Project Overview
Two Sigma’s Domain

• Financial investment company
• Disparate Datasets
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
Raw Data \( \xrightarrow{\text{derivedFrom}} \) Dataset \( \xrightarrow{\text{purchased Through}} \) Contract \( \xrightarrow{\text{hasA}} \) Vendor
OWL & RDF
RDF

• Represents relationships between data
• Atomic building block of the Semantic Web
• Good for human-understandable presentation
<sp://teams/Monadics> sp:teamOn <sp://SeniorProject>.
OWL

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

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

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

Java EE
Java EE

- Tomcat 6
- Spring
- Open source tools
Jena

- High performance ontology framework
- Built-in reasoner
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
Scoping Limitations
Scoping Limitations

• Do not have to write “collectors”
• Stub security system
Process
Lifecycle

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

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

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

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

- New methodology
- Updated metrics
- Adding features to prototype
- Department deliverables
Current Project Status

• Two delivered prototypes
• Deployed on Two Sigma test server
• Currently being tested by Two Sigma
Questions?