TMS

PROMOTING THE GLOBAL SCIENCE AND ENGINEERING PROFESSIONS CONCERNED WITH MINERALS, METALS, AND MATERIALS

Materials Informatics: Theory and Application

Krishna Rajan

Department of Materials Science and Engineering
NSF Intl. Materials Institute Combinatorial Sciences &
Materials Informatics Collaboratory
Iowa State University

MS&T'06

October 15th 2006

Materials Science & Technology 2006 Conference and Exhibition
October 15-19, 2006 - Duke Energy Center - Cincinnati, OH

learn
explore
experience

TMS

PROMOTING THE GLOBAL SCIENCE AND ENGINEERING PROFESSIONS CONCERNED WITH MINERALS, METALS, AND MATERIALS

NSF

National Science Foundation

WHERE DISCOVERIES BEGIN

NSF-IMI

COSMIC

International Materials Institute

MatDL

Pacific Northwest National Laboratory

Operated by Battelle for the U.S. Department of Energy

Thermo-Calc Software

1:00 **Introduction to workshop**
T. Marechaux and L. Peurrung

1:15: **Materials Informatics: Theory and Application**
K. Rajan, Iowa State University

1:30 **NSDL Materials Digital Library Pathway**
L. Bartolo, Kent State University

Processing information

2:15 **Thermo-Chemistry to Phase Diagrams and More**
P. Mason, ThermoCal

3:00 Break

3:20 **Data Patterns and Links to Materials Theory**
K. Ferris, Pacific Northwest National Laboratory

4:00 **Data Mining and Materials Informatics**
K. Rajan, Iowa State University

4:40 **Conclusion and Discussion**

TMS

PROMOTING THE GLOBAL SCIENCE AND ENGINEERING PROFESSIONS CONCERNED WITH MINERALS, METALS, AND MATERIALS

Materials Informatics: Growing from the Bio World

Warren H. Hunt, Jr.

Author's Note: [illegible]

The materials science community has seen explosive growth in the volume and complexity of materials information. Materials are becoming more engineered, complex, and multifunctional, and simple models and parametric approaches are often inadequate to predict their properties, performance, and reliability. Moreover, materials data is becoming increasingly complex, involving numerical values, textual annotations, spatial information (e.g., micrographs), and logical information. Materials information and how it is compiled and validated needs to evolve, or we as a community will not be able to produce new advances in materials rapidly nor take full advantage of existing materials capabilities in design.

When these same issues faced the biological sciences community, development of a new field known as bioinformatics resulted. Based on the National Institutes of Health definition, bioinformatics is the research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral, or health data, including those to acquire, store, organize, archive, analyze, or visualize such data. Materials informatics, drawing on principles developed in bioinformatics, is emerging as a new approach for addressing these needs in our field.

Materials informatics fuses collections of annotated materials data. The fused data is interrogated in a structured learning framework (i.e., with intelligent data mining tools that include physical models and get "smarter" as the search proceeds). In this way, not just data, but knowledge is extracted. Combined with high-throughput experimental techniques, materials informatics offers a more efficient manner to develop new materials with optimal properties.

On December 8–9, 2005, a group of individuals gathered in Washington, D.C., to discuss materials informatics. The participants developed a working definition, identified and prioritized needs, and outlined actions to advance this new field of materials science. The workshop was sponsored by Pacific Northwest National Laboratory, ASM International and TMS. Outcomes of the meeting are presented in the sidebar.

Since the December workshop, a steering committee has been meeting biweekly via conference call. The group worked to move forward on selected recommendations from the workshop and begin to build the materials informatics community. A key upcoming community event is the Materials Informatics workshop scheduled for Sunday, October 15 from 1–5 p.m. in conjunction with the Materials Science & Technology 2006 conference in Cincinnati, Ohio. Further information is available at www.matscitech.org or through organizer Krishna Rajan, a professor at Iowa State University, at krajan@iastate.edu. More information is also available from the TMS Technical Services Department at techstaliaison@tms.org.

