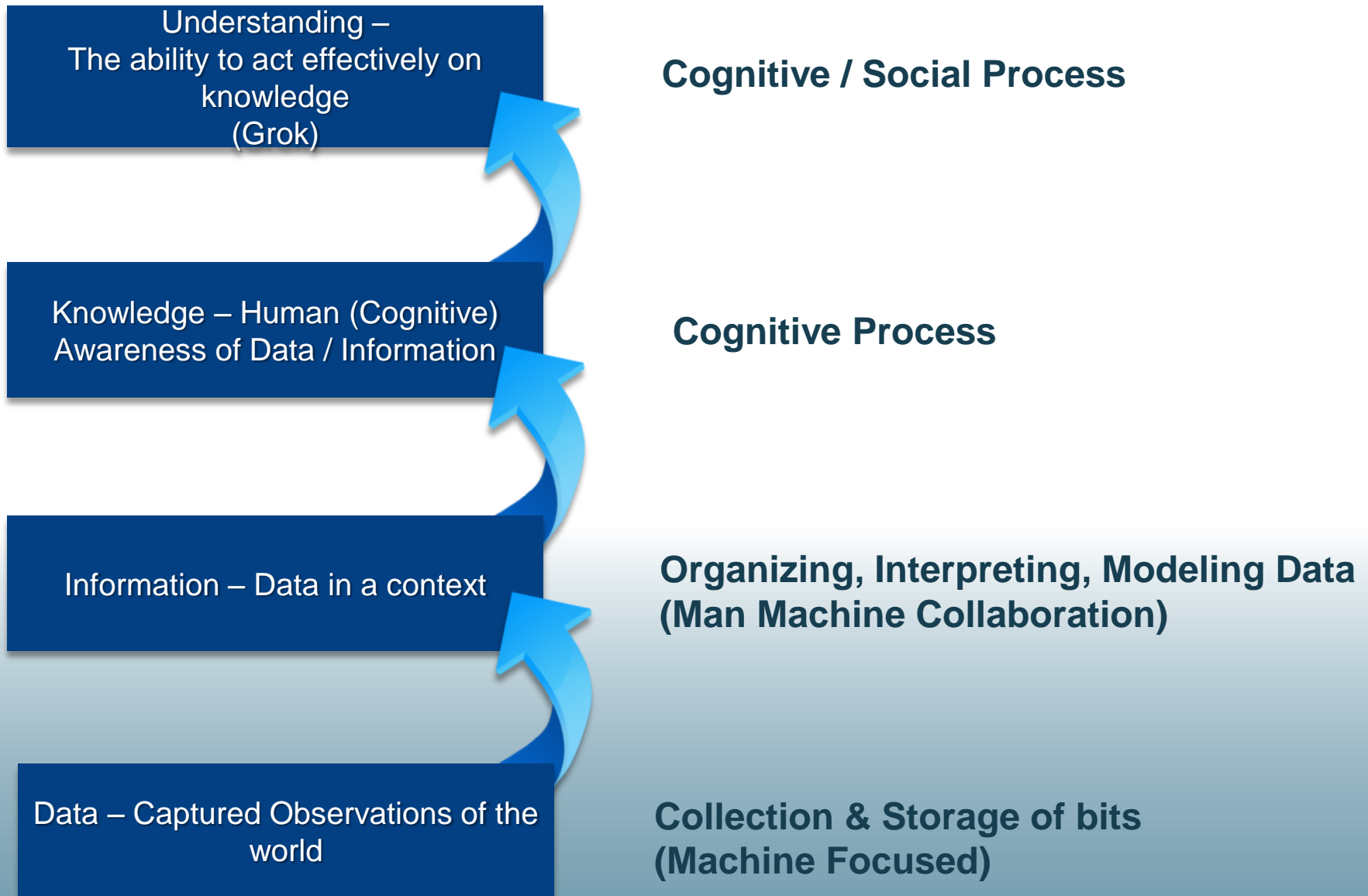# Open Source Analytics:

The Business Case

Dr. Dan Maxwell
KaDSci, LLC
dmaxwell@kadsci.com

Analysis is Data Rich & Insight Poor

Understanding – The ability to act effectively on knowledge (Grok)	**Cognitive / Social Process**
Knowledge – Human (Cognitive) Awareness of Data / Information	**Cognitive Process**
Information – Data in a context	**Organizing, Interpreting, Modeling Data (Man Machine Collaboration)**
Data – Captured Observations of the world	**Collection & Storage of bits (Machine Focused)**

Current Limitations of Analysis

» **Analyst developed models are usually developed using PC Based COTS Software**

  › Limits Scalability of models and analysis
  › Limits Sharability of models
    ↳ Inconsistencies in data
    ↳ Lack of documentation

» **Questions requiring formal analysis are often complex – Requiring multiple methods**

» **Analysis Teams are often geographically distributed**

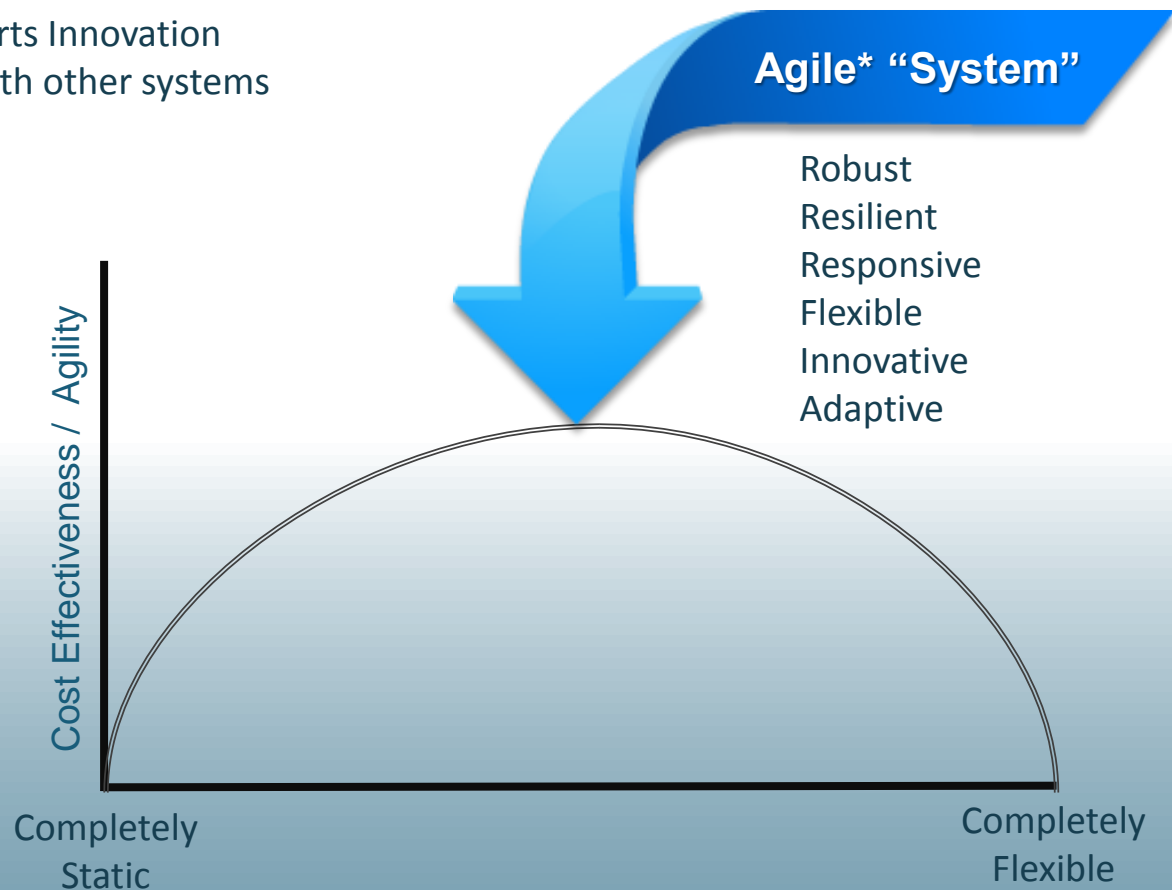» **Commercial Enterprise Level Analysis Software is sold as a "Solution"**
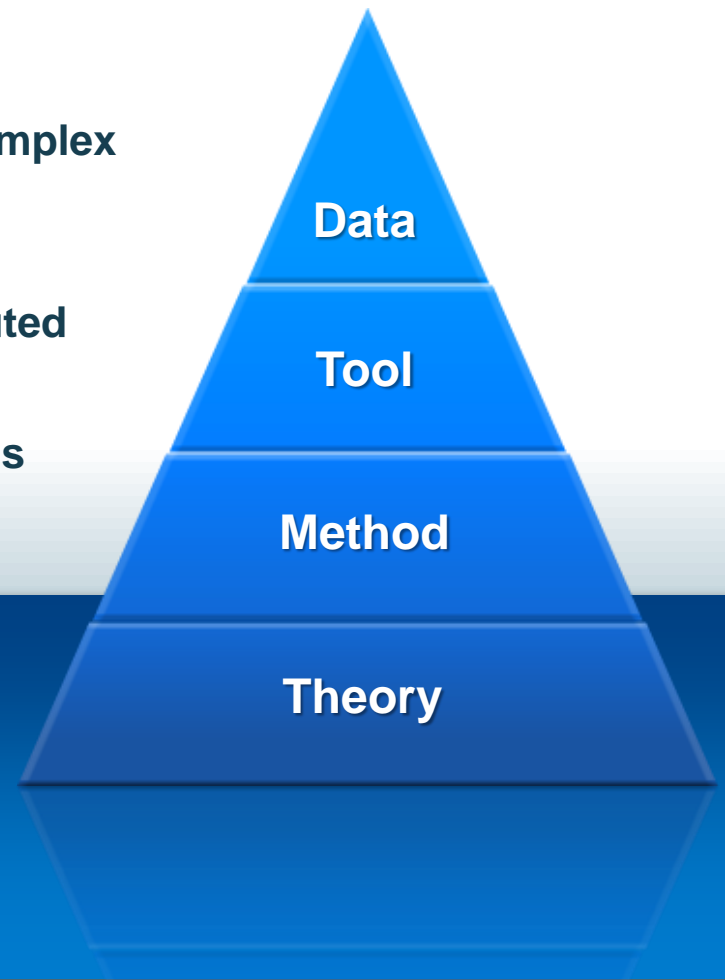
We are data rich and insight poor

