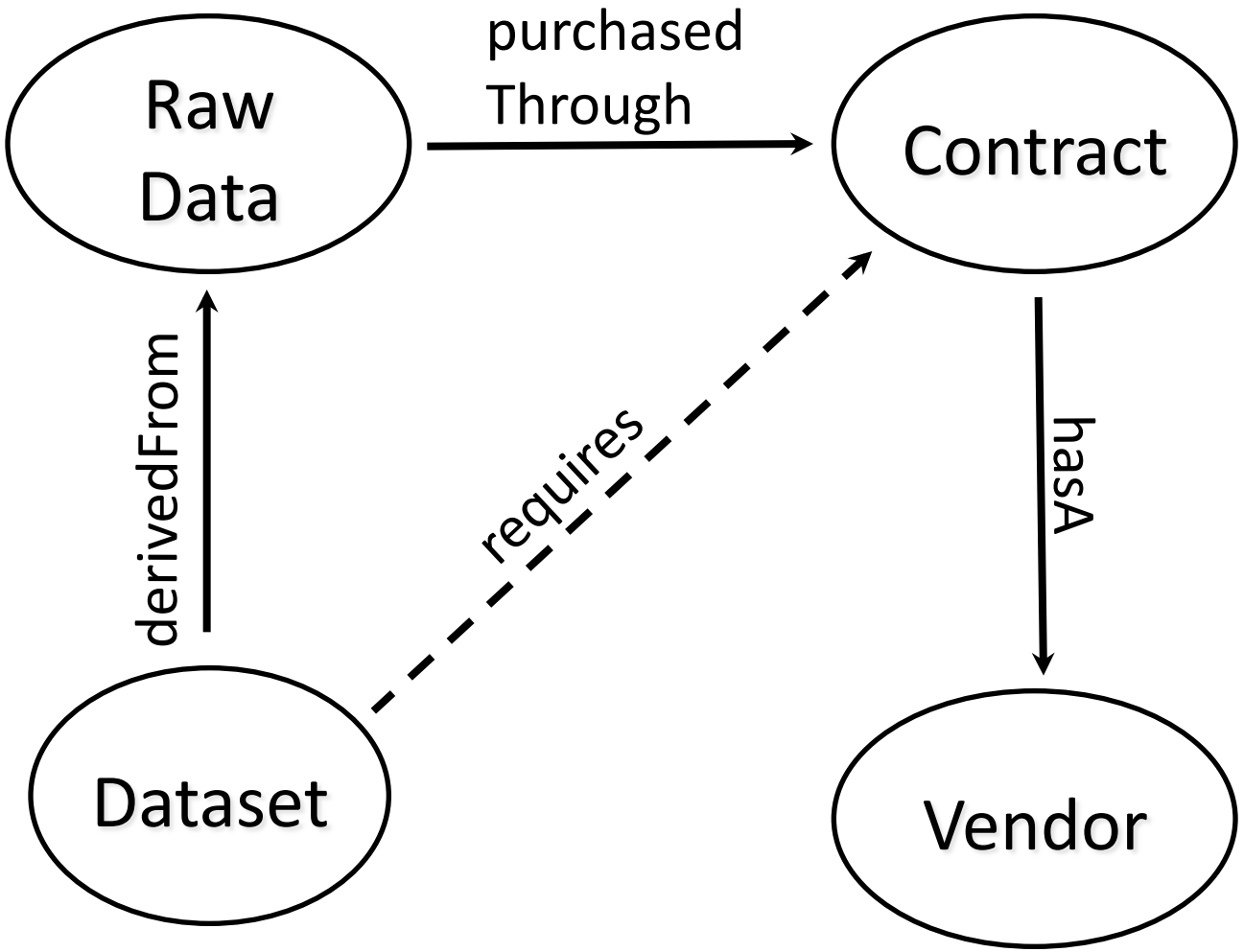
Ontologies

Ontology

(n.) a rigorous and exhaustive organization of some knowledge domain that is usually hierarchical and contains all the relevant entities and their relations



