If I was Thinking of My Own PhD Project, What would I Go For?

Vadim Ermolayev, PhD
Dept of Information Technologies
Zaporizhzhya National University

H2020 ICT, Big Data

SemData
FP7 Marie Curie IRSES

ICTERI

Kherson, Ukraine, 24 May, 2016

Myself

- **Assoc. Prof**., Dept of Information Technologies, Zaporizhzhya National Univ.
  - Also: **Expert**, Directorate General Information Society & Media, European Commission
  - Also: Industrial **Research Consultant**, several companies throughout my carrier

- More info and contact:
  - Personal Web Site: http://ermolayev.com/
  - LinkedIn Profile
  - ResearchGate Profile
  - Google Scholar Profile
  - Email address: vadim [at] Ermolayev [dot] com

- Current projects:
  - SemData (FP7 Marie Curie IRSES)
  - ICTERI Conference Series

Interests

- Broadly – in AI:
  - Dynamics, Change, Adaptability, Evolution – Evolving Knowledge
- Narrower – in Semantic Technologies
  - Ontology Engineering – refining ontologies to meet changing requirements
  - Ontology Learning – learning ontology (tokens) from data / texts
- Narrower – in Big Data / LOD Analytics
  - Using semantics / ontologies for improving quality and scalability of analytics
- Technologically – Distributed AI / Agents and MAS
  - Agent-based approaches to solve ST problems related to dynamics

Focal Problems of Interest

- Ontology Dynamics
- A Syndicated Ontology of Time ...
- Evolving Knowledge Ecosystems

Ontology Dynamics

vs

Ontology Change

(ST, Ontology Engineering)

Ermolayev, V.: The Law of Gravitation for Ontologies and Domains of Discourse. Computer Science Journal of Moldova, 23(2), 209 – 236, 2015 (PDF)

What is Changed, not Why …

- Is this one fit to go faster than 100 mph?

Ontology Change:
- Let's measure the distance passed in 1 hour
- If <= 100 mi, then NOT

Would be happy to have:
- Perhaps NOT, judging by its current shape
- Could be able however if thrown from there – up the cliff

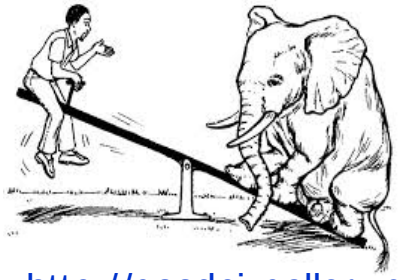
Open Issues?

- Ontology Change – studies the ways to change an ontology in response to a need
  - Assumed: Smb needs to **tell you** about the need
  - C.F.: **Kinematics** – studies the **motion** of objects without a reference to its causes
- Ontology Change:
  - In fact does Kinematics but misses the study of the causes that trigger the need
  - C.F.: **Dynamics** - is concerned with the study of **forces** and **torques** and their **effect on motion**

- **Ontology Change** misses **Ontology Dynamics**
  - What would be the force **causing** this going faster then 100 mph?

A Possible Way to Go …

- What if:
  - You take insights (so far) from … Newtonian **Mechanics**
    - E.g.: the Law of **Universal Gravitation** by Newton



http://goodpixgallery.com/

- Differences (simplifications / complications):
  - Knowledge representations are discrete in their nature:
    - Continuous models -> **discrete models**
  - Objects are immaterial:
    - Masses -> **Fitnesses**

- Desired output:
  - **OntoGrav**: A framework and tool to measure and visualize Ontologies in a Gravitation Field of a Domain of Discourse

Ermolayev, V.: Gravitation and Fitness in Ontology Dynamics. Inv. Talk at OEG, Universidad Politecnica de Madrid (ES), 11-Nov-2015, presentation slides (PDF)

A Syndicated Ontology of Time

(ST, Ontology Engineering)

Ermolayev, V., Batsakis, S., Keberle, N., Tatarintseva, O., Antoniou, G.: Ontologies of Time: Review and Trends. Int. J. of Computer Science & Applications. Vol. 11, Issue 3, 57–115, 2014 (PDF)

Why Important?

