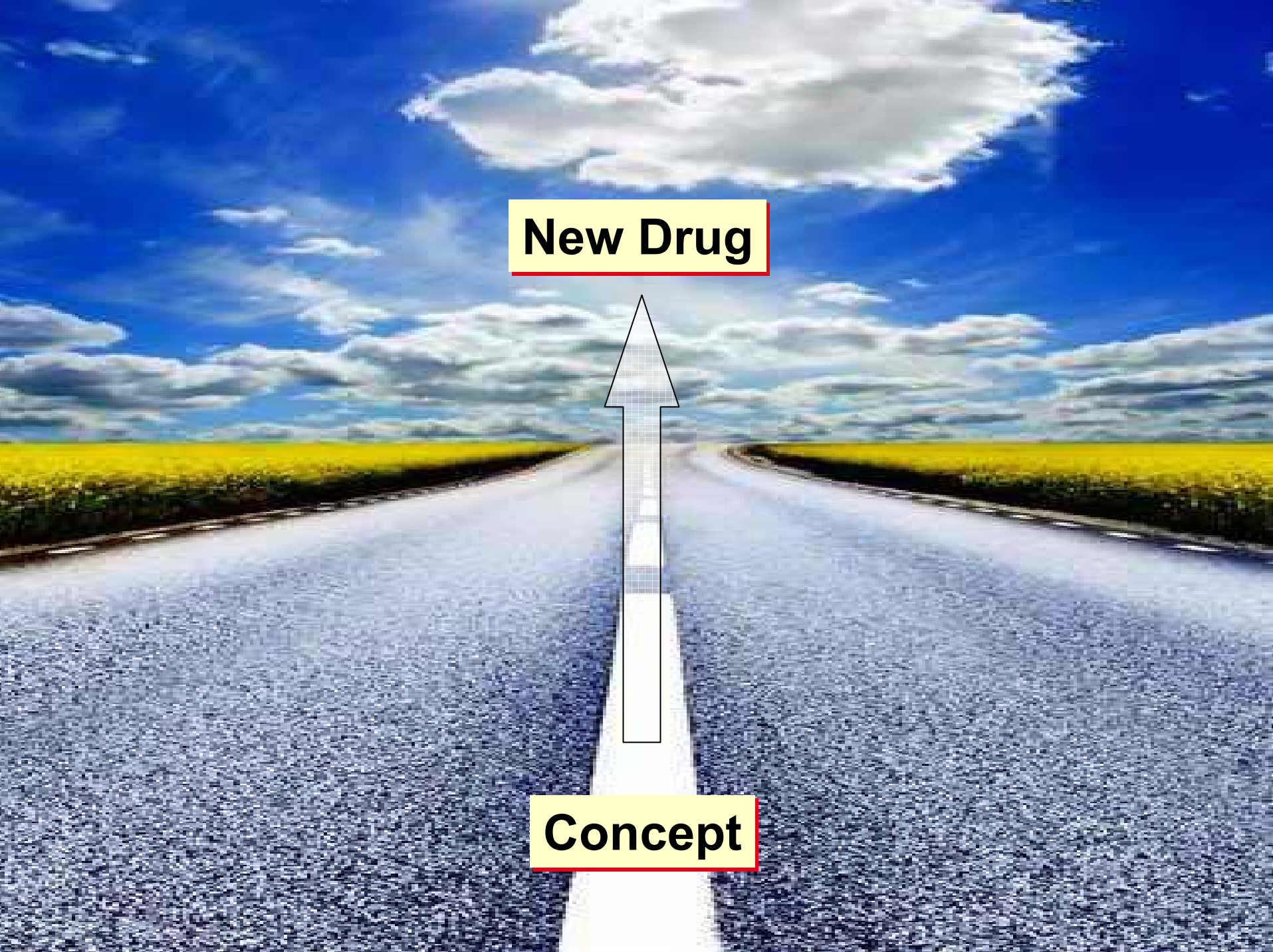
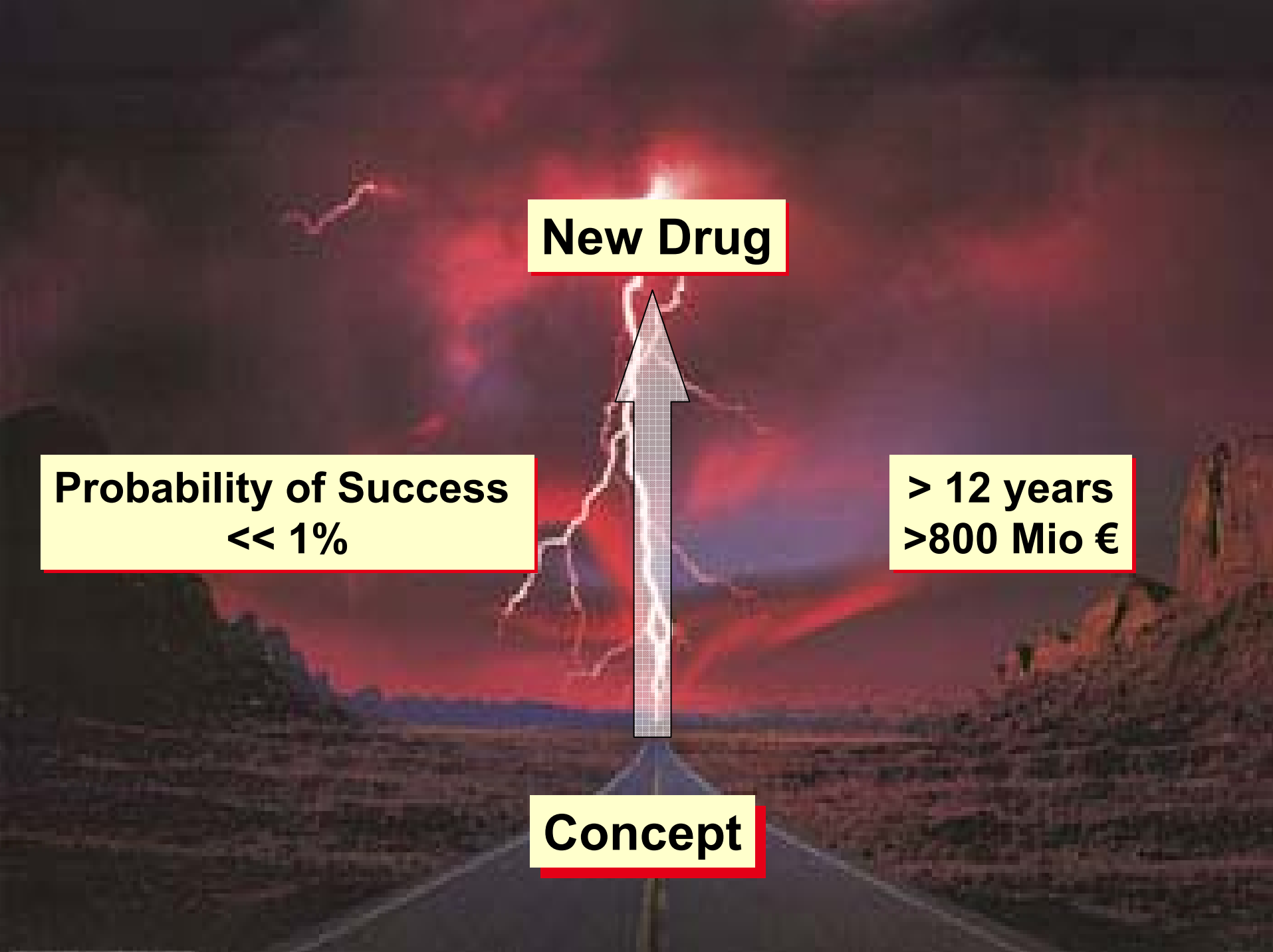
Automated knowledge discovery to support the development of new drugs