The Monadics! TWO SIGMA



The Monadics! TWO SIGMA

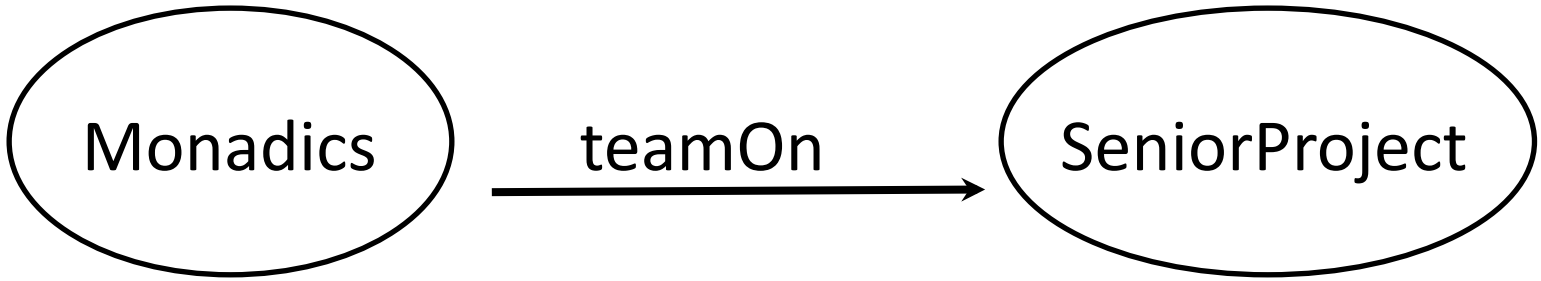
OWL & RDF

RDF

- Represents relationships between data
- Atomic building block of the Semantic Web
- Good for human-understandable presentation

The Monadics! TWO SIGMA

RDF Triples




```
<sp://teams/Monadics>
  sp:teamOn <sp://SeniorProject>.
```

The Monadics! TWO SIGMA

OWL

- Web Ontology Language
- Set Logic
- For when RDF needs to be machine-understandable

The Monadics! TWO SIGMA

SPARQL

SPARQL

- SPARQL Protocol and RDF Query Language
- Query language for ontologies

The Monadics! TWO SIGMA

SPARQL

```
SELECT * WHERE {
   ?x <teamOn> <SeniorProject>
}
```

Java EE

Java EE

- Tomcat 6
- Spring
- Open source tools

Jena

Jena

- High performance ontology framework
- Built-in reasoner

The Monadics! TWO SIGMA

TDB

Triple Database

- Ontology Datastore
- Fast Graph Database
- Scales to “Millions of Triples”

ActiveMQ

ActiveMQ

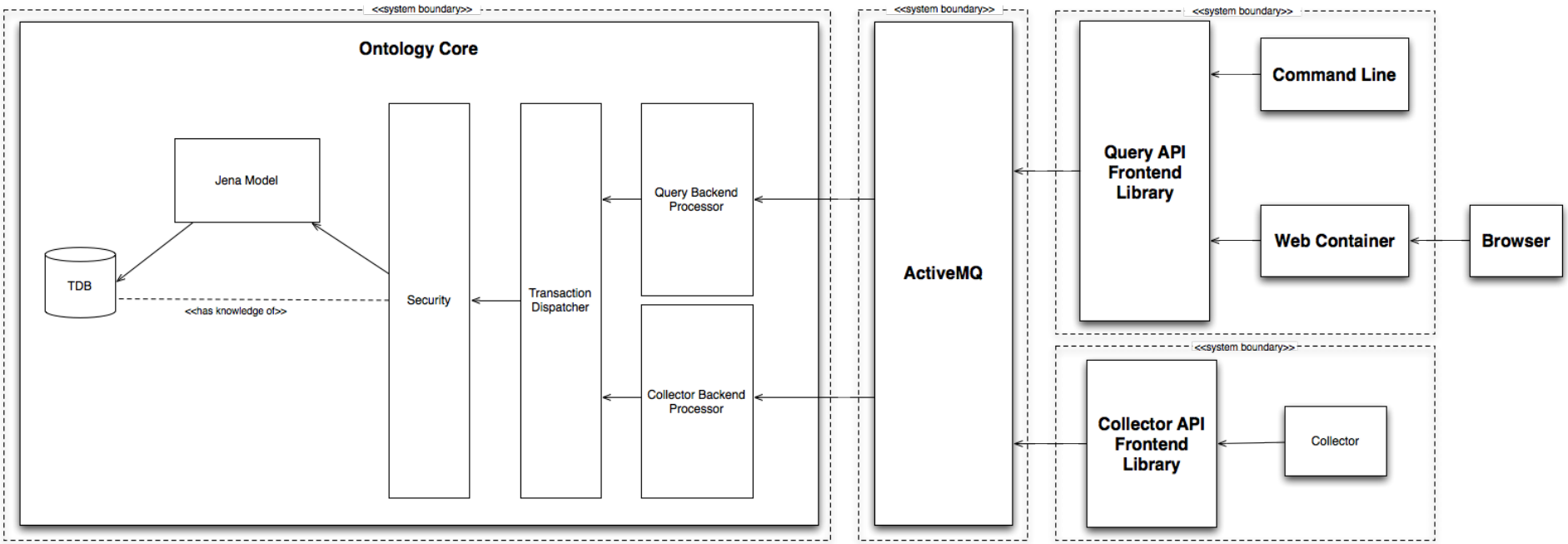
- Java Message Service implementation
- High performance
- High reliability
- High scalability