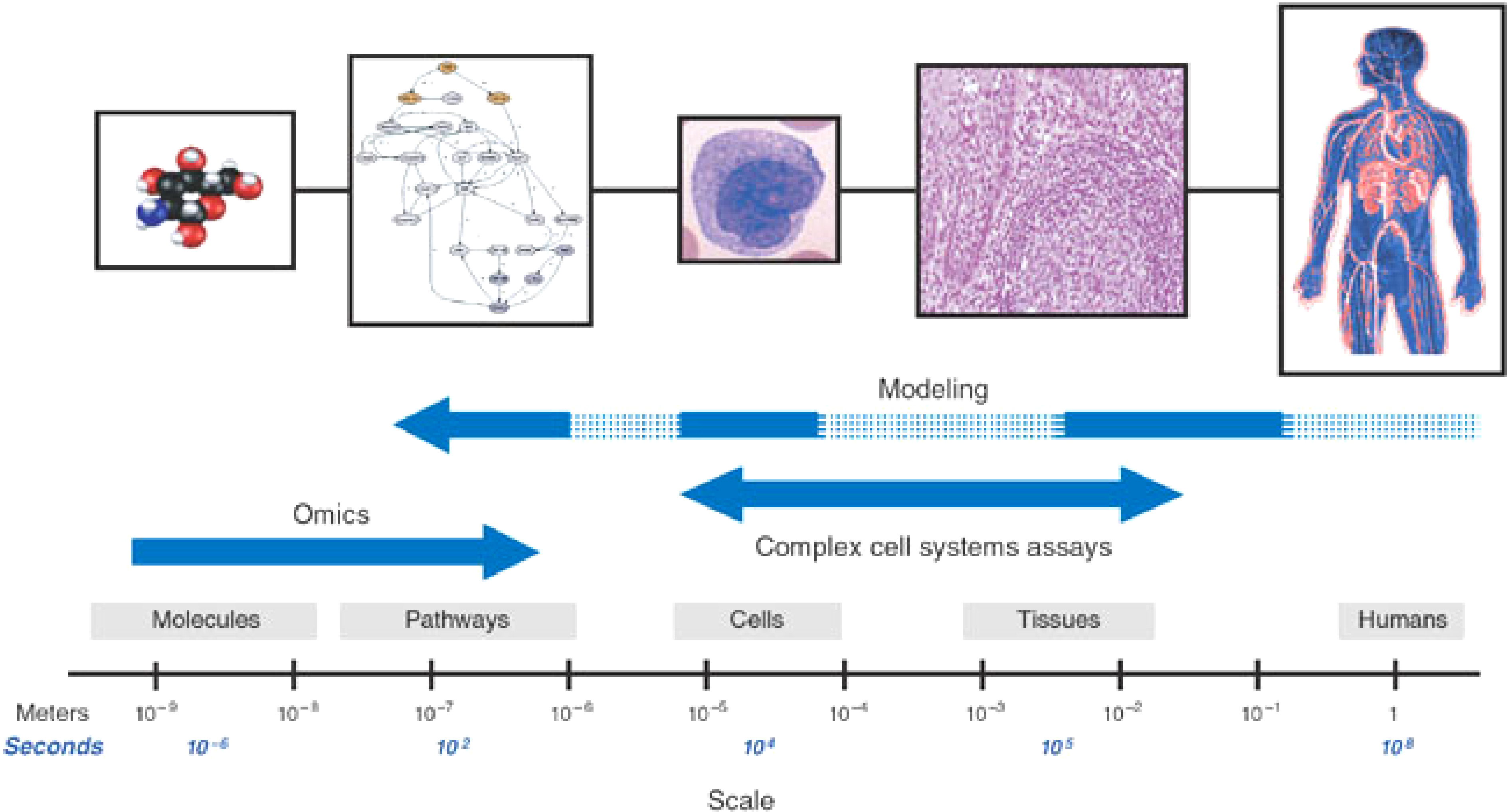
Warren H. Hunt is executive director of TMS.