Caroline Kant
Operational Excellence Research
Research Informatics & Knowledge Management
Merck Serono

Agenda

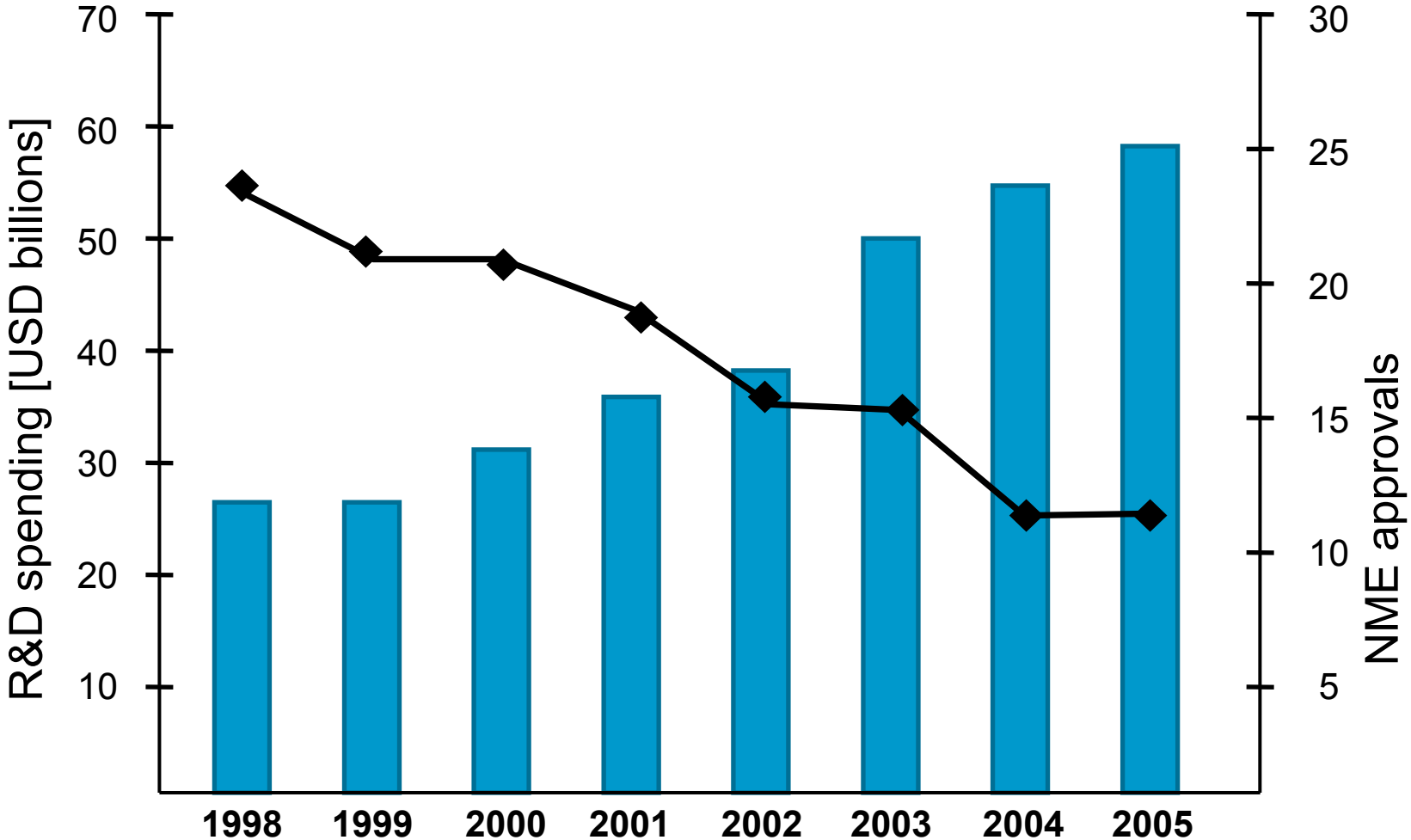
- The Pharma productivity challenge
- Automated Knowledge Discovery (AKD)
- AKD supports the whole pipeline
  - 1st case study: predicting safety issues
  - 2nd case study : supporting innovation
- Where we are heading
- Conclusions