Groovy

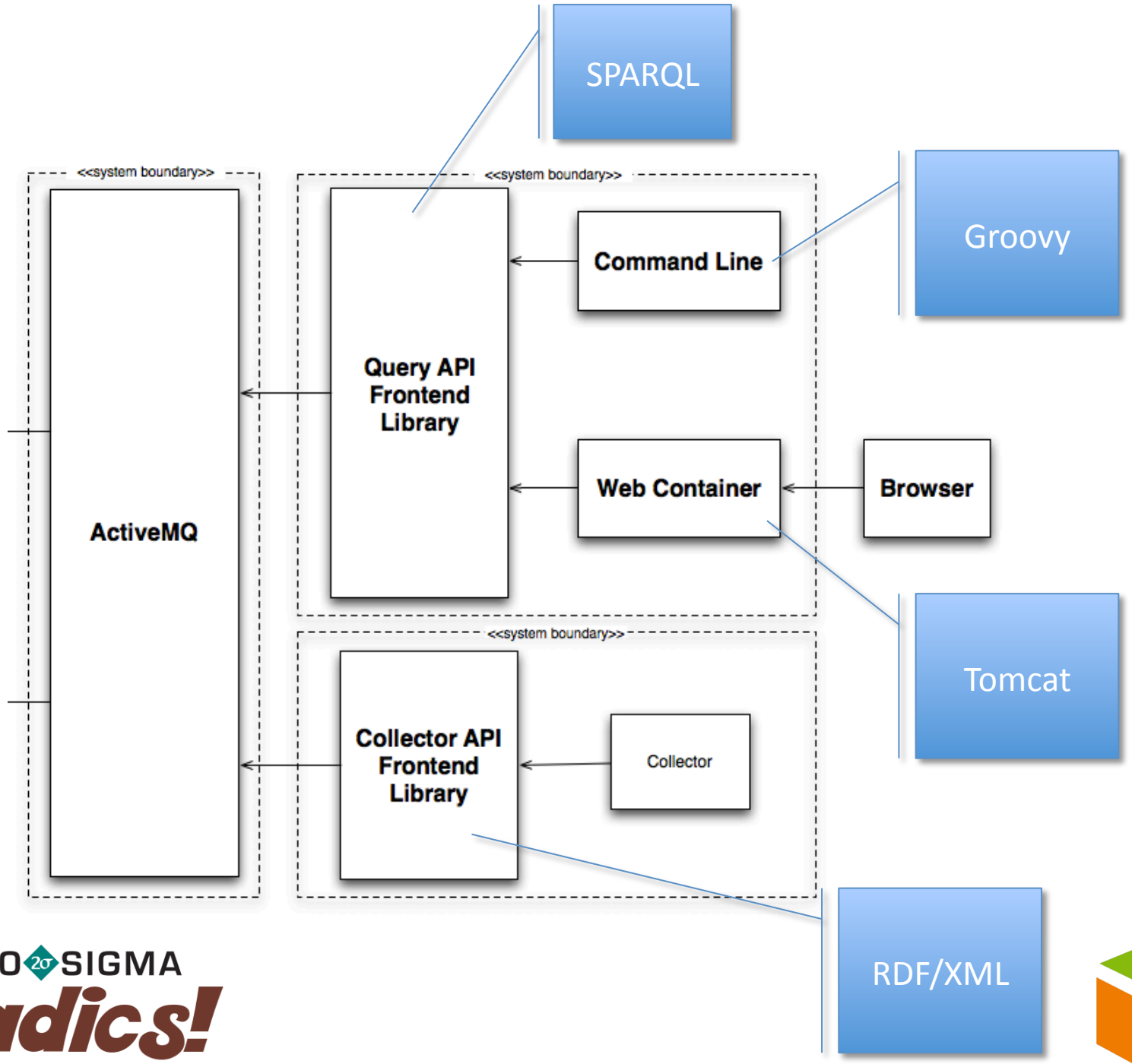
Groovy

- Dynamic scripting language