WORKSHOP OUTCOMES

One of the outcomes of the materials informatics workshop was agreement on a definition: materials informatics is "the high speed robust acquisition, management, analysis and dissemination of diverse materials data." Also, a number of areas requiring action to move materials informatics forward were identified. These were:

- Data acquisition: The barriers to data acquisition include finding sources of data, locating data within the sources, extracting sufficient metadata to make the property data meaningful, harmonizing the information from different sources, and collecting the results in a fully documented, searchable, and accessible data system. The group identified several means by which the materials community as a whole could contribute to improving data access and acquisition.
- Data standards: Specific needs included a listing and taxonomy of the elements of an information framework to describe materials as well as both de-facto and formalized standards for data formats.
- Dissemination/access: A requirement to define access policies was identified. It was envisaged that different tiers will probably be required to control access to the data depending on the status of who is trying to access it.
- Analysis tools: Knowledge extraction from the fused data sets through the detection of knowledge gaps, identification of anomalies, model development, pattern recognition, and mapping (forward and inverse) techniques is needed. These capabilities need to handle the various forms of information, conform to the likely data standards, and must be joined by a system architecture, as needed by the combinatorial community.
- Visibility/awareness: The practitioners of materials science and related fields need to become educated about informatics approaches, even to the point of developing materials informatics as a recognized new sub-discipline within materials science. The group identified a number of activities to raise awareness and visibility of materials informatics broadly and to foster a materials informatics community.
- Resources: As with any endeavor of this nature, financial and management support greatly help further progress on resolving the issues and needs described. There may also be a need to develop standard reference materials that can be used to support data validation.