New Drug
Concept


New Drug
Probability of Success
<< 1%
> 12 years
>800 Mio €
Concept

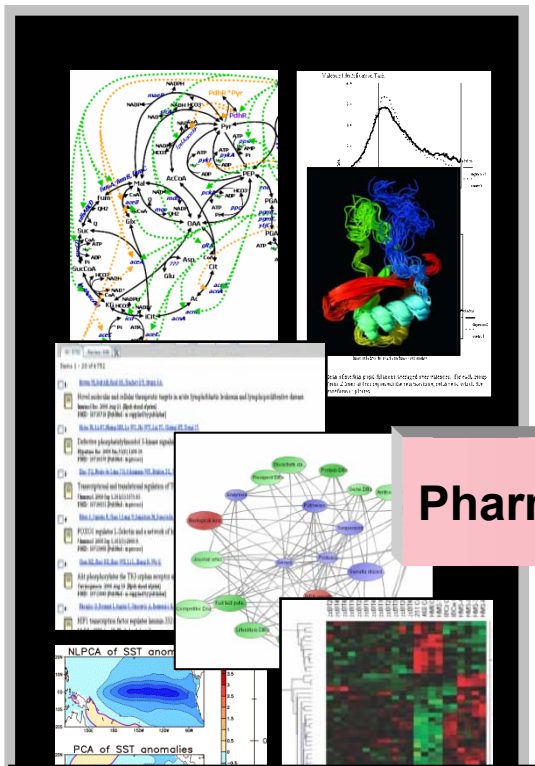
The Pharma Innovation Gap

Research Productivity is falling



PhRMA Pharmaceutical Industry Profile, 2006 – U.S. FDA Center for Drug Evaluation and Research

