IBM Healthcare & Life Sciences

Text Analytics and Chemical Annotation of Patents and Biomedical Literature in Biomarker-enabled Drug Discovery

International Chemical Information Conference (ICIC)
Sitges, Spain, 21-24 October 2007
Michael Hehenberger, hehenbem@us.ibm.com

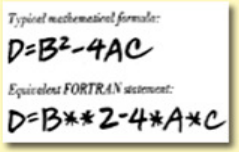
IBM Research

62 years of innovation

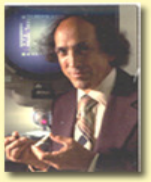
1944: Mark 1

1948: SSEC

1956: RAMAC



1957: FORTRAN



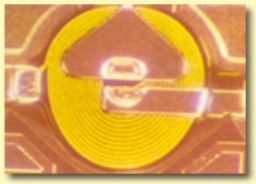
1966: One-Device Memory Cell

1967: Fractals

1970: Relational Database

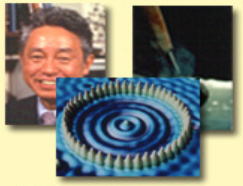
1971: Speech Recognition

1973: Winchester Disk



1979: Thin Film Recording Heads

1980: RISC

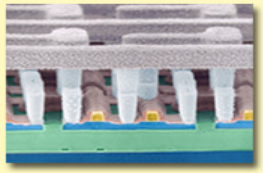


Nobel Prizes

1994: SiGe

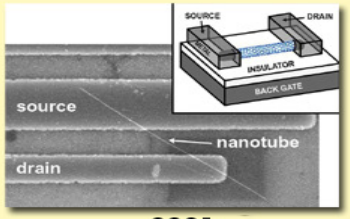
1993: RS/6000 SP
1996,97: Deep Blue

1997: Copper Interconnect Wiring

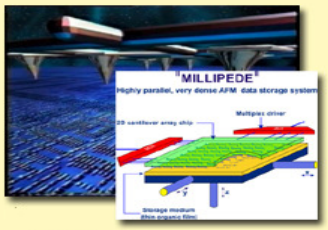


1998: Silicon-on-Insulator

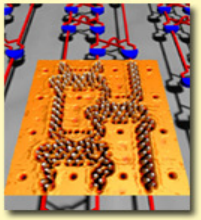
1998: Microdrive



2001: Nanotube Transistor



2002: Millipede



2002: Molecule Cascade Logic Circuit

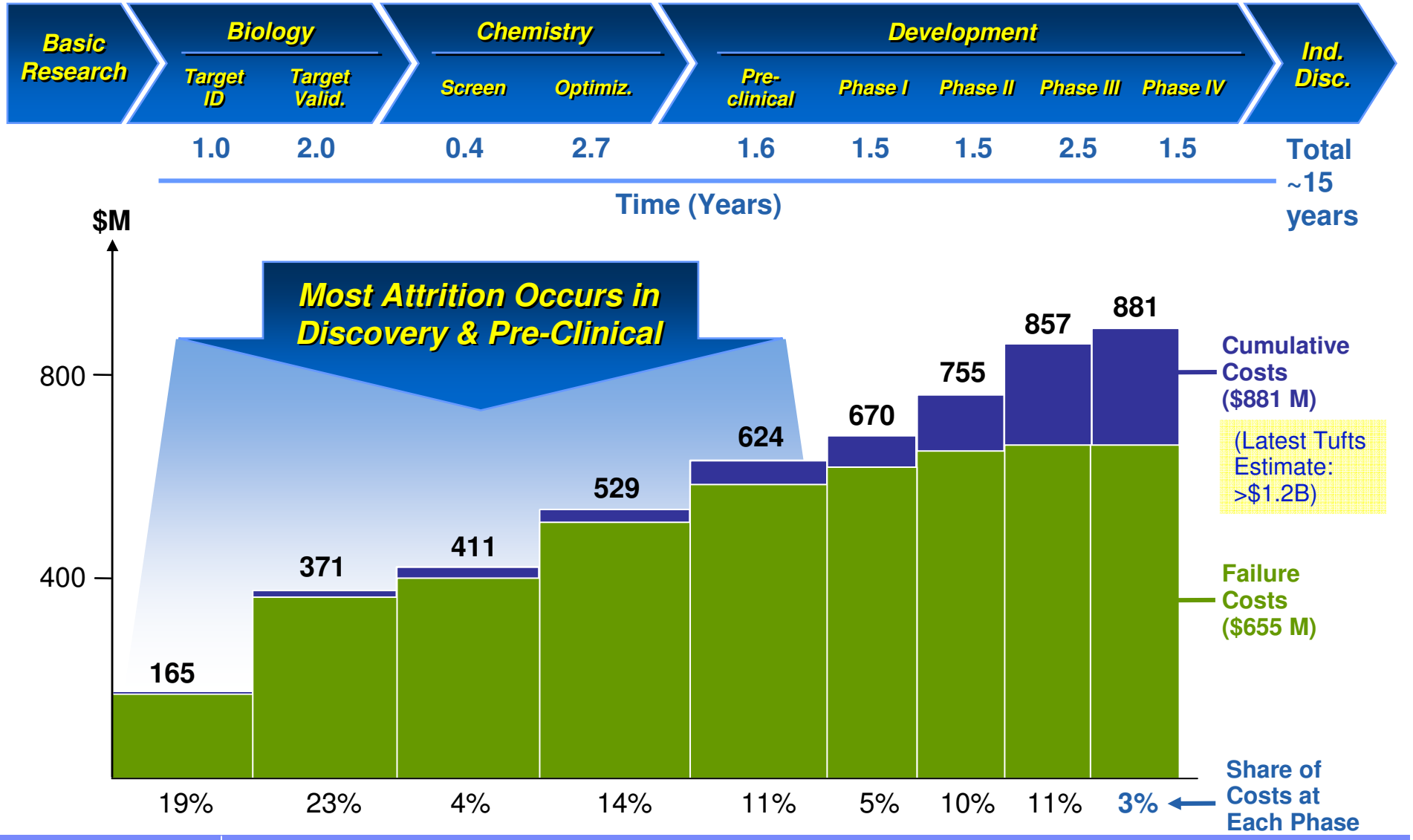
2004: Blue Gene/L
The fastest supercomputer in the world

Outline

- **Pharma R&D Productivity and Pipeline**
- **Biomarkers**
- **Role of Unstructured Information**
- **Unstructured Information Mining (UIM)**
- **Chemical Names Annotation**
- **Brief DEMO**
- **Blue Gene Supercomputing**
- ***in silico* Drug Discovery**

The current Pharma "Blockbuster" Model cannot be sustained:

Failure Costs Account for 75% of Drug R&D Costs: $881M / NME in 2003



Source: Tufts, Pfizer

Patent Expirations are threatening Bio-Pharma Industry ...

Company	2010		2011		2012		Share of Revenues (%)
AstraZeneca	Arimidex	($2.2bn)*	Seroquel	($4.7bn)	Symbicort	($3.7bn)	38**
BMS			US Plavix Avapro	($4.8bn) ($1.3bn)	Abilify	($2.1bn)	30
GSK	Advair	($3.8bn)			Avandia	($2.5bn)	23
Eli Lilly			Zyprexa	($4.8bn)			22
Merck	Cozaar/ Hyzaar	($3.2bn)			Singulair	($4.5bn)	22
Novartis	Femara	($1.1bn)			Diovan	($6.0bn)	14
Pfizer	Aricept	($800m)	Lipitor Xalatan	($12.1bn) ($1.6bn)	Viagra Detrol Geodon	($1.7bn) ($860m) ($1.1bn)	41
sanofi-aventis	Taxotere	($2bn)	US Plavix Avapro	($3.8bn) ($2.1bn)	Lovenox	($3.1bn)	34

Source: AXA Framlington
Notes: * Estimate of global sales in 12 months prior to patent signing
** Value of products losing patent protection as a percentage of total company sales over next five years

Source: PwC Pharma2020

Increased Bio-Pharmaceutical R&D Productivity will require joint action by Regulators, Industry & Academia / Public Sector

Ongoing efforts focused on improving R&D productivity:

NIH Roadmap

FDA Critical Path

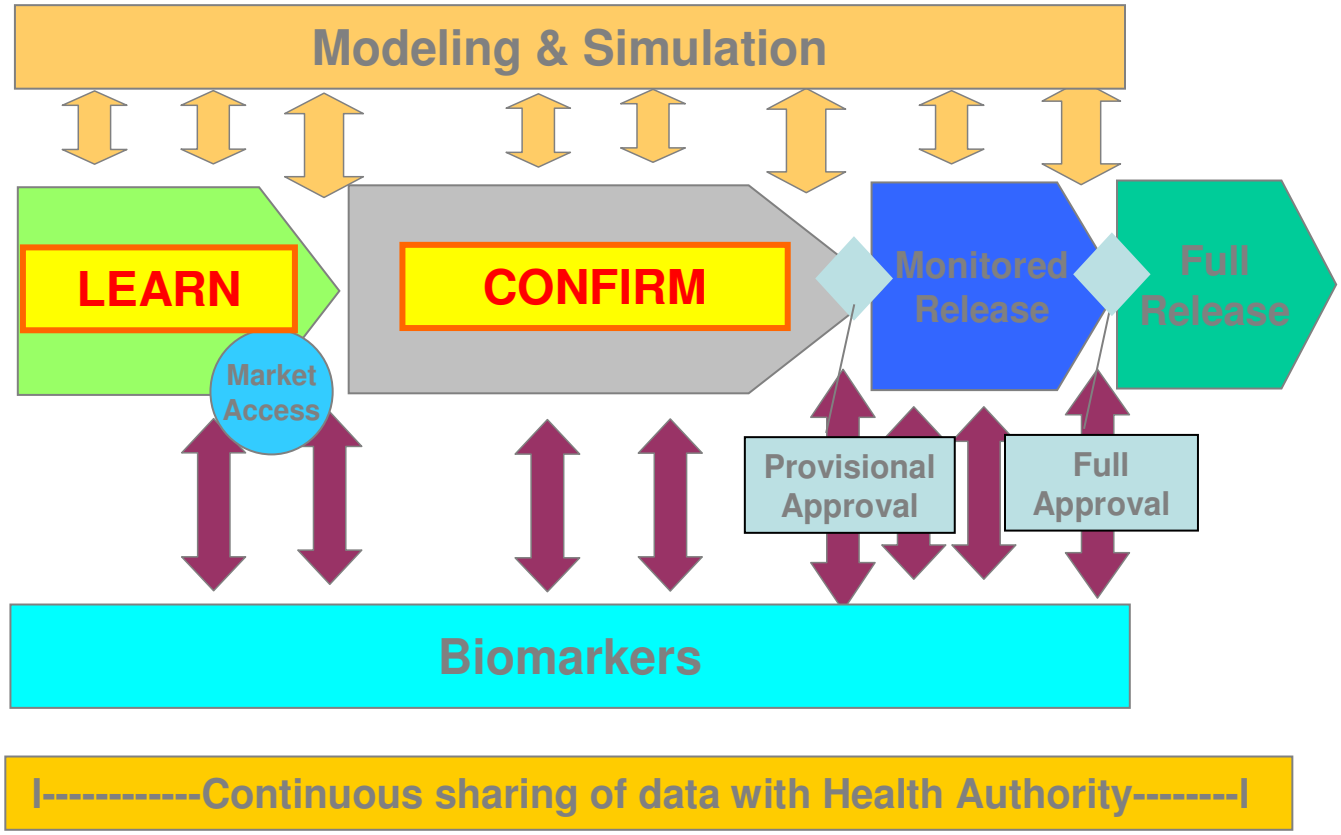
Pharmaceutical Innovation Steering Committee (PISC)