- Built on the Java Virtual Machine

System Architecture

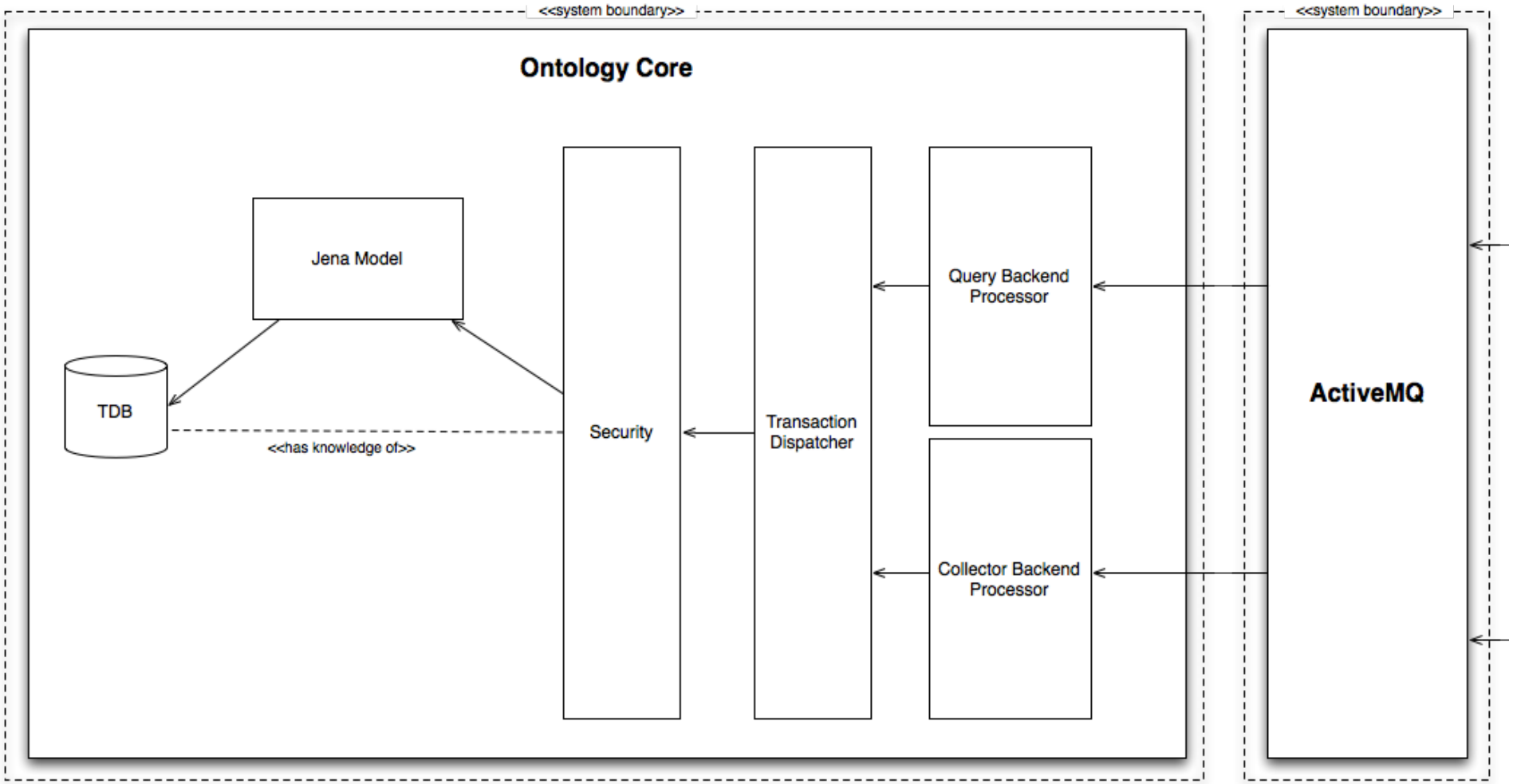