- **Has been** in the focus of scientific thought from ancient times –
  - e.g. **Plato**: the revolution of the celestial spheres

- **Continues to be** an important subject of research for philosophers, physicists, mathematicians, logicians, computer scientists, and even biologists

Geocentric celestial spheres; Peter Apian's Cosmographia (Antwerp, 1539)

- **One reason**: time is a <u>fundamental aspect</u> to understand and react to <u>change</u> in the World, including the broadest diversity of applications

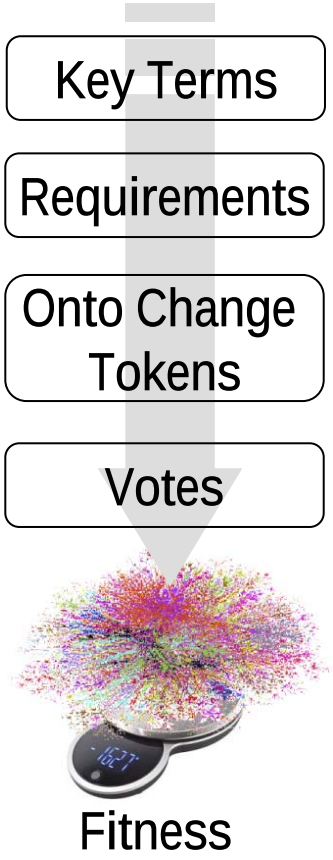
Methodology - OntoElect

- **OntoElect** *: understanding requirements as **votes** of the Domain Knowledge Stakeholders regarding the Ontology
  - Ontology fitness (**Φ**) is understood as proportional to the **ratio** of positive and negative **votes** of the Stakeholders
  - Votes collected **indirectly** – using a statistically representative Document Collection:

Requirements Elicitation
- Extract a **saturated** set of multi-word key terms
- Select the most **influential** key terms – Requirements

Conceptua-lization
- Transform the natural language definitions of the terms to formalized **structural contexts** – Ontology Change Tokens

Evaluation
- Map the structural contexts to the ontology – positive and negative Votes
- Compute the change in **Φ** – more or less positive Votes