- **Biomarkers**
- **Novel adaptive trial design**
- **Accelerating proof of concept**
- **Enriched patient population trial designs**
- **Rolling dose studies**
- **Exploratory IND**
- **Improving efficiency of late-stage clinical research**
- **SAE (Serious Adverse Events) data-mining validation**
- **Best regulatory practices and sponsor/regulator communication**
- **Predictive models for safety and efficacy**

Source: Pfizer

Key Drivers to transform the Pharma R&D Model

- **Biomarkers**
- **Modeling and Simulation**
- Rapid compound selection in man
- Innovative clinical trial design
- Innovative approaches to registration
- Integrated safety assessment & risk management
- Quality manufacturing



Source : Biomarker Summit III, Dr. Werner Kroll, 26 January 2007

Source: Novartis

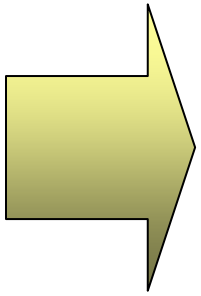
What is a Biomarker and why are Biomarkers important?

- What is a Biomarker?
  - "Objectively measured indicator of biological/ pathobiological process or pharmacologic response to treatment".
    - A substance or measurement whose detection indicates a particular disease state
    - Usually, a "biomarker" indicates a change in expression or the state of a protein that correlates with the risk or progression of a disease, or with the susceptibility of the disease to a given treatment
- What is a "Clinical Endpoint"?
  - "A Characteristic or variable that reflects patient feeling, function or survival"
  - Biomarkers can be "***Surrogate endpoints***":
    - Biomarker intended to substitute for a clinical endpoint (predict benefit or harm) based on epidemiologic, therapeutic, pathophysiologic or other scientific evidence
- Biomarker Examples: **Cholesterol, Blood pressure levels for heart disease, PSA (antigen) for prostate cancer, HbA1c in diabetes, CD4 lymphocyte count for AIDS**
- How are they used in drug development?
  - Once a proposed biomarker has been validated, it can be used to diagnose disease risk, presence of disease in an individual, or to tailor treatments for the disease in an individual
  - If a treatment alters the biomarker, which has a direct connection to improved health, the biomarker serves as a "surrogate endpoint" for evaluating clinical benefit.

Biomarkers can improve clinical research and patient care

Biomarker varieties:

- Genetics
- Genomics
- Other "- omics"
- Molecular Markers in Blood Serum
- Analysis of Cell Populations
- Physiological Properties
- Structural, Functional & Molecular Imaging



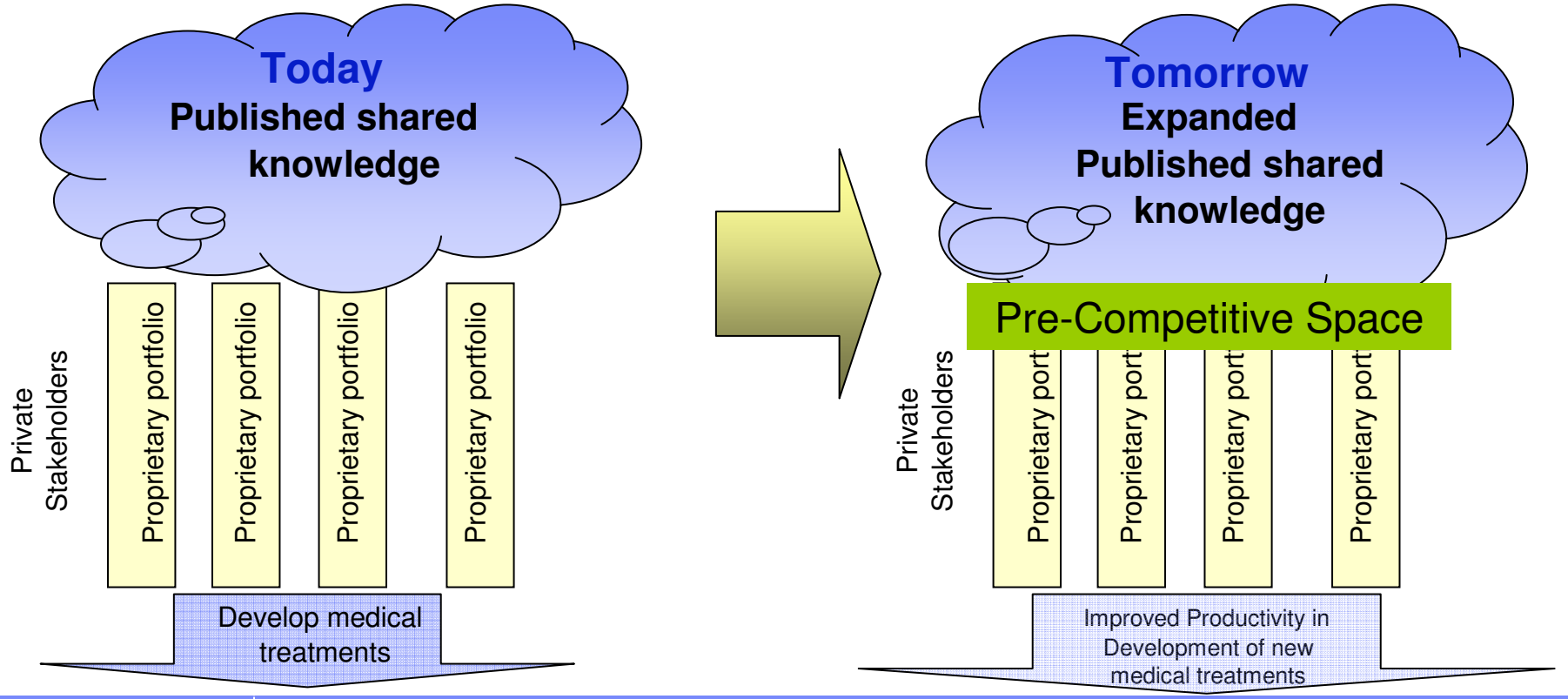
Improve Clinical Research & Patient Care via:

- Diagnostics
- Tracking of Biological Effect
- Risk assessment
- Patient selection based on safety factors
- Patient stratification based on response factors
- Monitoring of treatment
- Prognostic capability

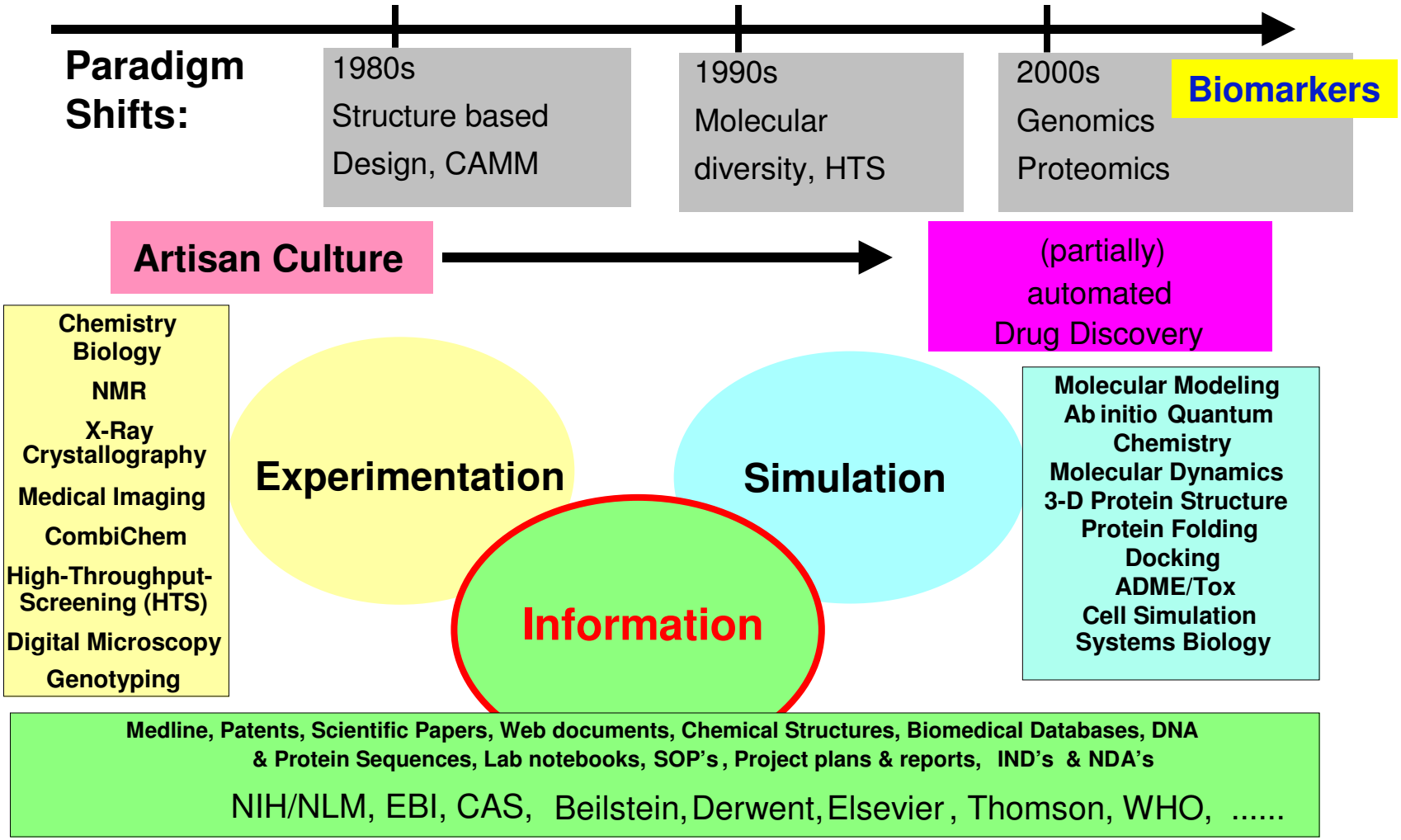
Public – Private Partnerships grow the shared knowledge while preserving Stakeholders' ability to build proprietary IP

Foundation for NIH (FNIH) managed example:

- ***Biomarkers Consortium***



Biomarker-enabled Drug Discovery requires the interplay of Experimentation, Simulation and Information



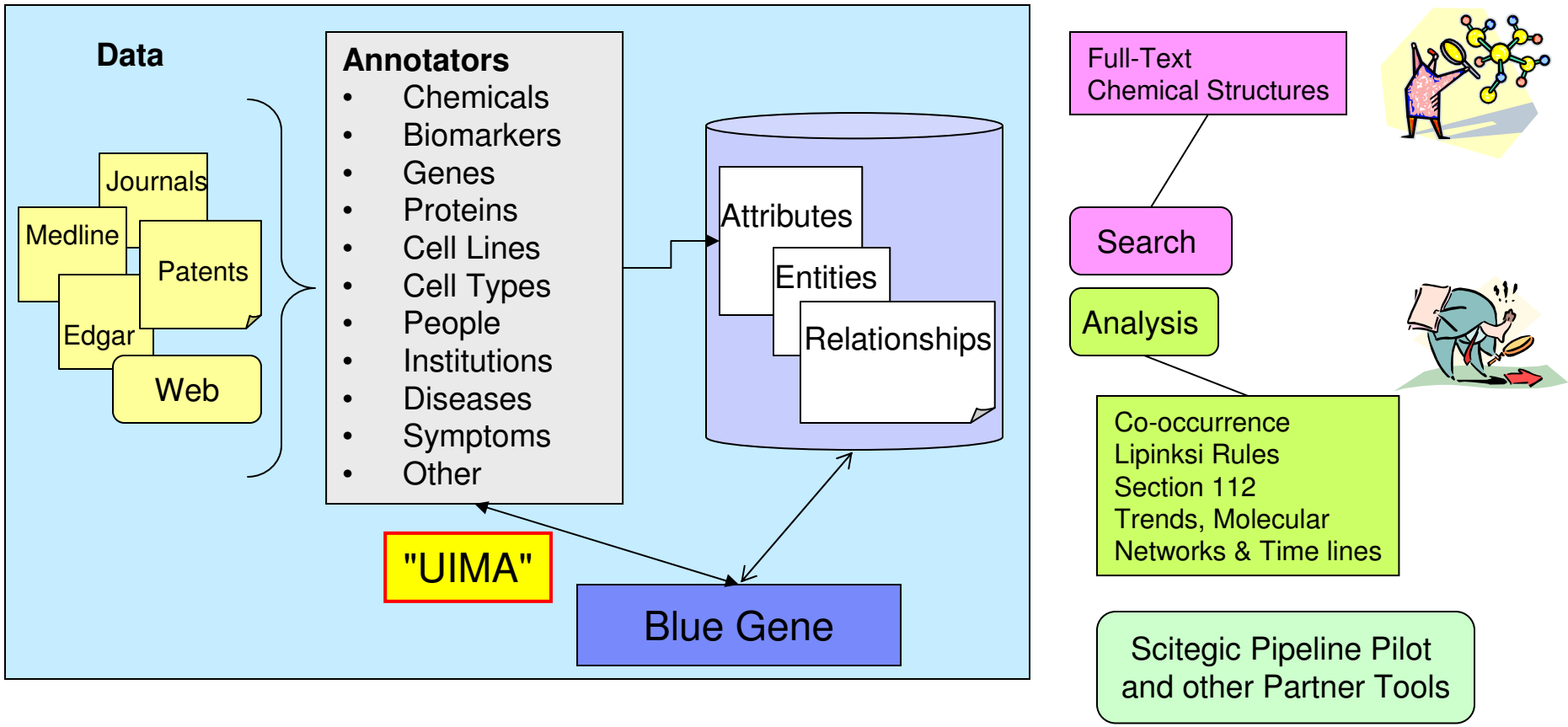
"The pharmaceutical industry
is an information industry"

Peter Drucker

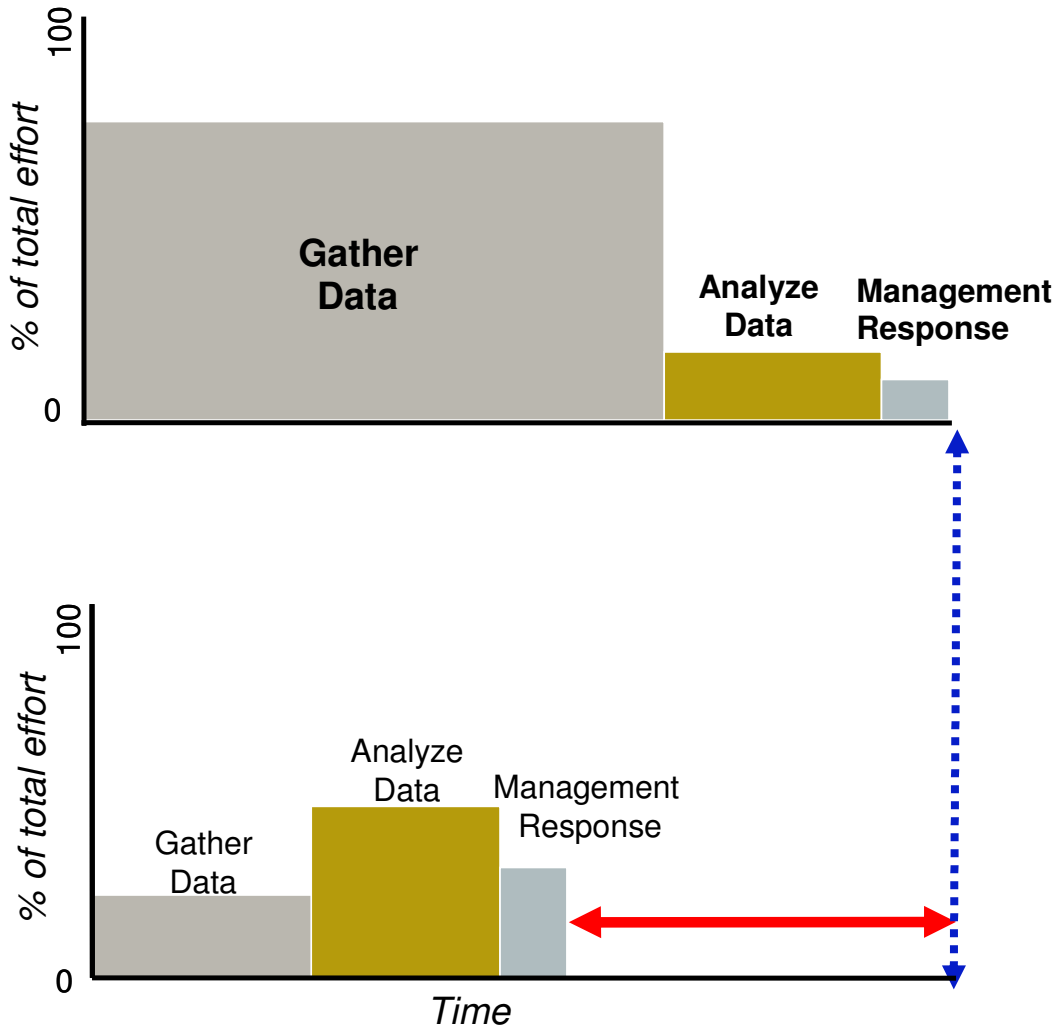
Source: Andrew Hopkins / Pfizer

IBM's Unstructured Information Mgmt/Mining (UIM) Solution

Data Warehouse Created.



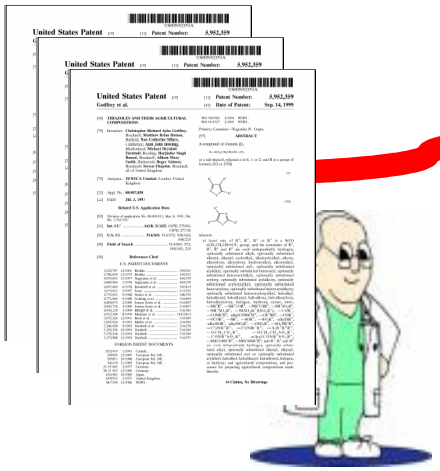
Text Analytics can provide significant business value, IF we

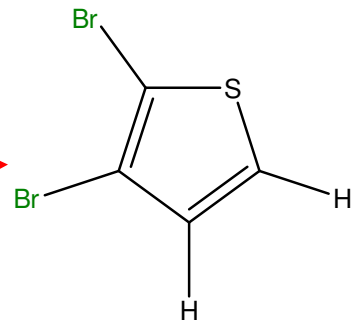


- **Automate the aggregation & search process**
- **Enhance quality using advanced text analytics and other analysis tools - to help determine relevance**
- **Base the solution on an Open and Scalable Architecture**
- **Significantly reduce time dependencies to mission critical information**

➢ **for DECISION SUPPORT**

Example IP Challenge



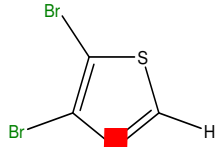


Additional Properties

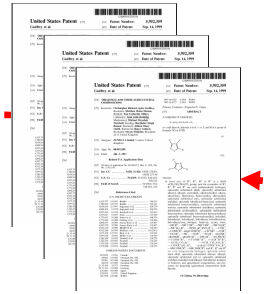
Relationships

How do I find entities from the docs?

How do I find entities' relationships?



New IP

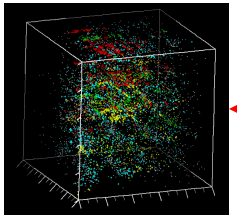


Web, Scientific & News

Worldwide Patents

Medline

High Throughput via Blue Gene



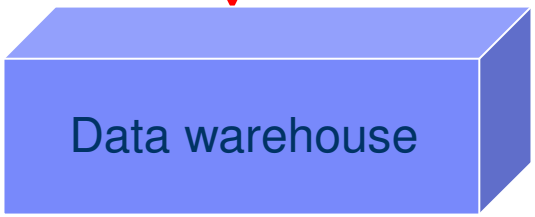
New Insights

How do I utilize other Information sources?

Why use Blue Gene?

- Find and compute the 3D structure of every molecule on every page of every patent (and Medline abs.)
- Identify every **protein** (from our dictionary of >350K proteins) on every page of every patent (and Medline abs.)
- Identify every **disease** (from our list of 14,500) on every page of every patent and map it to Medline MeSH codes
- Identify the occurrence of every **biomarker** (e.g. from a dictionary of biomarkers) on every page of every patent
- …….other entities / annotators (UIMA)

Compute properties & find relationships



Equivalent to 240K simultaneous Google searches!