The Monadics! TWO SIGMA



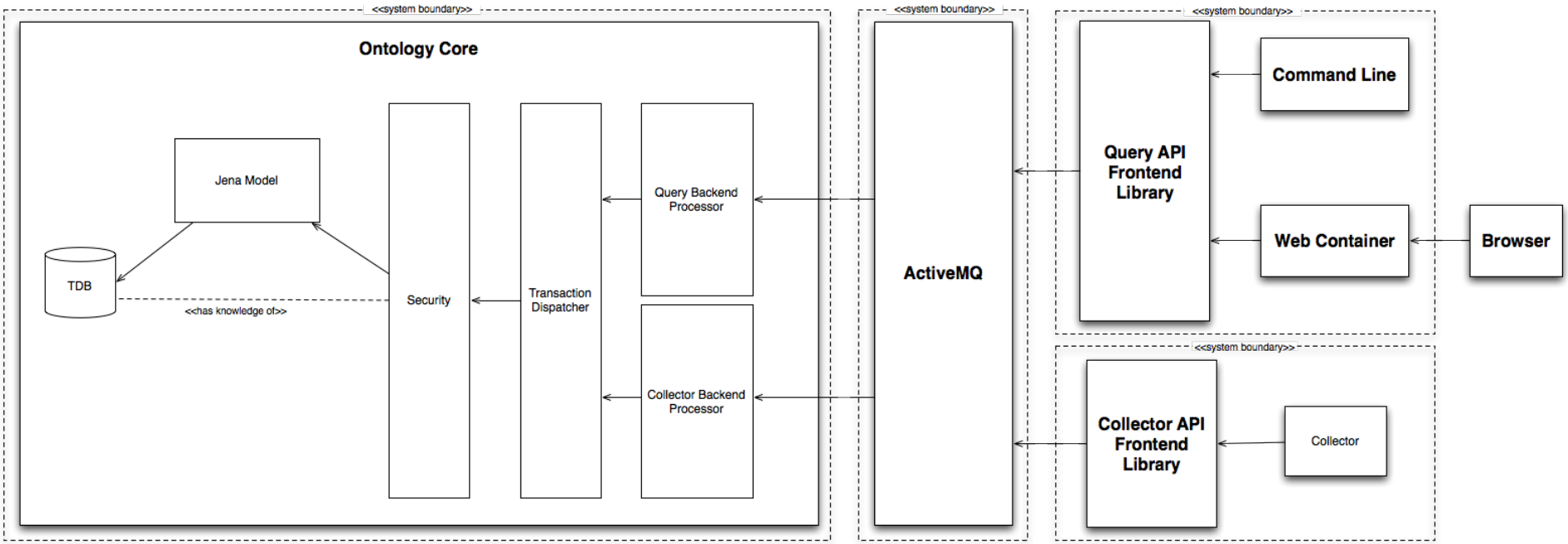
The Monadics!

TWO SIGMA



The Monadics!