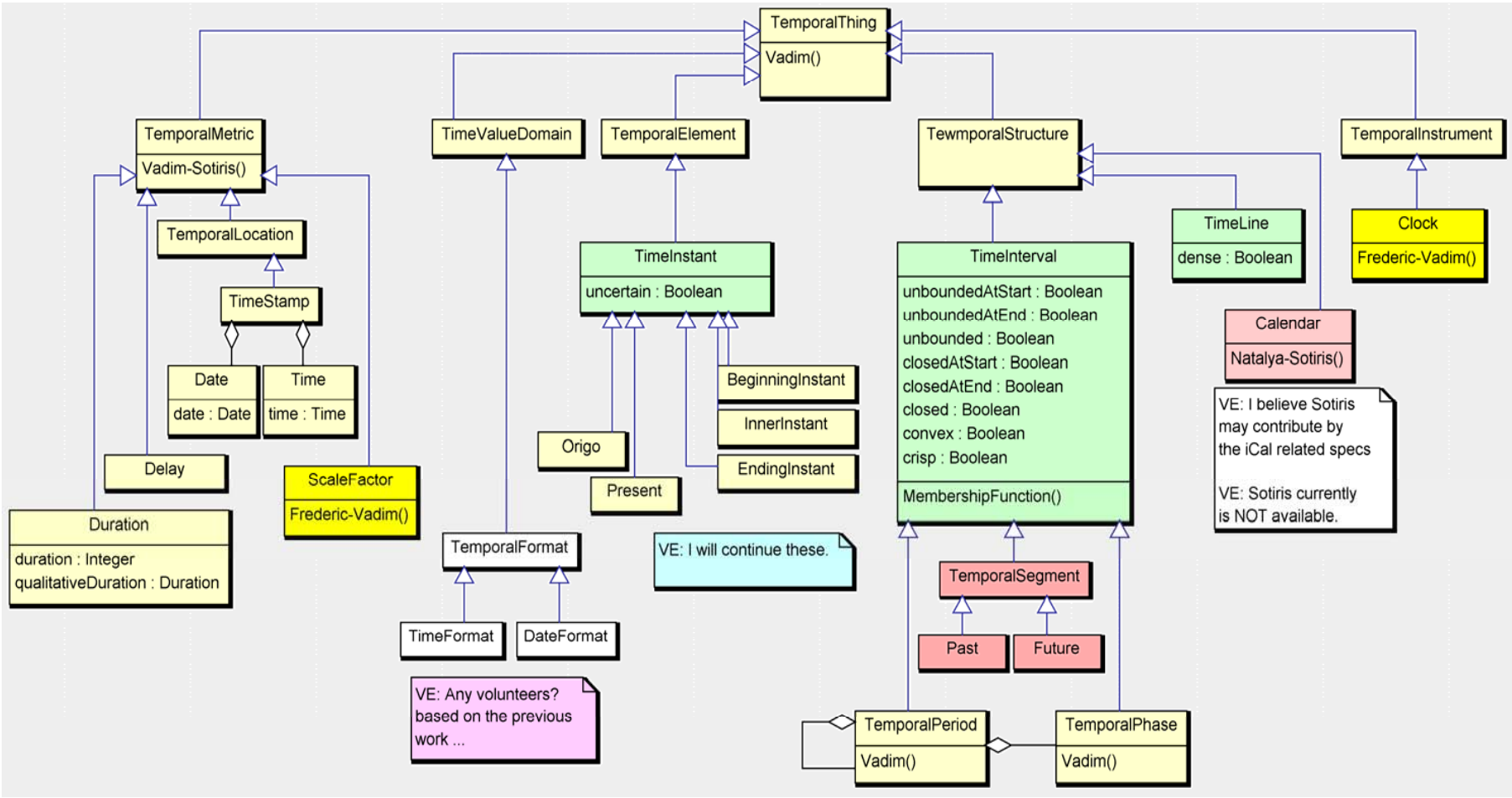
* Tatarintseva, O. et al. (2013) Quantifying Ontology Fitness in OntoElect Using Saturation- and Vote-Based Metrics. In: Ermolayev et al. (eds.) *ICT in Education, Research, and Industrial Applications. Revised Selected Papers of ICTERI 2013*, CCIS 412, pp. 136–162

Workflow

OntoElect **Conceptualization** Phase:

- Develop the **Backbone Taxonomy**
  - Based on the Requirements (**features**)
- Develop the **Seed**:
  - Focus on **Key Concepts** (Taxonomy)
    - E.g.: TimeLine, TimeInstant, TimeInterval, Clock
  - Develop/refine **theoretical descriptions**
    - Check if **implementable** using the available (W3C) languages
  - Harmonize – check **consistency**
  - Transform to Ontology (**Change Tokens**)
    - Visualize in a UML Class Diagram
    - Produce a **W3C compliant** code (OWL 2 DL + SWRL)
  - Document (**SOT-Wiki**)
  - Evaluate against required features (OntoElect: **Fitness,** Evaluation phase)
- **Expand**
  - Add concepts (Taxonomy)
  - Repeat the cycle until:
    - All the requirements are met (OntoElect: **Fitness,** Evaluation phase)
      OR
    - The limits of expressive power are reached (W3C **compliance**)

Backbone Taxonomy



Ermolayev, V.: Toward a Syndicated Ontology of Time for the Semantic Web. AIFB Oberseminar, Karlsruhe Institute of Technology (DE), 02-Feb-2016, presentation slides (PDF)

A Collab. Effort (SemData+)

- **Vadim Ermolayev** (ZNU) – req. analysis, temporal theories, SOT theory, key concept models, SOT-wiki, PSI-ULO, PSI-Time
- **Sotiris Batsakis** (HUD) - temporal reasoning frameworks, ontologies, TimeInstant, TimeInterval models, SOWL
- **Frederic Mallet** (UN-SA) – SOT theory, Clock model, OMG Clock
- **Natalya Keberle** (ZNU) - temporal reasoning frameworks, OWL+SWRL compliance check, PSI-Time
- **Olga Tatarintseva** (ZNU) OntoElect, req. elicitation, PSI-ULO





Oh, btw ...