Outline

- Pharma R&D Productivity and Pipeline
- Biomarkers
- Role of Unstructured Information
- Unstructured Information Mining (UIM)
- **Chemical Names Annotation**
- **Brief DEMO**
- **Blue Gene Supercomputing**
- ***in silico* Drug Discovery**

Can you find the “key molecule(s)” in an “unstructured”, complex scientific journal or patent?

a) (2P/4S)-4-[4-Amino-5-(4-benzyloxy-phenyl)pyrrolo[2,3-d]pyrimidin-7-yl]-2-hydroxymethyl-pyrrolidine-1-carboxylic acid tert-butyl ester prepared analogously to Example 18 starting from (2R/4S)-4-[4-amino-5-(4-benzyloxy-phenyl)-pyrrolo[2,3-d]pyrimidin-7-yl]-pyrrolidine-1,2-dicarboxylic acid 1-tert-butyl ester 2-ethyl ester (Example 20a). 1 H-NMR (CDCl3, ppm): 8.52 (s, 1H), 7.52-7.32 (m, 7H), 7.1 (d, 2H), 6.95 (d,1 H), 5.50 (m, 1H), 5.13 (s, 2H), 4.62-4.42 (m, 2H), 4.28 (m, 2H), 4.10 (m, 1H), 3.95-3.70 (m, 1H), 2.75 (m, 1H), 2.50 (m, 1H),1.49 (s, 9H).

b) (2R/4S)-{4-[4-Amino-5-(4-benzyloxy-phenyl)-pyrrolo[2,3-d]pyrimidin-7-yl]-pyrrolidin-2-yl}-methanol: 0.100 g of (2R/4S)4-[4-amino-5-(4-benzyloxy-phenyl)-pyrrolo[2,3-d]pyrimidin-7-yl]-pyrrolidine-1,2-dicarboxylic acid 1-tert-butyl ester is dissolved in 4 ml of tetrahydrofuran; 10 ml of 4M hydrogen chloride in diethyl ether are added, and stirring is carried out for 1 hour at room temperature. The product is filtered off and dried under a high vacuum. The dihydrochloride of the title compound is obtained. 1 H-NMR (CD3 OD, ppm): 8.4 (s, 1H); 7.60 (s, 1H), 7.5-7.10 (m, 9H), 5.65 (m, 1H), 5.18 (s, 2H), 4.32 (m, 1H), 4.00-3.65 (m, 4H), 2.60 (m, 2H).

EXAMPLE 24

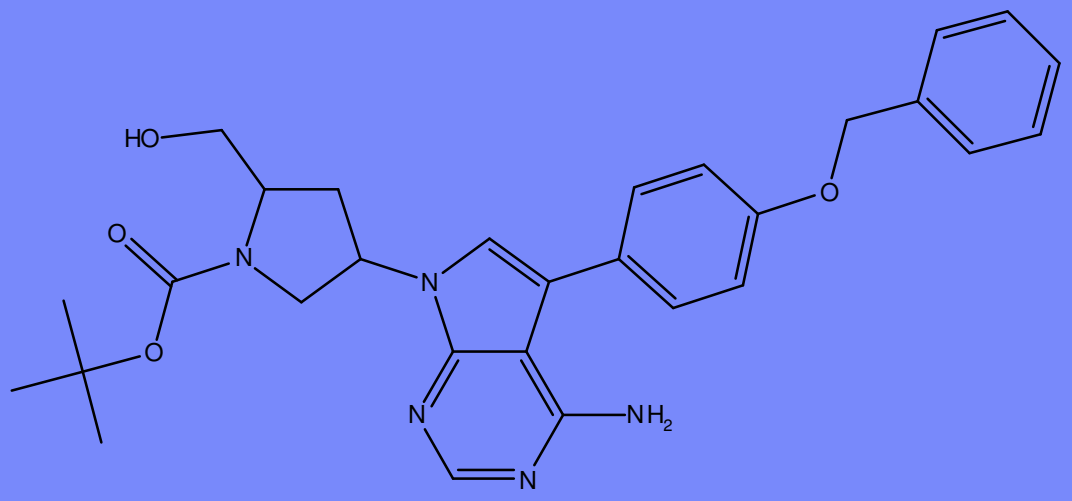
(2R/4S)-4-(4-Amino-5-phenyl-pyrrolo[2,3-d]pyrimidin-7-yl)-1-(2,2-dimethyl-propionyl)-pyrrolidine-2-carboxylic acid ethyl ester 0.130 g of (2R/4S)-4-(4-benzyloxycarbonylamino-5-phenyl-pyrrolo[2,3-d]pyrimidin-7-yl)-1-(2,2-dimethyl-propionyl)-pyrrolidine-2-carboxylic acid ethyl ester is dissolved in 8 ml of methanol, and the solution is hydrogenated over 0.030 g of palladium-on-carbon (10%) for 1 hour at normal pressure. The catalyst is removed by filtration, the filtrate is concentrated by

Use Text Analytics / Chemical Names Annotation!

a) **(2P/4S)-4-[4-Amino-5-(4-benzyloxy-phenyl)pyrrolo[2,3-d]pyrimidin-7-yl]-2-hydroxymethyl-pyrrolidine-1-carboxylic acid tert-butyl ester** prepared analogously to Example 18 starting from (2R/4S)-4-[4-amino-5-(4-benzyloxy-phenyl)-pyrrolo[2,3-d]pyrimidin-7-yl]-pyrrolidine-1,2-dicarboxylic acid 1-tert-butyl ester 2-eth (Example 20a). 1 H-NMR (CDCl3, ppm): 8.52 (s, 1H), 7.52-7.32 (m, 7H), 7.1 (d, 2H), 6.95 (d 5.13 (s, 2H), 4.62-4.42 (m, 2H), 4.28 (m, 2H), 4.10 (m, 1H), 3.95-3.70 (

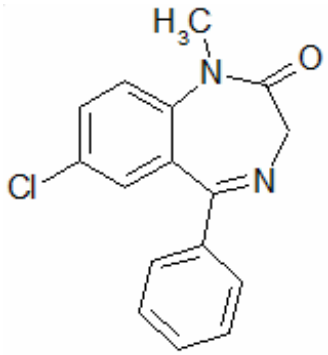
b) (2R/4S)-{4 methanol: 0.10 pyrrolidine-1,2- of 4M hydrogen c temperature. T the title compound l), 5.65 (m, 1H), 5 EXAMPLE 24 (2R/4S)-4-(4-A -2- carboxylic acid 3- d]pyrimidin-7-yl in 8 ml of methanol, r 1 hour at normal pre

What is this compound ??



Problem – I need to find information about Valium

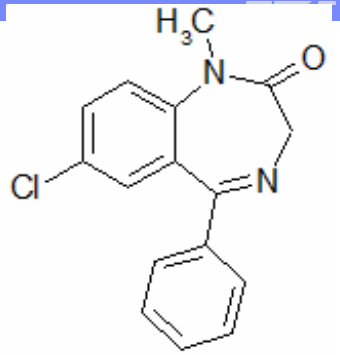
Valium (Trade Name) = Diazepam (Generic Name) = CAS # 439-14-5 (Chemical ID #) =



Valium has > 149 "names"

ALBORAL, ALISEUM, ALUPRAM , AMIPROL ,ANSIOLIN , ANSIOLISINA , APAURIN, APOZEPAM, ASSIVAL , ATENSINE , ATILEN , BIALZEPAM , CALMOCITENE, CALMPOSE , CERCINE, CEREGULART, CONDITION, DAP, DIACEPAN, DIAPAM , DIAZEMULS , DIAZEPAN , DIAZETARD , DIENPAX, DIPAM , DIPEZONA, DOMALIUM , DUKSEN, DUXEN, E-PAM, ERIDAN, EVACALM, FAUSTAN, FREUDAL , FRUSTAN, GIHITAN, HORIZON, KIATRIUM, LA-III , LEMBROL, LEVIUM, LIBERETAS , METHYL DIAZEPINONE, MOROSAN , NEUROLYTRIL NOAN NSC-77518 PACITRAN PARANTEN PAXATE PAXEL PLIDAN QUETINIL QUIATRIL QUIEVITA RELAMINAL RELANIUM RELAX RENBORIN RO 5-2807 S.A. R.L. SAROMET SEDAPAM SEDIPAM SEDUKSEN SEDUXEN , SERENACK SERENAMIN SERENZIN SETONIL SIBAZON SONACON STESOLID STESOLIN , TENSOPAM TRANIMUL TRANQDYN TRANQUASE TRANQUIRIT , TRANQUO-TABLINEN , UMBRIUM UNISEDIL USEMPAX AP VALEO VALITRAN VALRELEASE VATRAN VELIUM, VIVAL VIVOL WY-3467

There are many different chemical names for Valium



Valium = Diazepam = CAS # 439-14-5 =

7-CHLORO-1-METHYL-5-PHENYL-2H-1,4-BENZODIAZEPIN-2-ONE

7-CHLORO-1-METHYL-5-PHENYL-3H-1,4-BENZODIAZEPIN-2(1H)-ONE

7-CHLORO-1-METHYL-5-PHENYL-1,3-DIHYDRO-2H-1,4-BENZODIAZEPIN-2-ONE

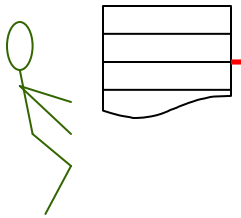
7-CHLORO-1-METHYL-2-OXO-5-PHENYL-3H-1,4-BENZODIAZEPINE

1-METHYL-5-PHENYL-7-CHLORO-1,3-DIHYDRO-2H-1,4-BENZODIAZEPIN-2-ONE

7-CHLORO-1,3-DIHYDRO-1-METHYL-5-PHENYL-2H-1,4-BENZODIAZEPIN-2-ONE

7-CHLORO-1-METHYL-5-3H-1,4-BENZIODIAZEPIN-2(1H)-ONE

Problems of 'taxonomy" & name normalization

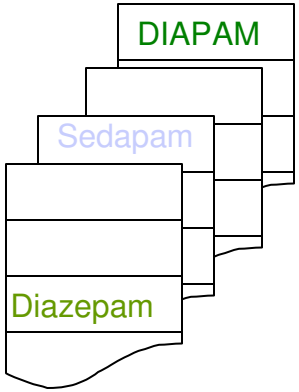


Valium

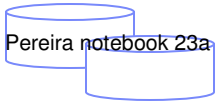
Choose keywords

Taxonomies & Dictionaries

Medline



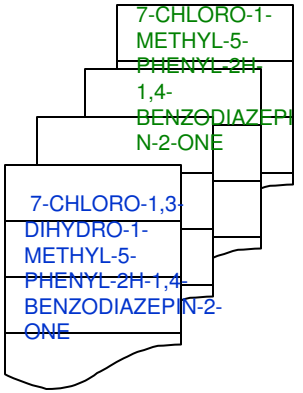
In-house database



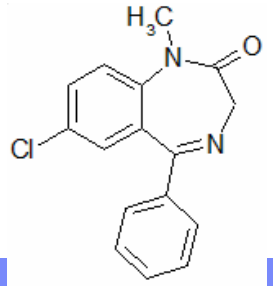
Chem. Abstracts

439-14-5
(Chemical ID)

Patent database



Multiple documents contain Information about Valium



Chemical Structure Annotation DEMO:

Example:

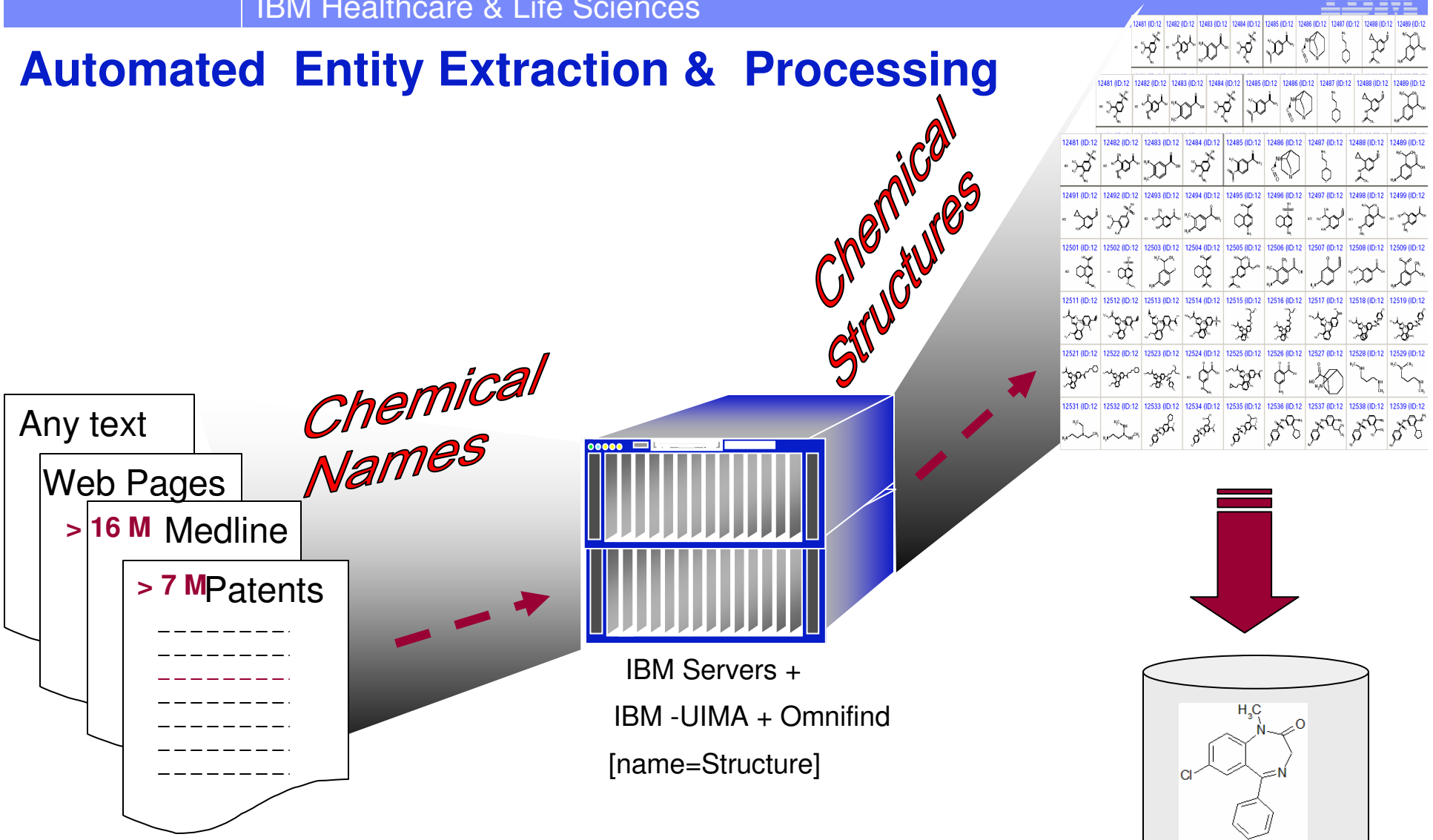
US6051577: **N-7-heterocyclyl pyrrolo[2,3-D]pyrimidines** and the use thereof

Title: "New 7-(pyrrolidinyl or piperidinyl)-pyrrolo[2,3-d]pyrimidine derivatives - useful as protein-tyrosine-kinase inhibitors for treatment of e.g. osteoporosis, Paget's disease, cardiovascular diseases and cancers"

Assignee: Novartis

Inventor: Altmann, Eva; Reinach, Switzerland

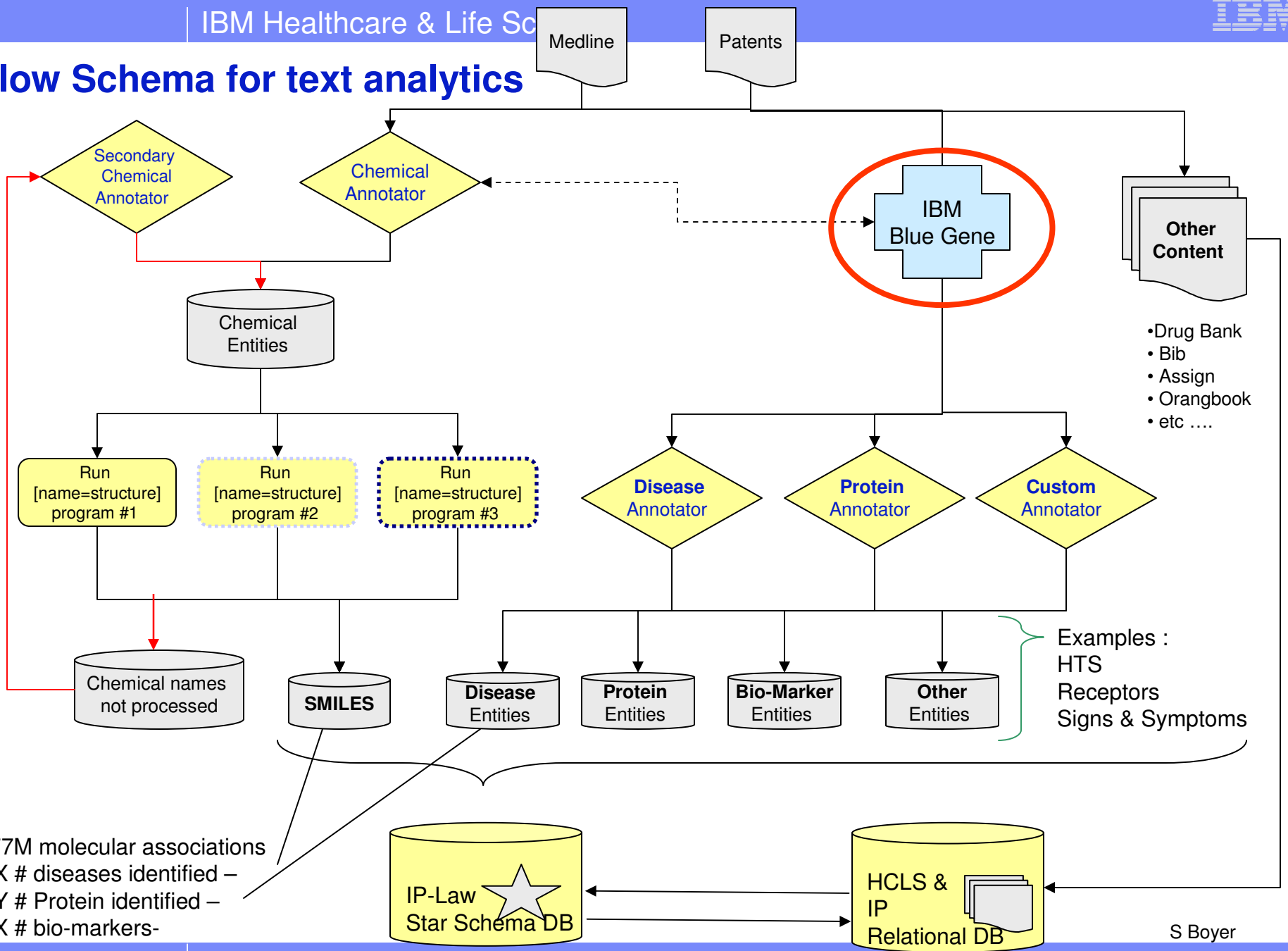
Automated Entity Extraction & Processing



Build large chemical data bases - & selectively monitor literature for chemical entities

S Boyer

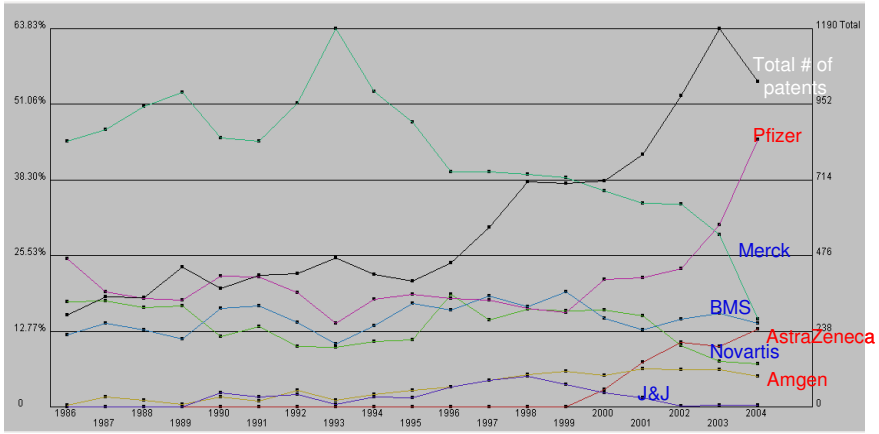
Flow Schema for text analytics



Example: Identify Emerging IP/Patent Patterns (using UIM)

- Looked at 18 years of US patent data to show how pharmaceutical market positioning
- Pfizer, AZ, and Amgen are increasing their patent activity
- Comparing most relevant words/terms in patent data emerging patterns identified
- Genentech is staking out white space in the areas not covered by the other major pharmaceuticals.