TWO SIGMA



The Monadics! TWO SIGMA

Scoping Limitations

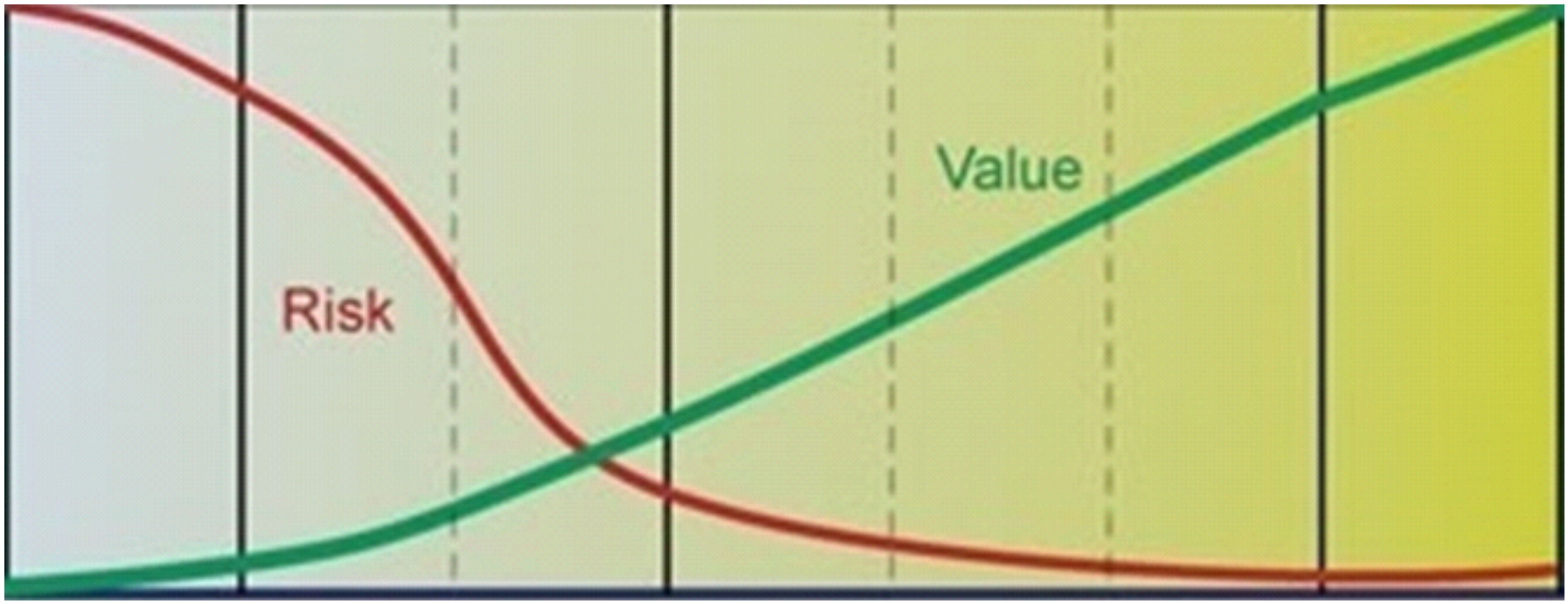
Scoping Limitations

- Do not have to write “collectors”
- Stub security system

Process

Lifecycle

- First Quarter: Evolutionary prototyping
- Second Quarter: Time-boxed iterations



Made available under EPL v1.0

The Monadics! TWO SIGMA

Lifecycle

- First Quarter: Evolutionary prototyping
- Second Quarter: Time-boxed iterations

Testing Strategy

Testing Strategy

- Performance
- Auto-generated test data
- Unit testing
- 80% Code coverage

Process Tools

- Trac
- Mailman
- Mercurial
- Hudson
- Ant

Risks

Risks

- High learning curve
- Great dependency on open source software
- Unclear or changing requirements
- Aggressive scheduling

Metrics

First Quarter Metrics

- Requirements volatility
- Risk volatility
- Hours worked

The Monadics! TWO SIGMA

Requirements Volatility

- High volatility in weeks 1-5
- Low volatility in weeks 6-8
- Expected higher volatility next quarter