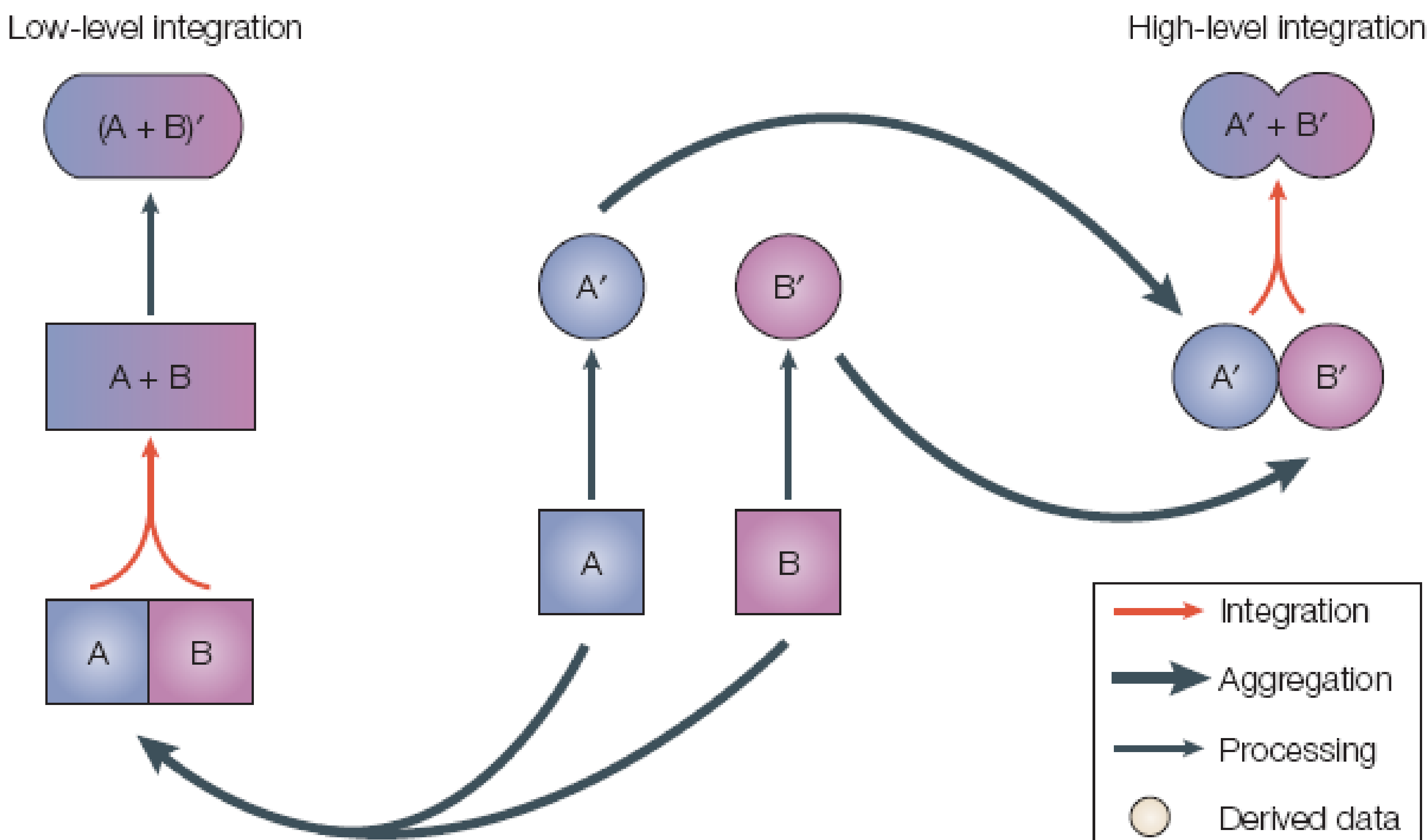
THE “OMICS” OF MATERIALS SCIENCE



Eugene C Butcher, Ellen L Berg & Eric J Kunkel. Systems biology in drug discovery. *Nature Biotechnology* vol 22, 1253-1259 (2004).

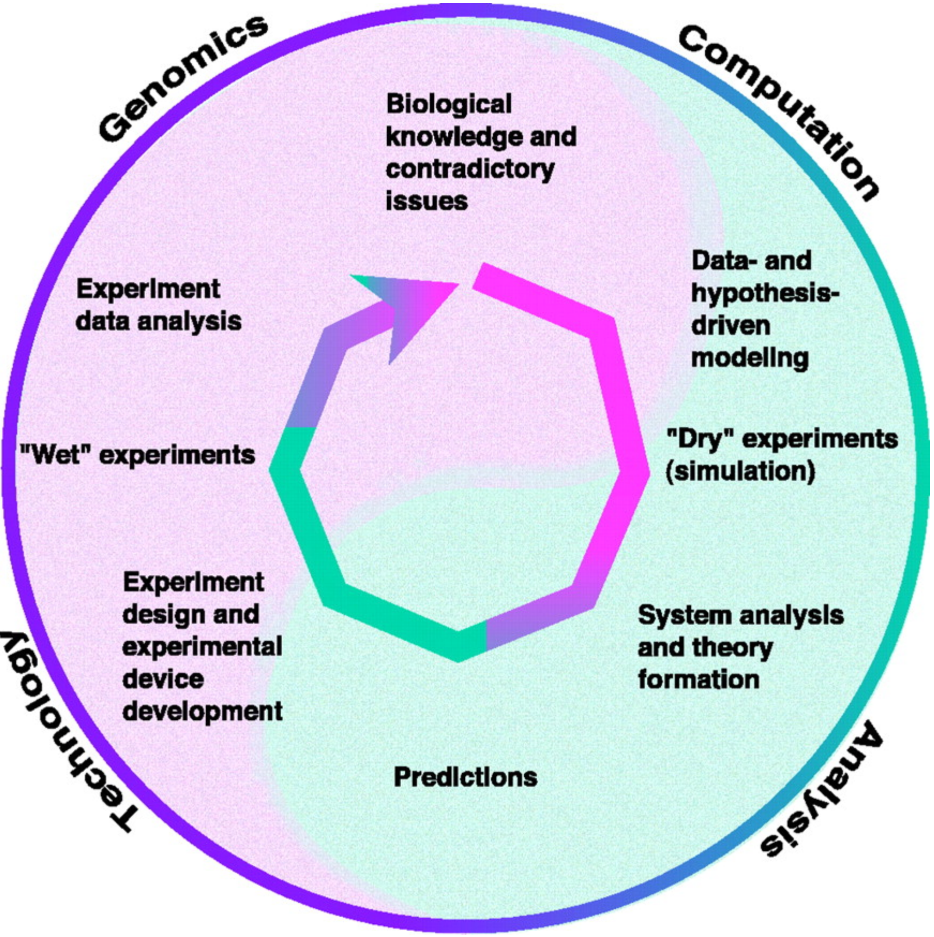
TMS
PROMOTING THE GLOBAL SCIENCE AND ENGINEERING PROFESSIONS CONCERNED WITH MINERALS, METALS, AND MATERIALS