The Ultimate Challenge
Transforming Data into Knowledge



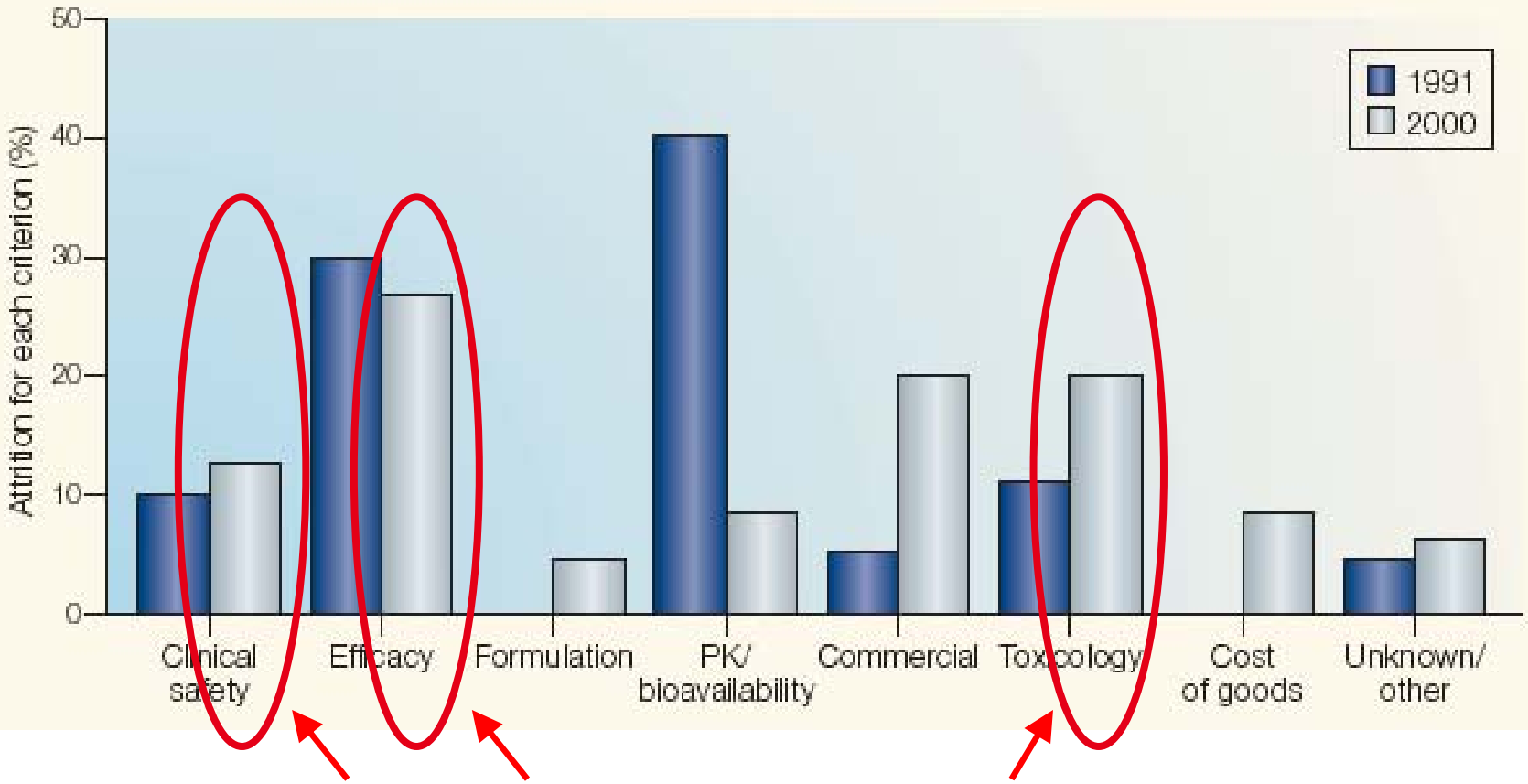
Pharma invest heavily into data

« If only we knew what we know we would be a company that is three time more profitable »

Lew Platt, former CEO of Hewlett-Packard

Why drugs fail?

Kola I & Landis J (2004) Nature Reviews Drug Discovery; 3, 711-714



Translational science challenge

Automated **K**nowledge **D**iscovery
Tools and Strategies

An *in silico* approach leveraging on a collection of techniques and strategies to identify and exploit knowledge on large amounts of heterogeneous data/text

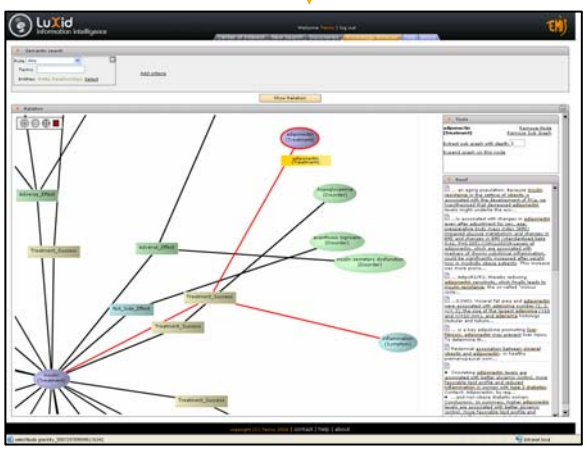
- Text Mining
- Visual Text Analytics
- Semantic search
- Pathway analysis
- Semantic mapping
- Literature Base Discovery
- **Combination** of tools and strategies

Knowledge Mining
Supporting the pipeline

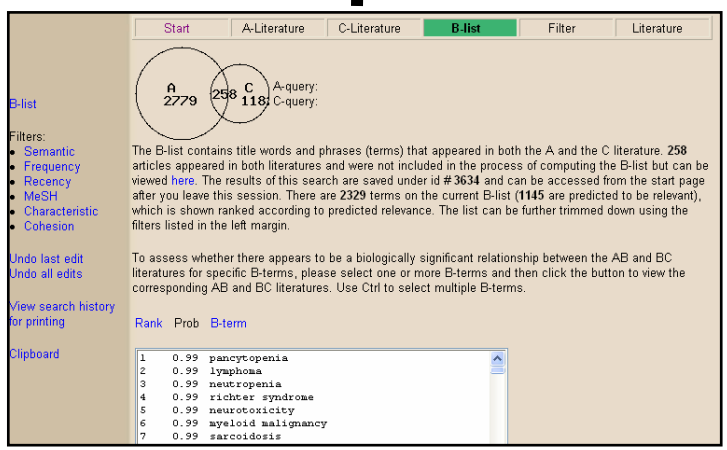
- Mechanism of Action
  - Which molecular mechanism shall be perturbed in order to cure the disease X?
  - Who is the best indication for drug Y?
- SAFETY
  - What is the rationale for the adverse event X associated with the administration of Y ?
  - What „adverse events“ observed with drugs in the same class should be considered „risk“ for compound x ?
- BIOMARKERS
  - Which biochemical feature could be used to identify a disease and monitor the effect of treatment Y for disease X ?

Integration of Visual analytics and Literature Base Discovery **–Case Study #1**

What „**adverse events**“ observed with drugs in the same class should be considered „**risk**“ for compound x ?