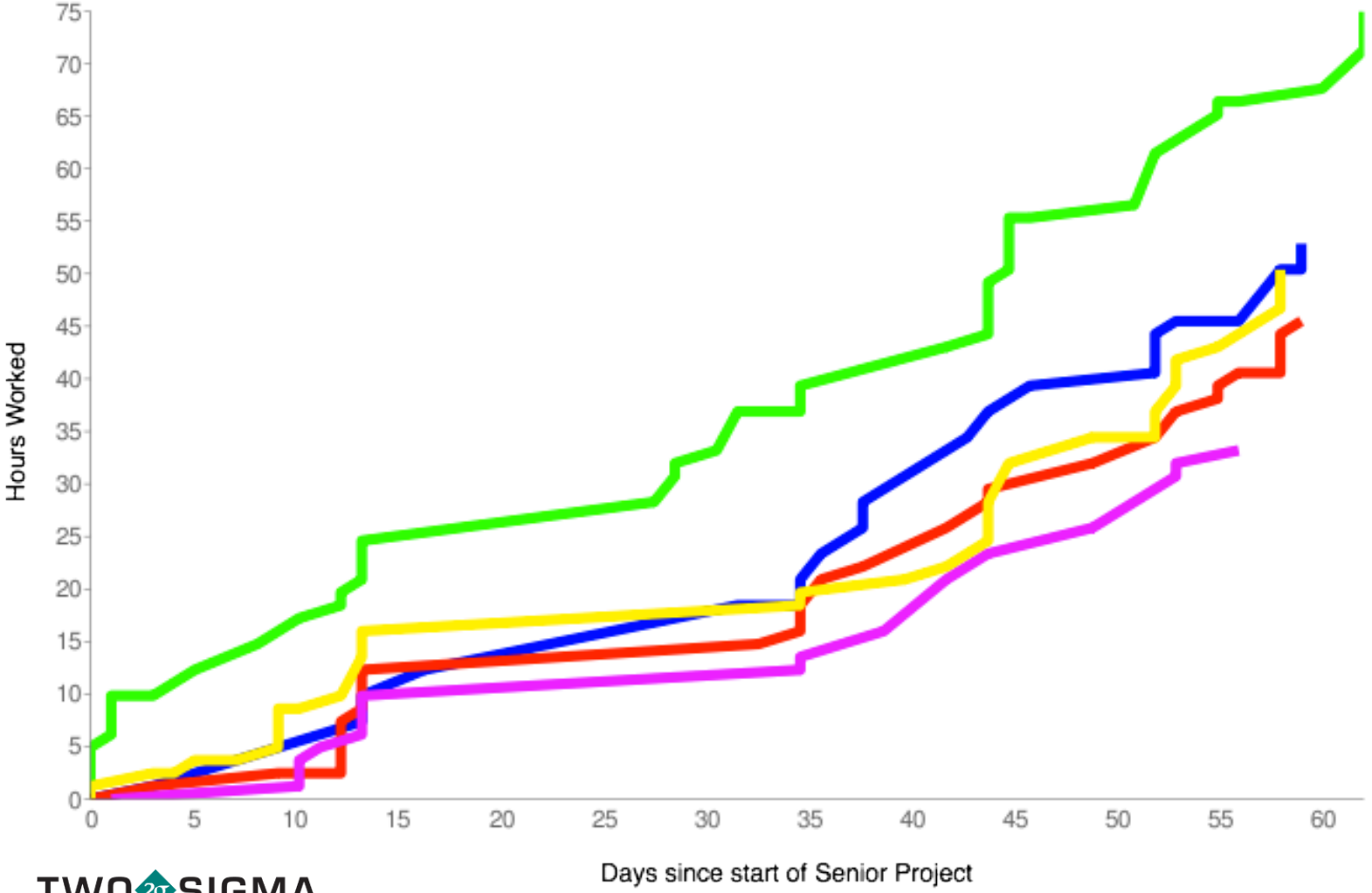
The Monadics! TWO SIGMA

Risk Volatility

- Low volatility in weeks 1-5
- High volatility in weeks 6-9
- Lower volatility next quarter

The Monadics! TWO SIGMA

Hours Worked



The Monadics!

TWO SIGMA

Commitments

First Quarter

- Thin vertical slice of functionality
- Stable deployable prototype
- Performance testing

The Monadics! TWO SIGMA

Second Quarter

- New methodology
- Updated metrics
- Adding features to prototype
- Department deliverables

The Monadics! TWO SIGMA

Current Project Status

- Two delivered prototypes
- Deployed on Two Sigma test server
- Currently being tested by Two Sigma

Questions?