- This is not **necessarily** about Temporal...
- There are many **OTHER** Domains where the approach of **OntoElect** could be used ...
- One possible:
  - **Springer Publications** – in discussion with Springer Verlag, Heidelberg

Evolving Knowledge Ecosystems

(Big Data and Analytics)

Ermolayev, V., Akerkar, R., Terziyan, V., Cochez, M.: Toward Evolving Knowledge Ecosystems for Big Data Understanding. In: Akerkar, R. (ed.) Big Data Computing, pp. 3--56, Taylor & Francis, 2013, ISBN 978-1-46-657837-1

BIG is a Problem …

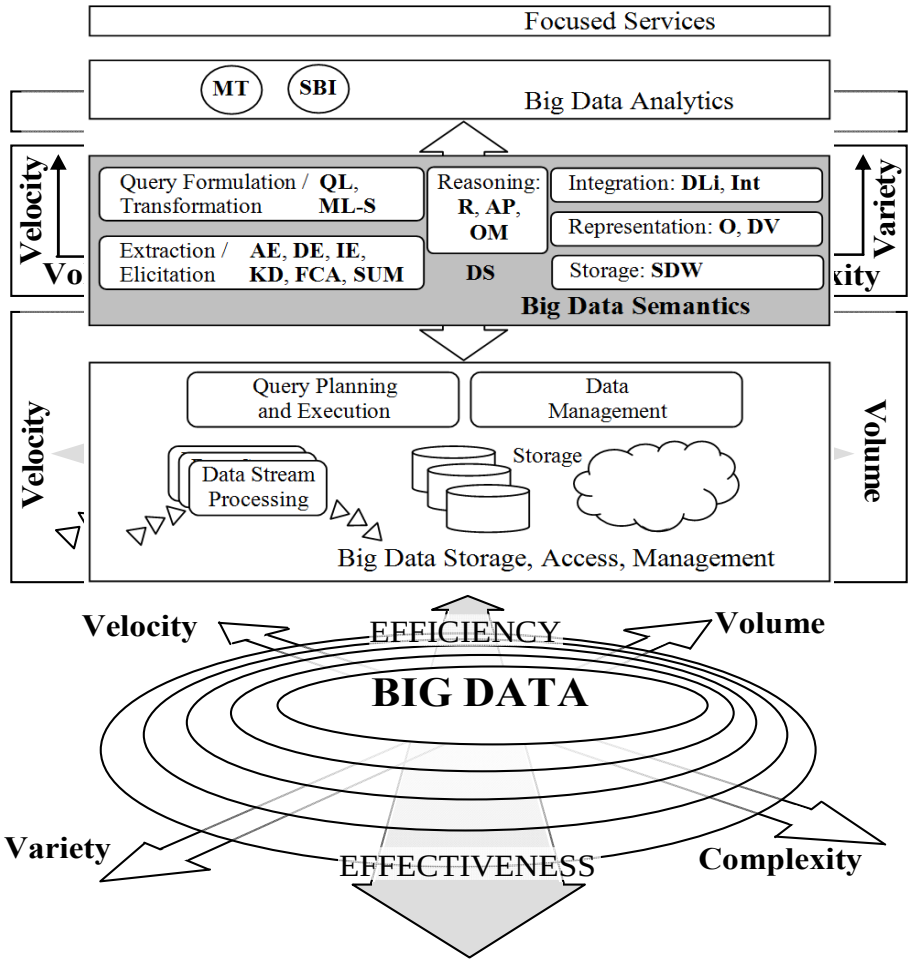
Courtesy of Vladimir Ermolayev

- Mined correlations (for Big Data analytics), though very useful, may hint about an answer to a “what” but not “why” kind of questions