Visual Analytics



Literature Base Discovery

Complementary techniques, logic knowledge extraction workflow

Visual analytics using LUXID – State-of-the-art Text Mining

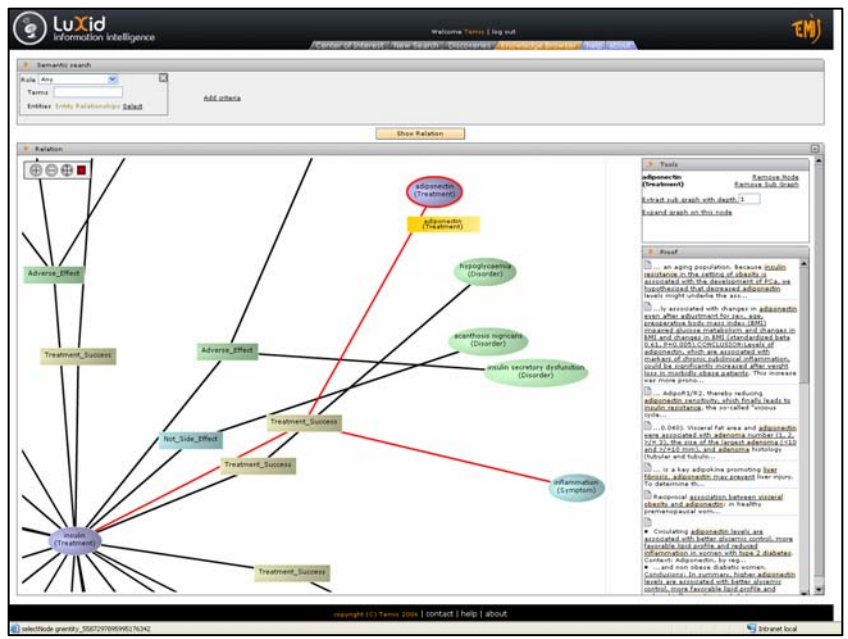
- Search delivered 2901 Documents
- Processed automatically in 9 h on one Central Processing Unit (CPU)

Microsoft Excel - codeine_adverse_effects.xls

E17: Codeine can cause dependence and may induce psychosis

	D	E
14		codeine 60 mg was associated with any significant increase in any patient experiencing an adverse event
15	patient (Species)	codeine 60 mg was associated with any significant increase in any patient experiencing an adverse event
	DT (Guessed Treatment)human (Species)effect Cyclooxygenase inhibitor (Drug Related Treatment)dosage Double-Blind method drug (Treatment)drug therapy (Drug Related Treatment)effect tooth extraction (Treatment)76-67-3 (Guessed Treatment)dosage Anti-Inflammatory agent (Treatment)dosage codeine (Treatment)Non-Steroidal (Guessed Treatment)surgery (Non Drug	adverse effects Administration, Oral *Analgesics: AD, administration & dosage Anti-Inflammatory Agents, Non-Steroidal: AD, administration & dosage Codeine: AD, administration & dosage Codeine: AE, adverse effects Cyclooxygenase Inhibitors: AD, administration & dosage Double-Blind Method Drug Combinations Humans *Molar, Third: SU, surgery *Pain, Postoperative: DT, drug therapy Randomized Controlled Trials Review Literature *Tooth Extraction: AE, adverse effects Tooth Extraction: SN, statistics & numerical data
16	Related Treatment)	103-90-2 (Acetaminophen); 76-67-3 (Codeine)
17	Psychosis (MeSH Nomenclature Disorder)	Codeine can cause dependence and may induce psychosis
18	tramadol (Pharmacological Substance) Vomit (MeSH Nomenclature Symptom)	adverse events for tramadol/APAP and codeine/APAP was found, except for constipation (0% vs 10.9%) and vomiting
19	APAP (Guessed Treatment) Constipation (MeSH Nomenclature Symptom)	adverse events for tramadol/APAP and codeine/APAP was found, except for constipation (0% vs 10.9%) and vomiting
20		poisoning Codeine
21		Female Adult Aspirin: PO, poisoning Codeine: PO, poisoning Humans Iatrogenic Disease *Intubation, Intratracheal: AE, adverse effects
22		poisoning Codeine
23	Adverse Effect Poison (MeSH Nomenclature Disorder) Adverse Effect	Female *Acidosis, Renal Tubular: CI, chemically induced Adult *Anti-Inflammatory Agents, Non-Steroidal: PO, poisoning Codeine: PO, poisoning Humans *Hypokalemia: CI, chemically induced *Ibuprofen: PO, poisoning
24		Pancreatitis is a rare adverse effect of codeine
25		Pancreatitis due to codeine
26		Pancreatitis is a rare adverse effect of codeine
27	Pancreatitis (MeSH Nomenclature Disorder)	Pancreatitis due to codeine
28		codeine/acetaminophen with fewer gastrointestinal and nervous system adverse events
29	acetaminophen (Pharmacological Substance)	codeine/acetaminophen with fewer gastrointestinal and nervous system adverse events
		codeine/acetaminophen group had more patients with 1 or more