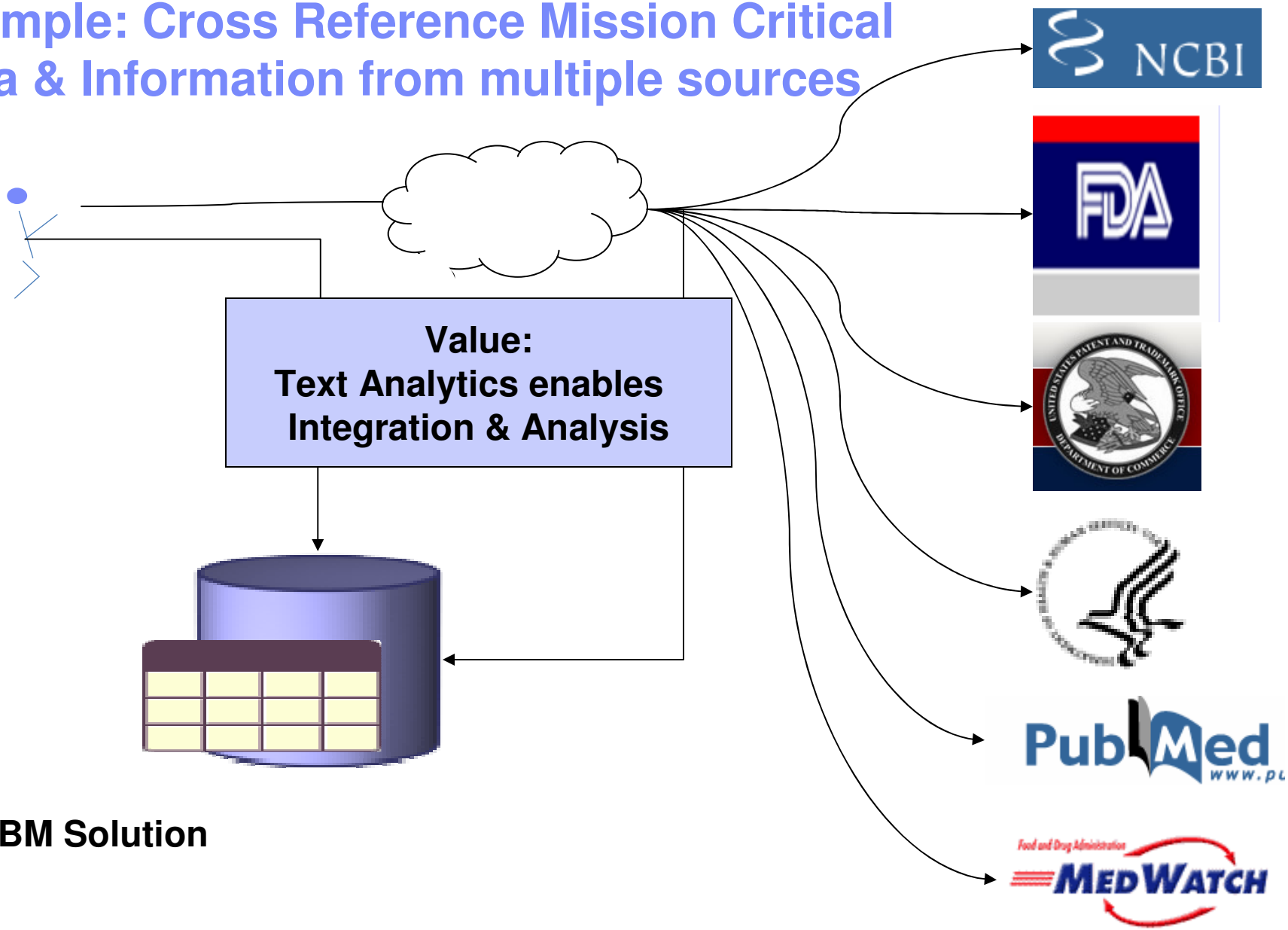
- **Value:**
  - **Identify potential new market areas**
  - **Competitive intelligence**
  - **Make Go / No Go Decisions**



Very High Affinity = Moderate Affinity = Low Affinity = No Affinity =

Term	Count	PFIZER I...	ASTRAZ...	AMGEN	GENENT...	Novartis+	MERCK ...	BRISTOL...	Johnson ...
alzheimer	321	304 (0.0)	7 (1.0)	18 (6.75384	1 (1.0)	2 (1.0)	36 (1.0)	43 (0.27108	0 (1.0)
anti-inflam...	367	115 (1.3328	11 (0.55007	10 (0.88916	5 (1.0)	50 (0.87330	130 (0.4197	41 (1.0)	5 (0.521284
arthritis	577	232 (1.3305	29 (8.89878	36 (2.71071	57 (1.0)	36 (1.0)	128 (1.0)	59 (1.0)	0 (1.0)
asthma	552	213 (1.4128	26 (8.12254	17 (0.48193	52 (1.0)	28 (1.0)	164 (1.0)	52 (1.0)	0 (1.0)
breast	1384	62 (1.0)	4 (1.0)	3 (1.0)	1238 (0.0)	13 (1.0)	41 (1.0)	23 (1.0)	0 (1.0)
cancer	785	253 (6.0049	19 (1.0)	41 (2.63881	115 (1.0)	30 (1.0)	233 (1.0)	94 (0.64142	0 (1.0)
cardiovasc	416	156 (2.8534	15 (0.15002	1 (1.0)	23 (1.0)	5 (1.0)	140 (0.9309	76 (1.01206	0 (1.0)
cartilage	474	8 (1.0)	0 (1.0)	1 (1.0)	449 (0.0)	4 (1.0)	7 (1.0)	5 (1.0)	0 (1.0)
cervical	985	5 (1.0)	0 (1.0)	0 (1.0)	979 (0.0)	0 (1.0)	0 (1.0)	1 (1.0)	0 (1.0)
coding_se...	1782	11 (1.0)	1 (1.0)	12 (1.0)	1740 (0.0)	0 (1.0)	7 (1.0)	2 (1.0)	0 (1.0)
colon	1307	48 (1.0)	2 (1.0)	7 (1.0)	1215 (0.0)	0 (1.0)	7 (1.0)	20 (1.0)	0 (1.0)
delivery	268	37 (1.0)	13 (0.01380	25 (3.62070	14 (1.0)	26 (1.0)	108 (0.0167	34 (0.52423	11 (4.89102
dna	2473	47 (1.0)	2 (1.0)	187 (0.0)	1907 (0.0)	27 (1.0)	196 (1.0)	57 (1.0)	0 (1.0)
gastrointes...	397	177 (2.7908	33 (8.39105	9 (1.0)	7 (1.0)	15 (1.0)	116 (1.0)	40 (1.0)	0 (1.0)
gene	1169	72 (1.0)	16 (1.0)	75 (3.03985	745 (0.0)	43 (1.0)	129 (1.0)	48 (1.0)	1 (1.0)
growth_hor...	312	67 (0.26885	4 (1.0)	8 (1.0)	113 (4.1137	2 (1.0)	112 (0.3564	6 (1.0)	0 (1.0)
heart	461	107 (0.0)	2 (1.0)	3 (1.0)	37 (1.0)	32 (1.0)	132 (1.0)	44 (1.0)	4 (1.0)
immune	352	109 (8.5430	1 (1.0)	18 (0.00292	82 (5.85665	7 (1.0)	85 (1.0)	50 (0.10208	0 (1.0)
kinase	245	48 (0.82418	10 (0.11448	26 (2.54970	44 (0.55085	10 (1.0)	44 (1.0)	41 (0.00900	2 (1.0)
liver	1329	47 (1.0)	2 (1.0)	12 (1.0)	1204 (0.0)	9 (1.0)	41 (1.0)	15 (1.0)	0 (1.0)
lung	1466	76 (1.0)	5 (1.0)	13 (1.0)	1368 (0.0)	7 (1.0)	68 (1.0)	19 (1.0)	0 (1.0)
pain	529	268 (0.0)	51 (2.14510	21 (0.04642	2 (1.0)	27 (1.0)	138 (1.0)	19 (1.0)	3 (1.0)
rheumatoid	425	154 (6.2229	25 (7.12472	30 (5.78526	53 (1.0)	22 (1.0)	99 (1.0)	42 (1.0)	0 (1.0)
stroke	405	116 (0.0)	19 (0.00463	16 (0.08691	12 (1.0)	12 (1.0)	82 (1.0)	48 (0.80072	0 (1.0)
tumor	1908	98 (1.0)	10 (1.0)	42 (1.0)	1333 (0.0)	42 (1.0)	183 (1.0)	200 (1.0)	0 (1.0)
vaccine	178	41 (0.17239	3 (1.0)	7 (0.265987	17 (1.0)	3 (1.0)	107 (3.5736	0 (1.0)	0 (1.0)
vascular	350	118 (1.6240	9 (0.944719	2 (1.0)	91 (1.56106	17 (1.0)	82 (1.0)	28 (1.0)	3 (1.0)
virus	317	68 (0.26957	0 (1.0)	10 (0.53945	27 (1.0)	33 (1.0)	137 (2.0156	42 (0.31203	0 (1.0)
Total	17701	3370	445	462	2930	2362	5922	2028	182

Example: Cross Reference Mission Critical Data & Information from multiple sources



IBM Solution

Outline

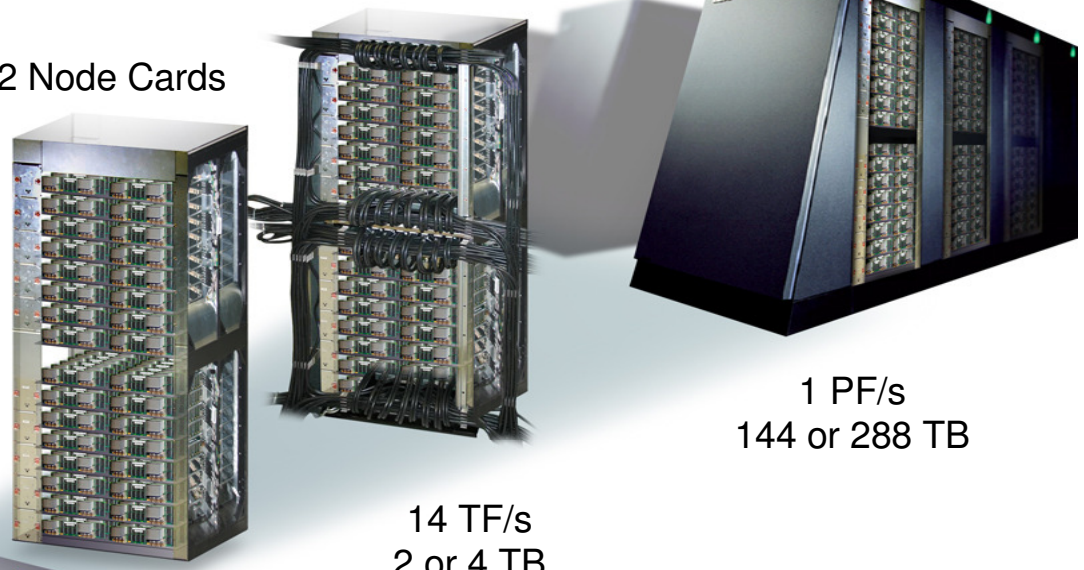
- Pharma R&D Productivity and Pipeline
- Biomarkers
- Role of Unstructured Information
- Unstructured Information Mining (UIM)
- Chemical Names Annotation
- Brief DEMO
- **Blue Gene Supercomputing**
- ***in silico* Drug Discovery**