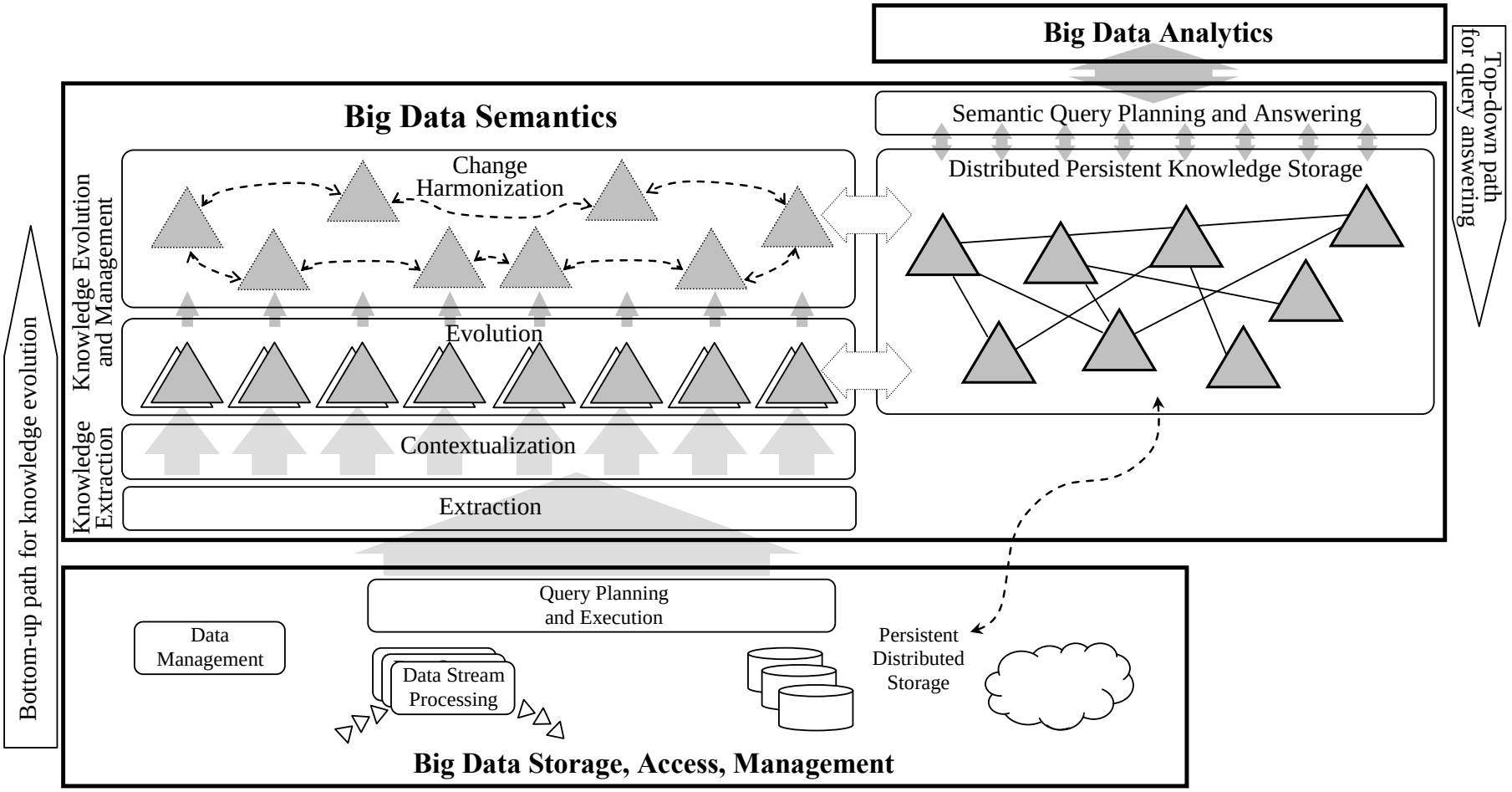
◇ For example, if Big Data about Royal guards and their habits had been collected in the 1700s France **one could mine today that all musketeers who used to have red Burgundy regularly for dinners have not survived till now**. Pity, red Burgundy was only one of many and a very minor problem.