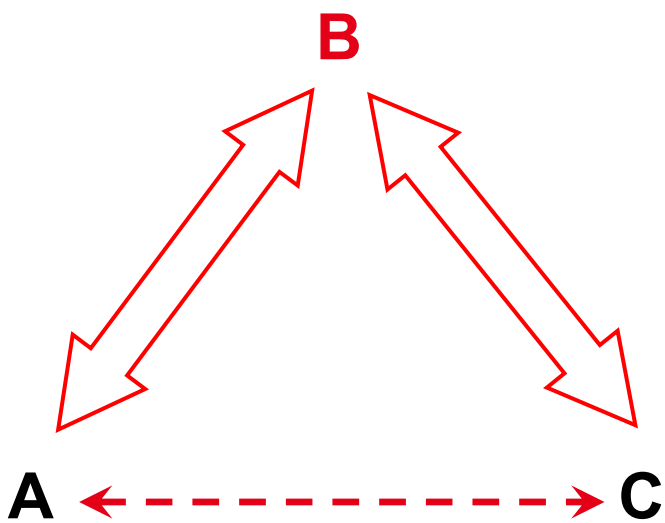
Relation__Adverse_Effect___with



Clustering of knowledge in Luxid

Literature Based Discovery
Swanson‘s ABC method

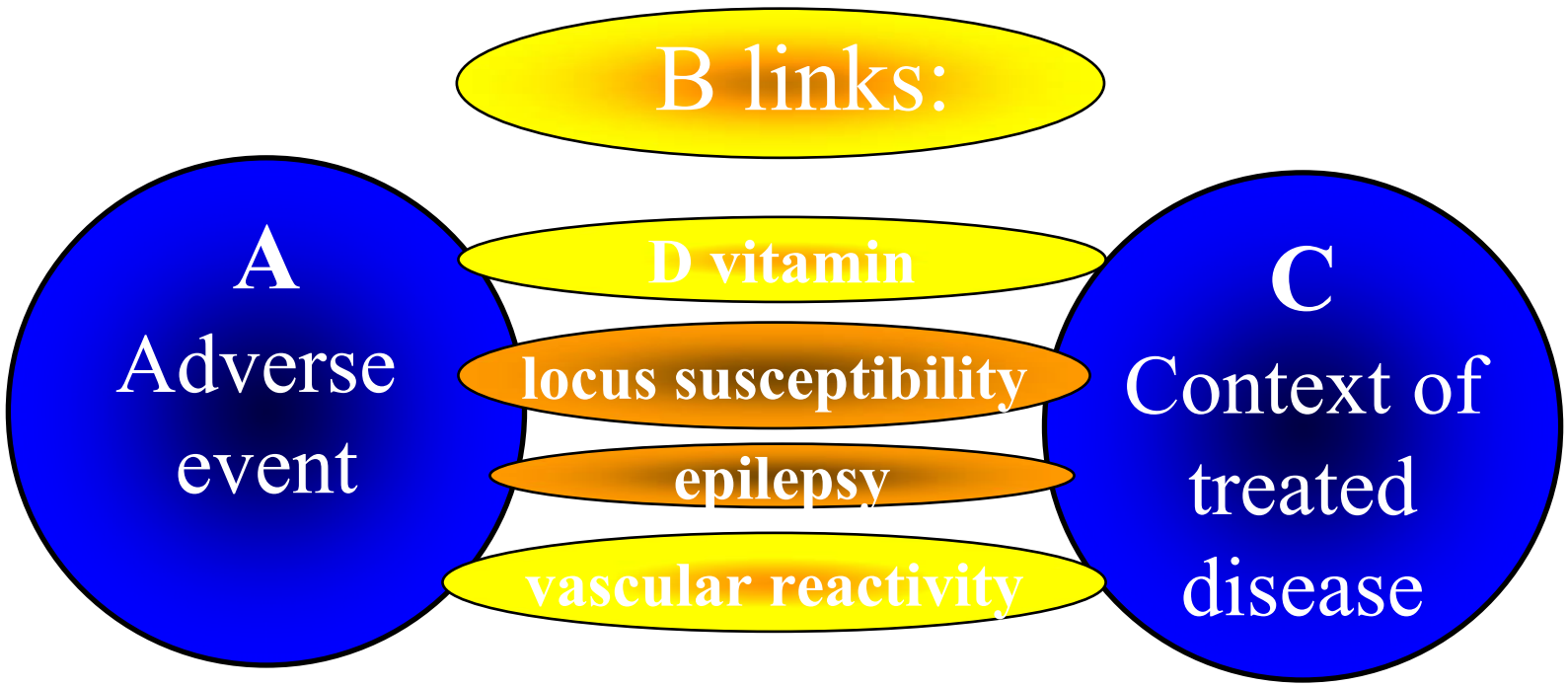
ABC method: An analyze base on statistical Natural Language Processing pioneered by Dr. Don Swanson and colleagues



“ The relationships AB and BC are known, the implicate AC relationship is a putative discovery” Marc Weeber, 2001

Literature Based Discovery: „B“
Tool: Arrowsmith

- What is linking A with C ?



AKD supports
Toxicology and Safety predictions

What „**adverse events**“ observed with drugs in the same class should be considered „**risk**“ for compound x ?

- Continuous interaction with the project teams
- Design experimental assays to test potential risks
- Early implementation of a risk management plan for the projects
- Follow toxicology and safety concerns all along the discovery and development process
- Better understand the biological context of the indication
- Support an informed design of the clinical trial protocol