Blue Gene – Ultrascalable multi-PetaFlop Performance

System scales to 256 racks achieving 3.56 PF/s peak

System
72 Racks, 72x32x32

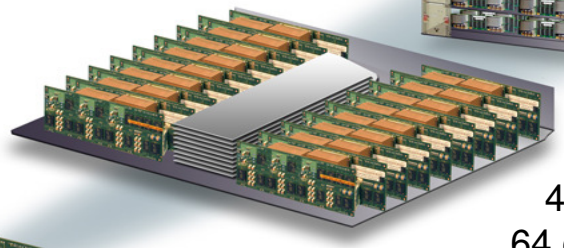
Rack
Cabled 8x8x16
32 Node Cards



1 PF/s
144 or 288 TB

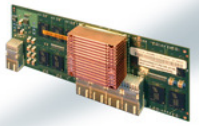
14 TF/s
2 or 4 TB

Node Card
(32 chips 4x4x2)
32 compute, 0-1 IO cards



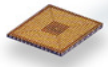
435 GF/s
64 or 128 GB

Compute Card
1 chip, 20 DRAMs



13.6 GF/s
2 or 4 GB DDR2

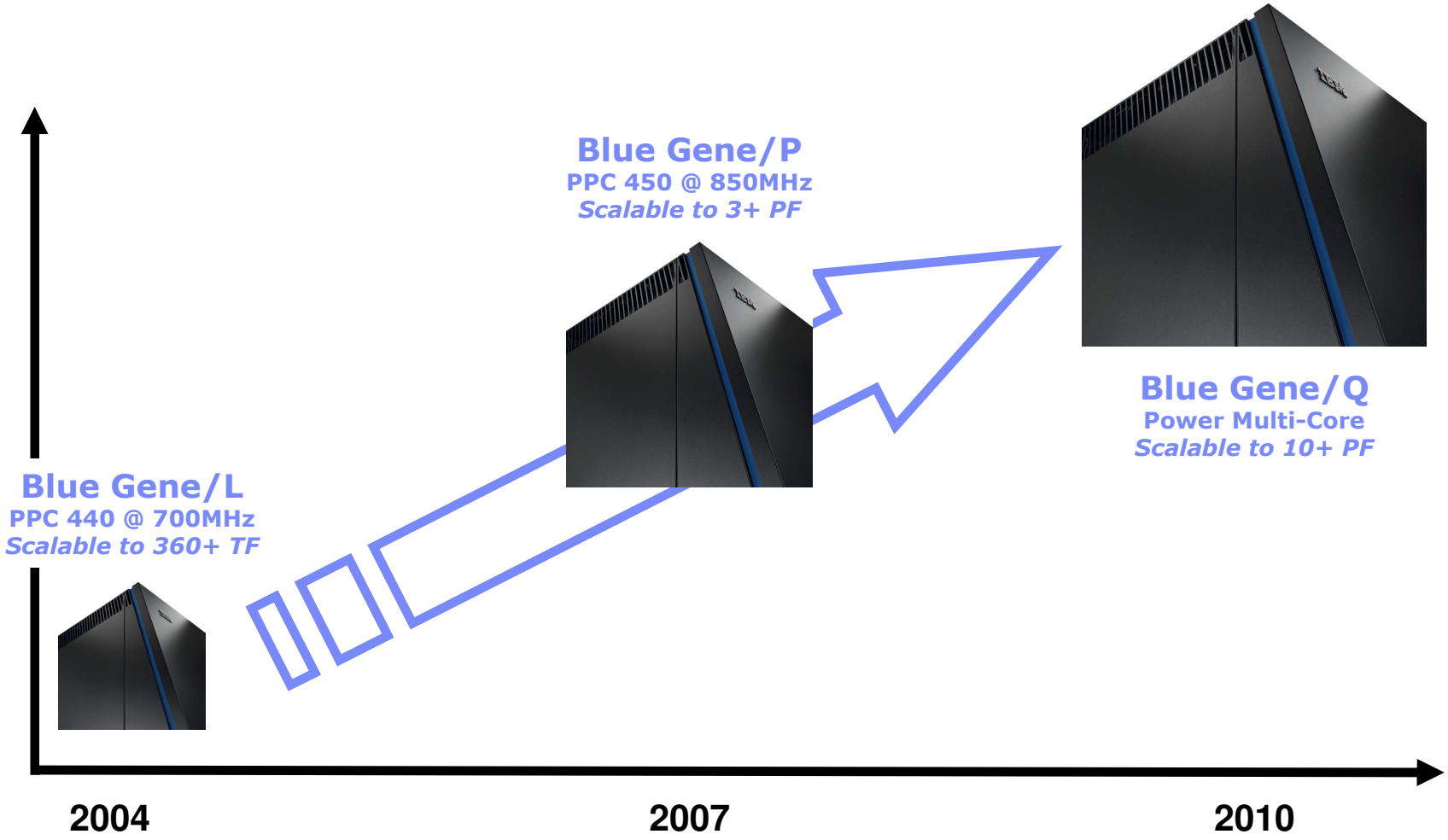
Chip
4 processors



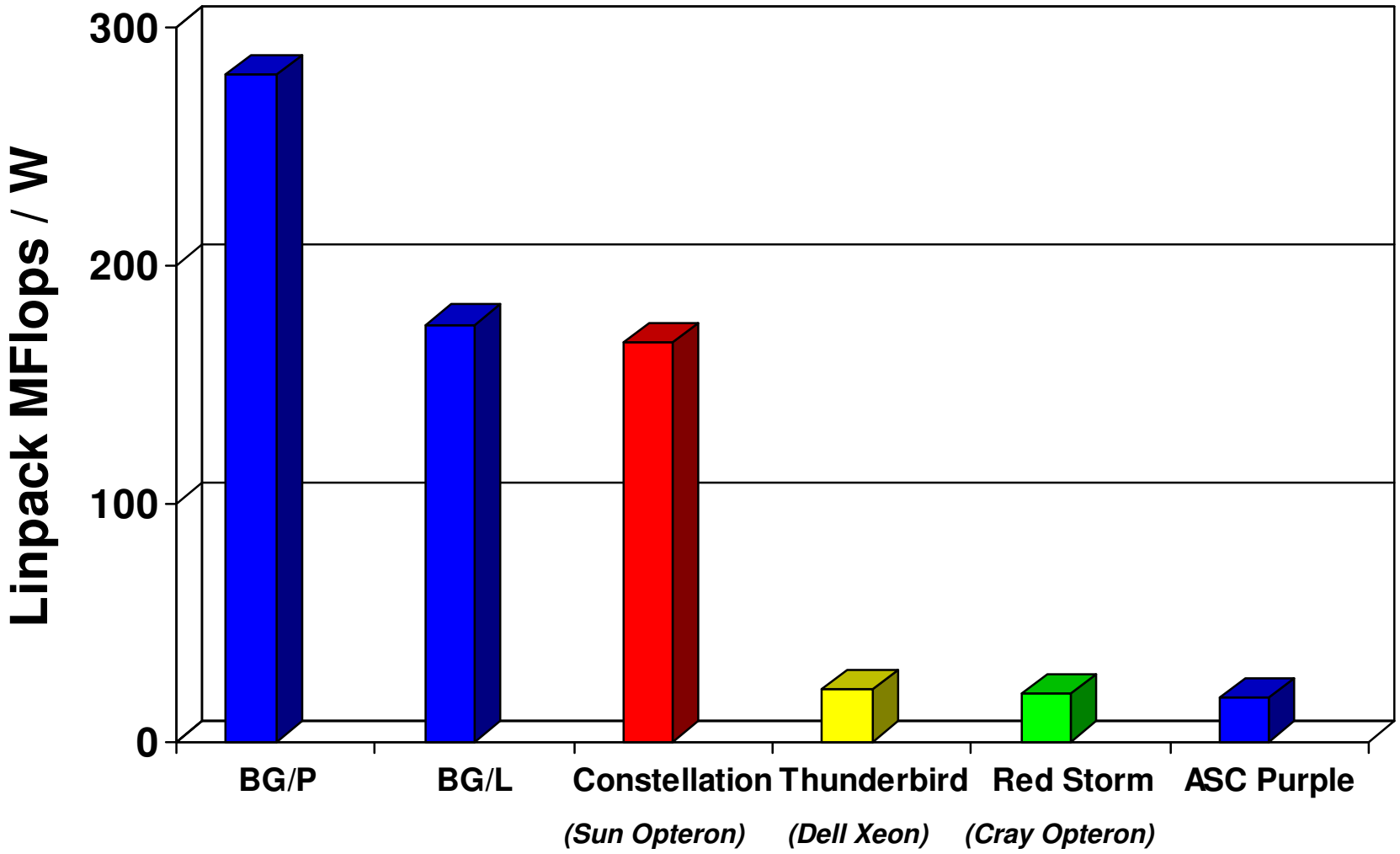
13.6 GF/s
8 MB EDRAM

Blue Gene continues its leadership performance in a space-saving, power-efficient package for the most performance demanding applications