Efficiency vs Effectiveness

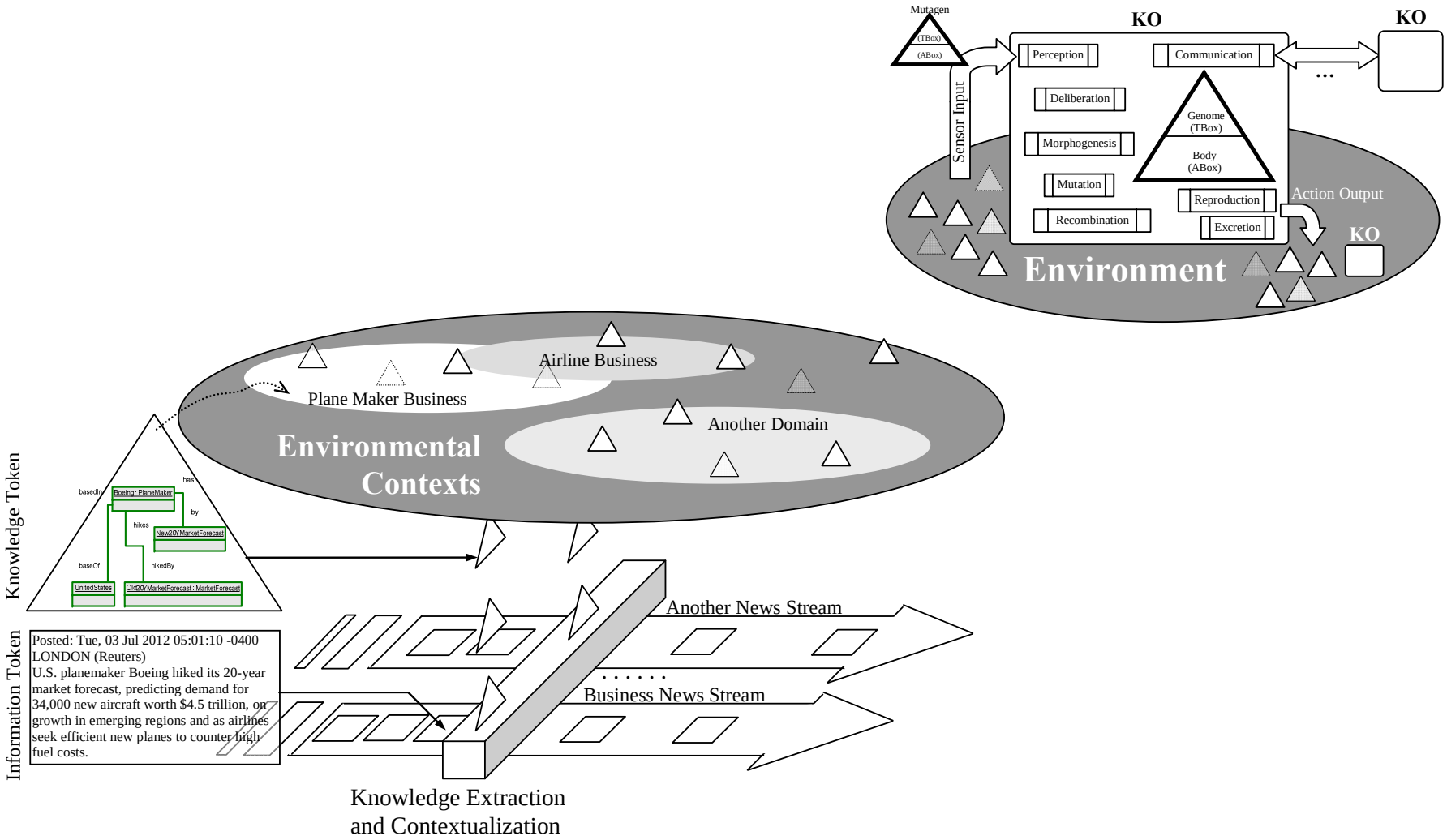
- A major conceptual complication for analytics is that efficiency is anti-correlated to effectiveness
- Should be balanced
- A **Big Data Semantics** layer needs to be introduced in the Processing Stack



Bottom-Up + Top-Down



Knowledge Ecosystems



Skills … also to Develop …

- Ontology Dynamics
  - Math, Java, UML, OWL, relevant tools, knowledge extraction from texts, NLP
- A Syndicated Ontology of Time
  - Math., Description Logics, OWL, SWRL, Java, UML, Graph processing
- Evolving Knowledge Ecosystems
  - Math., Agents, Java, JADE, OWL, UML, Graph models, Text stream processing

- Communication: good command of English

Will be happy to answer your questions ...

Will be also happy to continue discussions

vadim@ermolayev.com