Integration LBD and Visual analytics
Benefits and shortcomings

Visual Analytics

+ + + Fast output

+ + Integration int / ext data

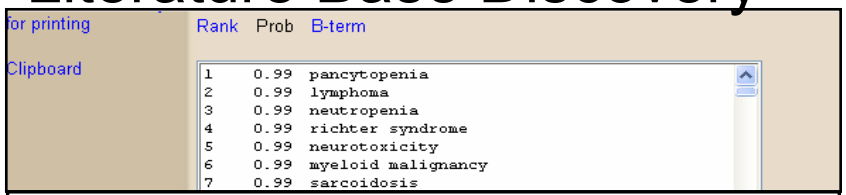
+ + Identification of patterns

- - - Incomplete picture

- - - Precision of recall

- - - Data hard to interpret

Literature Base Discovery



- - - Tedious / time consuming

- - - Only external data considered

+ + Wide converge of external data

+ + + Implicit hypothesis generation

complementary

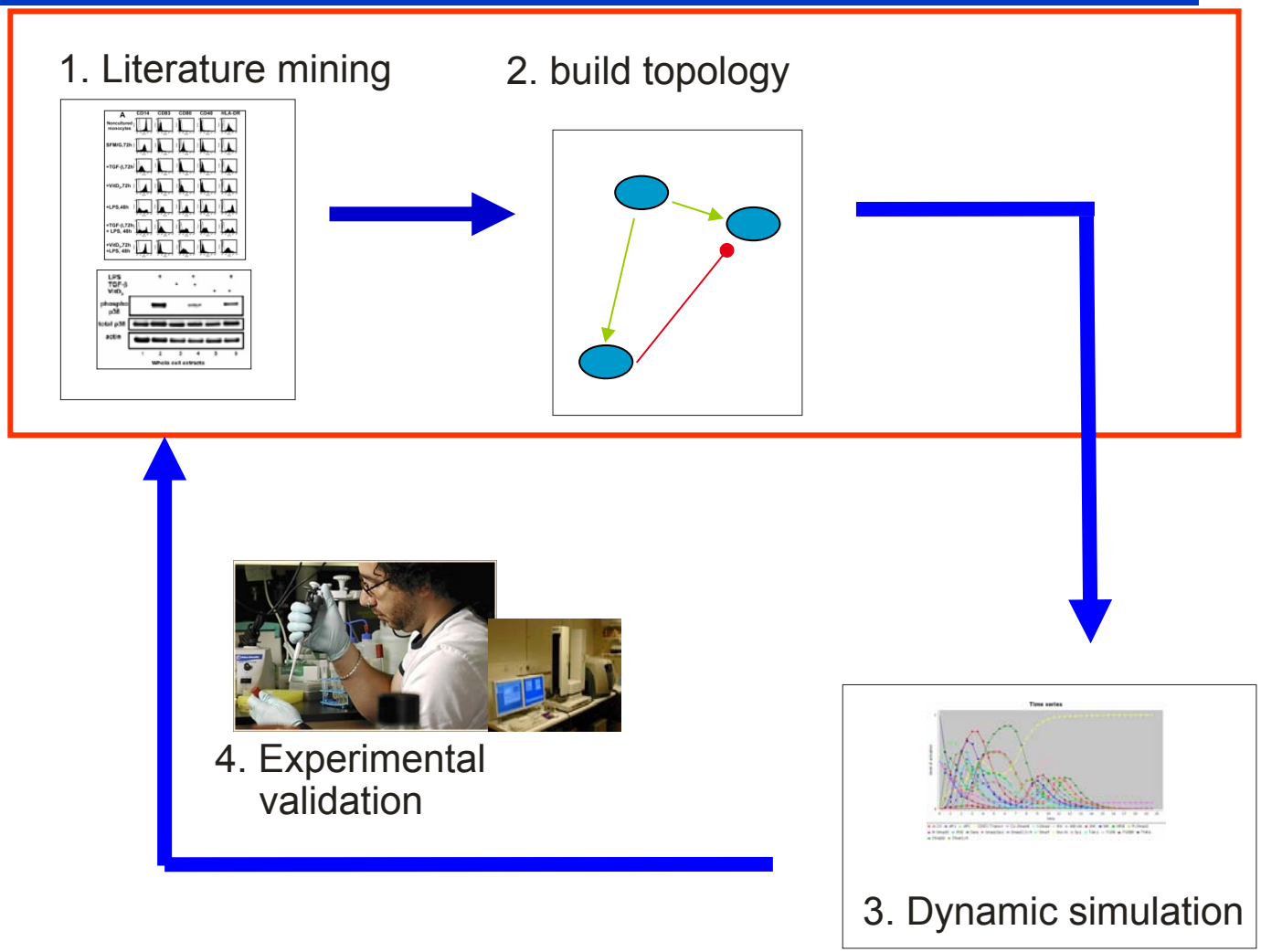
Biology modeling and simulation
Supporting innovation – **Case Study #2**

Quantitative modeling:

By perturbing the network with the over-expression or the knockout of components we can study how a network behaves and the role of its components.

- Understand the disease mechanism
- Identify and validate target
- Identify alternative pathways
- Predict efficacy
- Potential biomarker identification

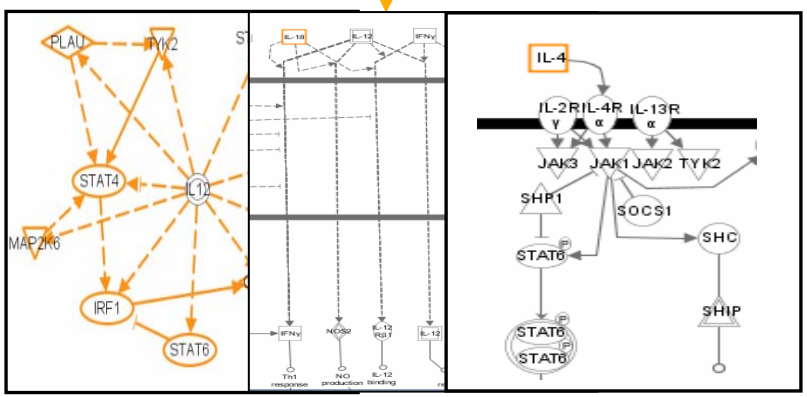
AKD to build a network topology for Modeling & Simulation – **Case Study #2**

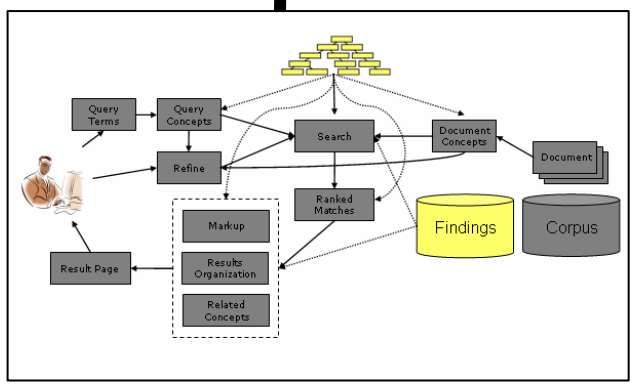


Integration PAT and Semantic search

Building the topology of molecular networks

Which „**molecular mechanism**“ should be „perturbed“ in order to cure the **disease X**?