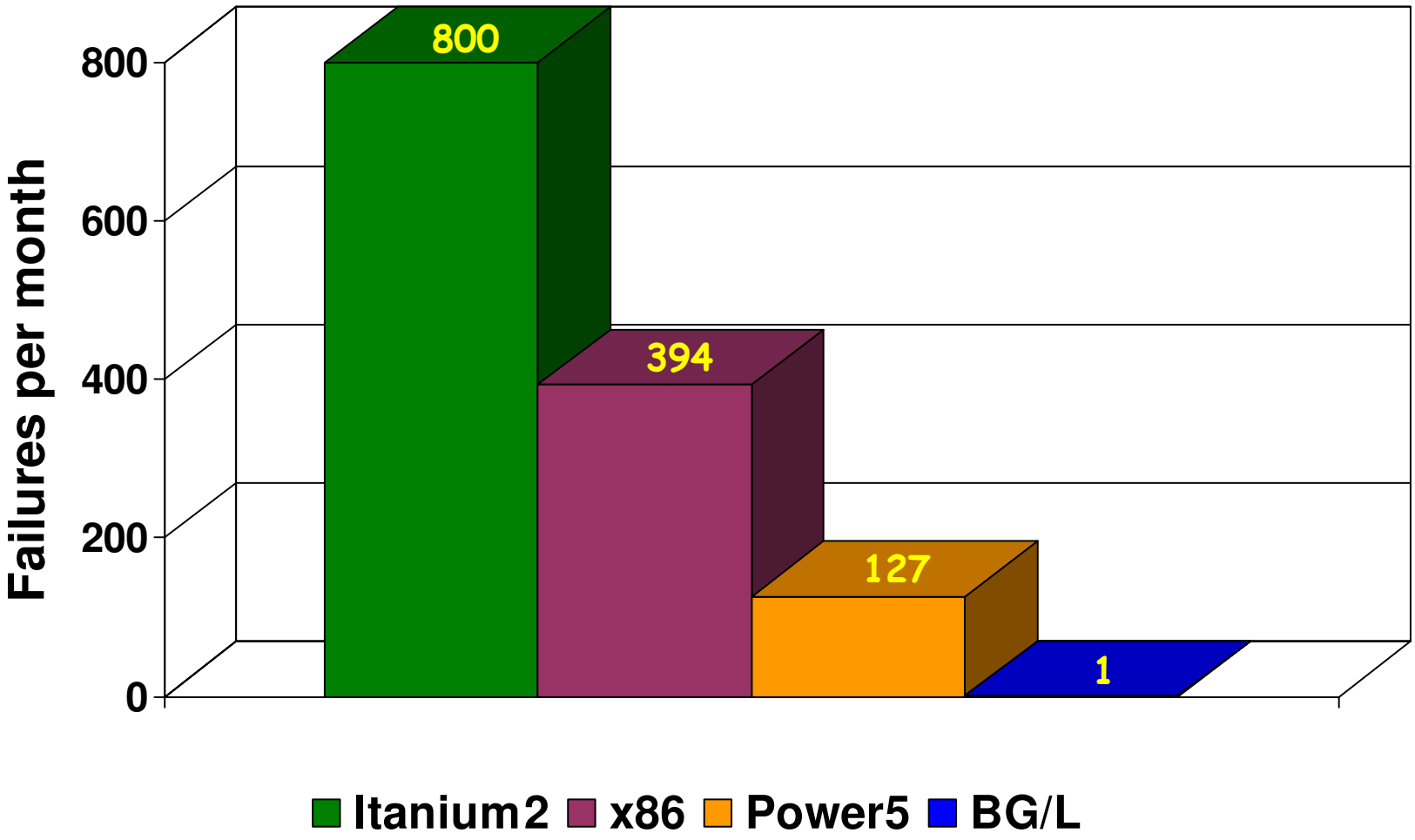
Blue Gene technology roadmap



Blue Gene/P's power efficiency is even better



Relative reliability @ 100 TFlops



Results of survey conducted by Argonne National Lab on 10 clusters ranging from 1.2 to 365 TFlops (peak); excluding storage subsystem, management nodes, SAN network equipment, software outages

BG/L @ LLNL

- **4 in TOP20**
- **22 in TOP100**
- **34 in TOP400**

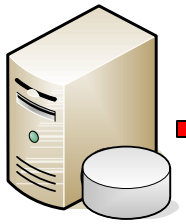
www.top500.org (6/07)

Blue Gene's Award-Winning Performance

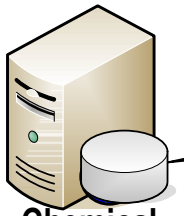
June 6, 2006 WASHINGTON, D.C. — The Department of Energy's National Nuclear Security Administration (NNSA) and IBM teamed up to announce that a new mark was achieved on the world's fastest supercomputer named Blue Gene/L (BG/L). This world record for a scientific application was set by achieving a sustained performance of **207.3 trillion floating-point operations per second (teraFLOPS) on the "Qbox" computer code** for conducting materials science simulations critical to national security.

The reference implementation… (annotation process)

Patent Database

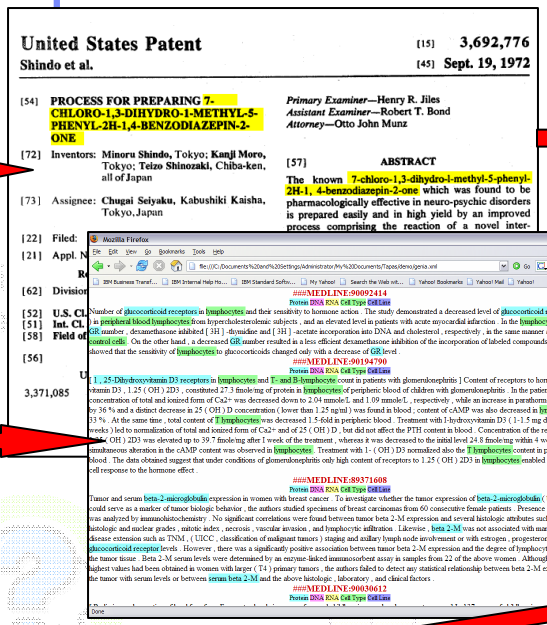


Medline Database



Chemical Annotations

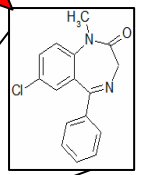
Chemical Properties

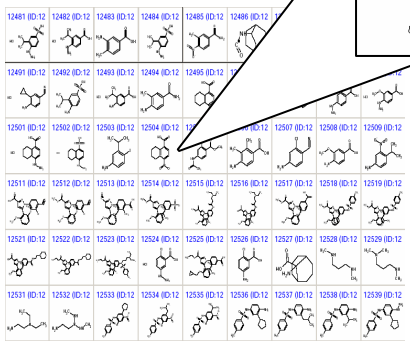


Name → Structure Conversion

SMILES strings:
c1ccccc1

Language Neutral Format

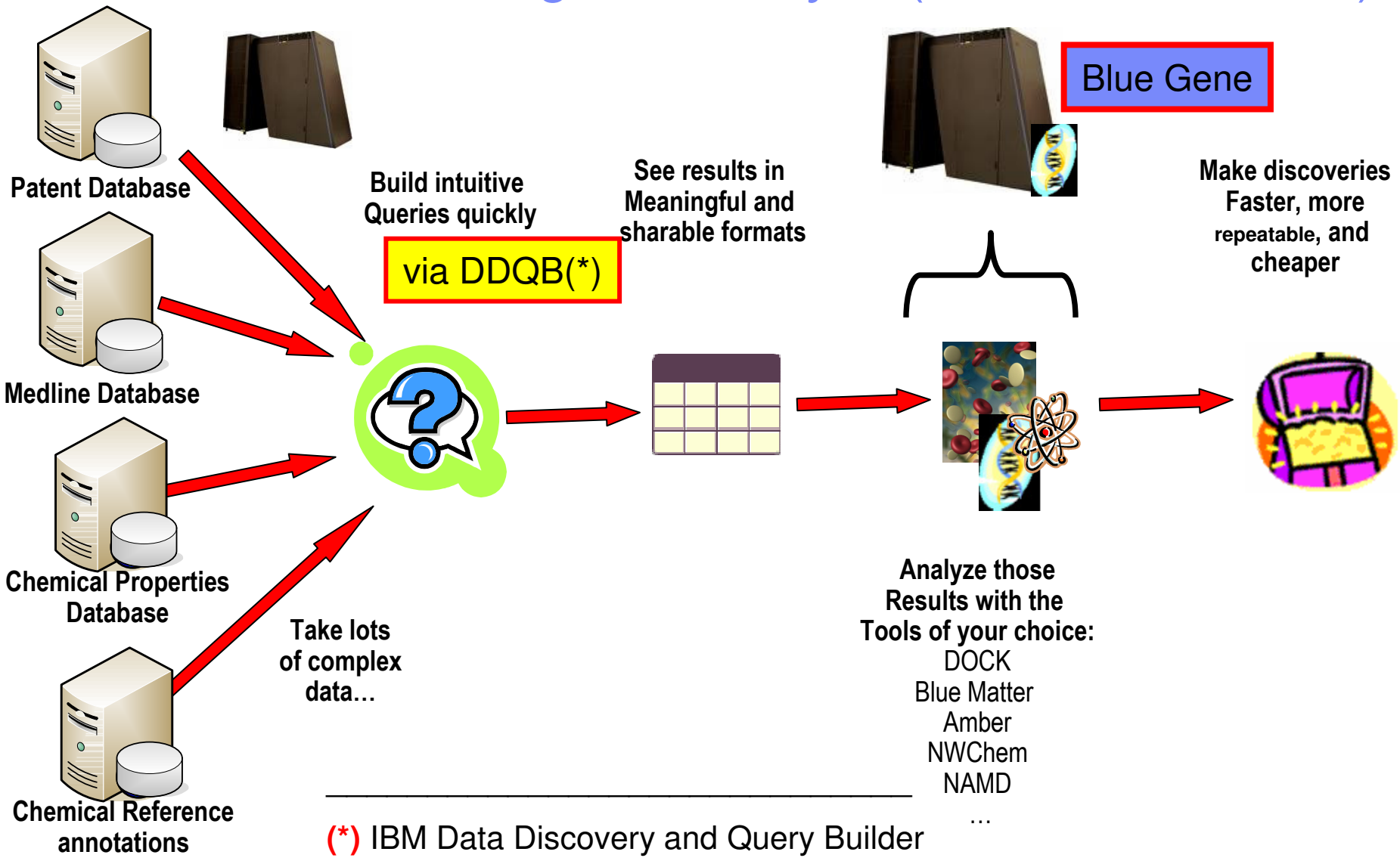




Connection tables

```
  6  6  0  0  0  0  0  0  0  0999 V2000
    6.7092    5.6087    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    6.7076    4.5056    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    7.6607    3.9551    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    8.6160    4.5062    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    8.6121    5.6136    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    7.6583    6.1591    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  2  0  0  0  0
  2  3  1  0  0  0  0
  3  4  2  0  0  0  0
  4  5  1  0  0  0  0
  5  6  2  0  0  0  0
  6  1  1  0  0  0  0
M  END
```

Towards in silico Drug Discovery… (UIM + Simulation)



(*) IBM Data Discovery and Query Builder

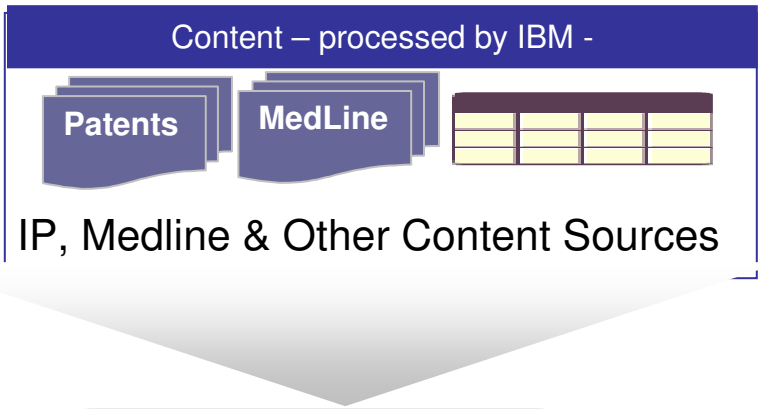
Acknowledgments (IBM)

- Steve Boyer, Almaden, CA
- James Rhodes, Almaden, CA
- David Martin, Almaden, CA
- Jeff Kreulen, Almaden, CA
- Tom Griffin, Almaden, CA
- Brad Wade, Almaden, CA
- David Zirl, Almaden, CA
- Jim Cooper, Yorktown, NY
- Anni Coden, Yorktown, NY
- Richard Dettinger, Rochester, MN
- Amanda Peters, Rochester, MN
- Carl Obert, Rochester, MN

Note:

INFOCHEM
is providing similar
capabilities based on
both Public Domain
and Infochem CONTENT!

Public databases



Available via InfoChem's on-line hosted Services

IBM Text Analytics & OmniFind System

IBM Chemical database
Derived from unstructured text

InfoChem databases

InfoChem Legacy data + Spresi db

InfoChem database
Derived from Spresi + Springer content