Data

Tool

Method

Theory

Business Goal of Analysis Software

1. To help inform leadership decisions

2. To be cost effective!

Controlled Life Cycle Costs
Usable by Analysts
Flexible – Supports Innovation
Interoperable with other systems
Secure

Agile* "System"

Robust
Resilient
Responsive
Flexible
Innovative
Adaptive

Cost Effectiveness / Agility

Completely
Static

Completely
Flexible

*From "Power to the Edge", Alberts & Hayes 2003

Myths About Open Source Software

» **It is Free**

> Sometimes There are support costs
> There are labor costs to support implementation (Learning Curve)

» **Quality is low**

> Quality is "Lumpy"
> Really no different than COTS

» **It is Unsecure – Linus' Law says different**

» **It is a Panacea –**

> It doesn't fix bad processes
> There is a place for COTS and custom proprietary software
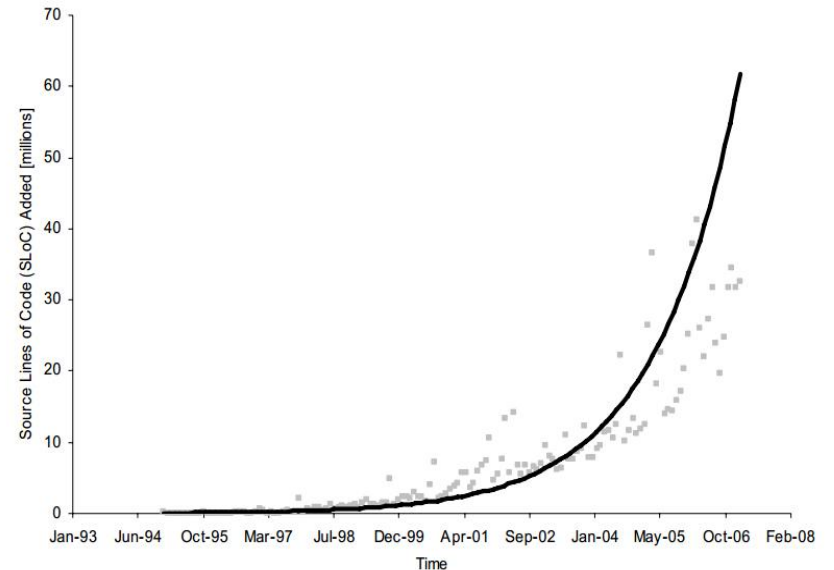
Assertions about Open Source

1. It Moderates cost because software development and maintenance is truly a service

2. It encourages innovation because there is no advantage to sitting on existing techniques – Always seeking Better, Faster, Cheaper

Open Source Availability is Growing

Millions of Lines of Code (1993-2008)
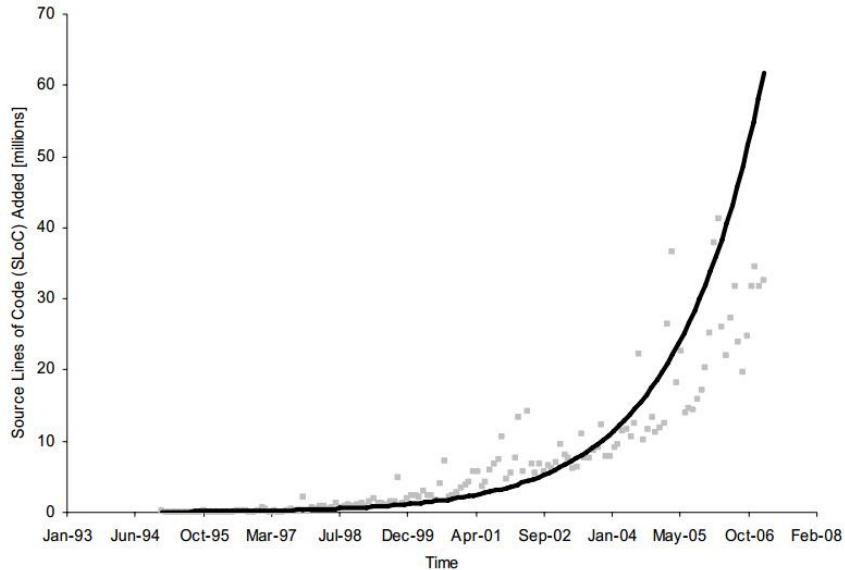
Number of Projects (1993-2007)

Extracted from "The Total Growth of Open Source", Riehle & Deshpande
http://dirkriehle.com/2008/03/14/the-total-growth-of-open-source/

Considerations for Transition

» **Consider "Life Cycle" Cost**

> Software Development
> Software Maintenance
> Data Management Costs
> Training

» **Look at Organization wide benefits**

> Agility
> Quality of Analysis

↳ Interoperability
↳ Analytic Foundation
↳ Improved Collaboration
↳ ......

Do a process oriented Life Cycle Cost Benefit Analysis

Backup

Open Source Analysis Software

» Database

- › Postgres
- › Hbase
- › MongoDB
- › Neo4j

» Statistical Analysis

- › "R" -- Implements the S Language
- › DEP
- › WEKA
- › PSPP

» Bayesian Networks

- › UnBBayes
- › JavaBayes

» Simulation

- › ASCEND
- › Facsimile
- › Galatea
- › NS2
- › SymPy

» Network Analysis

- › Gephi
- › Cytospcape
- › Cuttiefish
- › Jung

» Decision Analysis

- › MCDA

» GIS

- › GRASS
- › NASA World Wind

This is only a small sample…