Pathway analysis tools (PAT)

Semantic search

Complementary techniques, logic knowledge extraction workflow

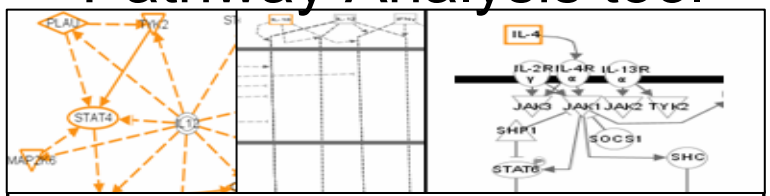
AKD supports
MOA and efficacy predictions

Which „**molecular mechanism**“ should be „perturbed“ in order to cure the **disease X**?

- Continuous interaction with the project teams
- Support the validation of the project’s biological rational
- Help an informed design of the project
- Design experiments to predict efficacy and simulate the compound behavior in humans
- Capture de complexity of the human disease context

Integration PAT and Semantic Search
Benefits and shortcomings

Pathway Analysis tool



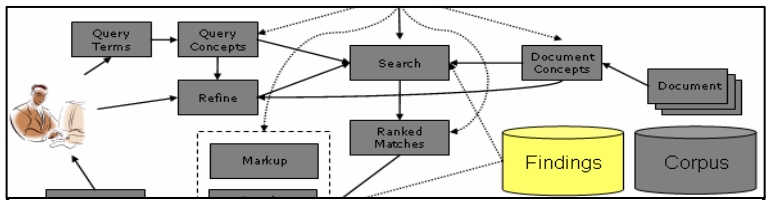
+ + + Quick access to a manually curated pathways

+ + Integration of pathways to build the network framework

- - - Incomplete picture

- - - Misleading annotation

Semantic Search



- - - Time consuming

+ + + Focused search

+ + + Enrich the network with wide coverage of data

+ + + Validation of the network

complementary

How to describe the best system?
Sustained knowledge acquisition

Genomics Data

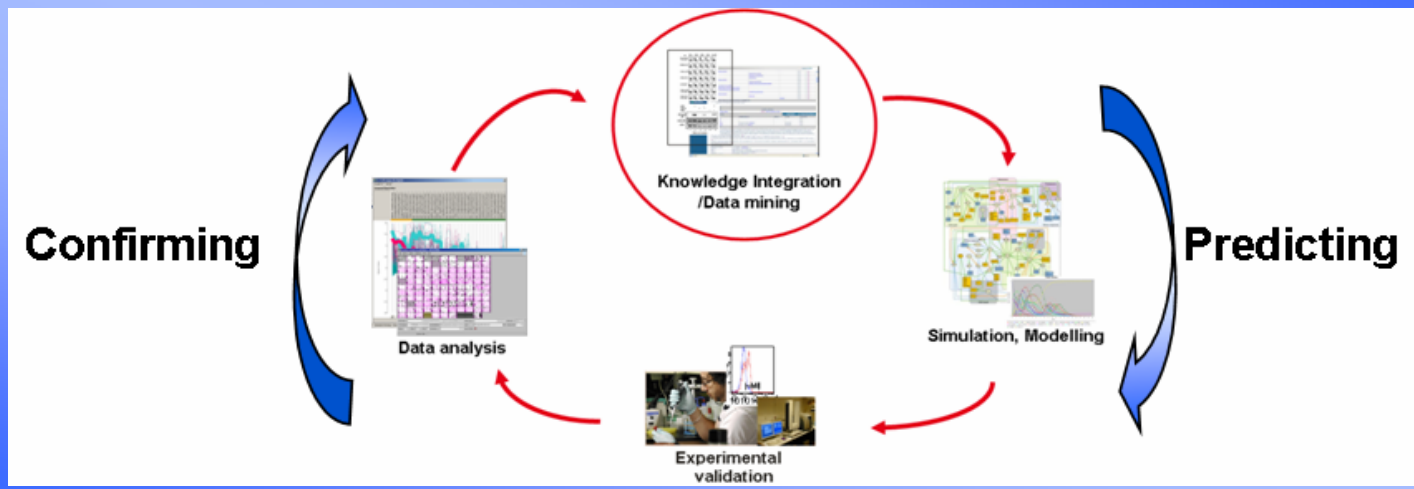
Chemistry & SAR, PK/PD

INTEGRATION

Clinical data

Extraction & analysis of external data

Knowledge acquisition workflow



Better decision-making/ increase confidence in the drug

Conclusions

- AKD is a methodology which uses several techniques and strategies in hands of expert how helps predict a safety as well as efficacy issue and supports innovation

- AKD Assists, but does not replace human experts

- The final goal is to implement an integrated Knowledge Management System which empowers scientist to continuously convert data into knowledge

Acknowledgment

- Merck Serono – Research Informatics & Knowledge Management group

  - **Alexander Scheer**
  - **Luca Toldo**
  - Claudia Schepers
  - Grayce Shomade
  - Natalia Novac
  - Nicole Gullu

Thank you for your attention!