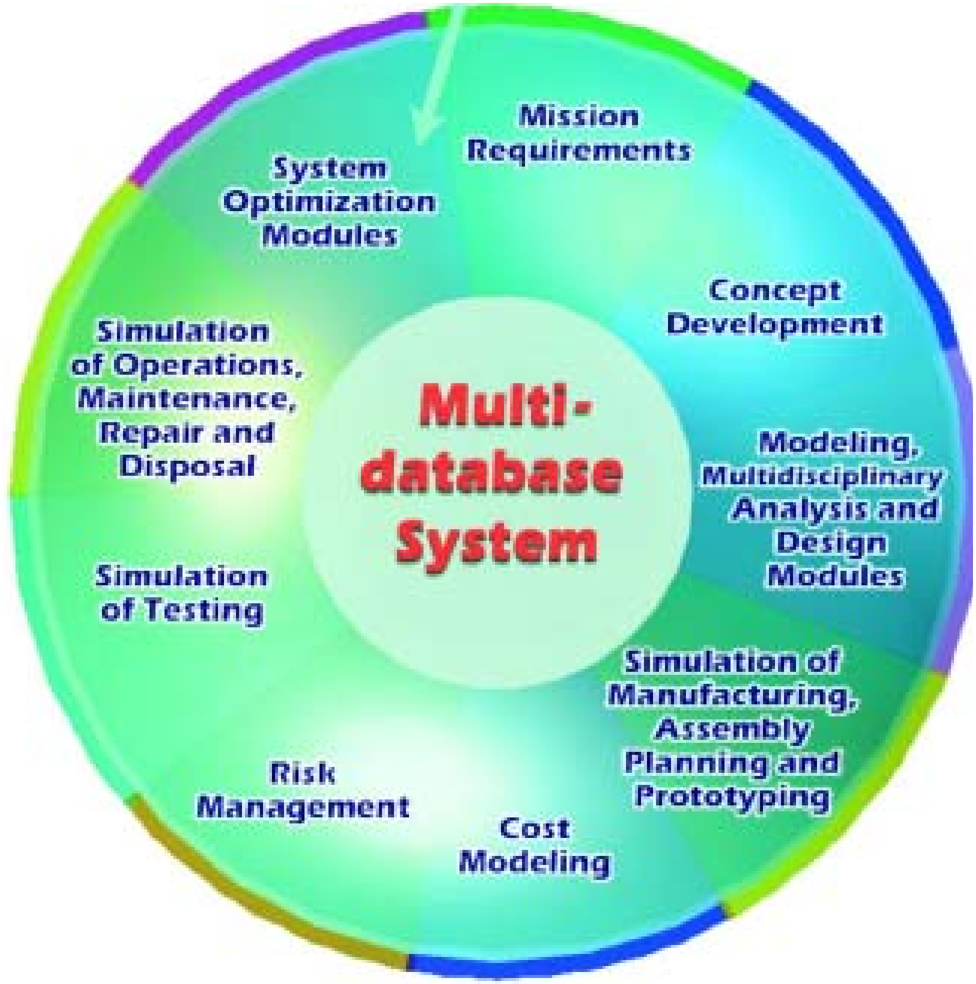
DATA AND DATA INTEGRATION



David B. Searls. Data Integration: Challenges for Drug Discovery.
Nature Reviews Drug Discovery, vol 4, 45-58 (2005).

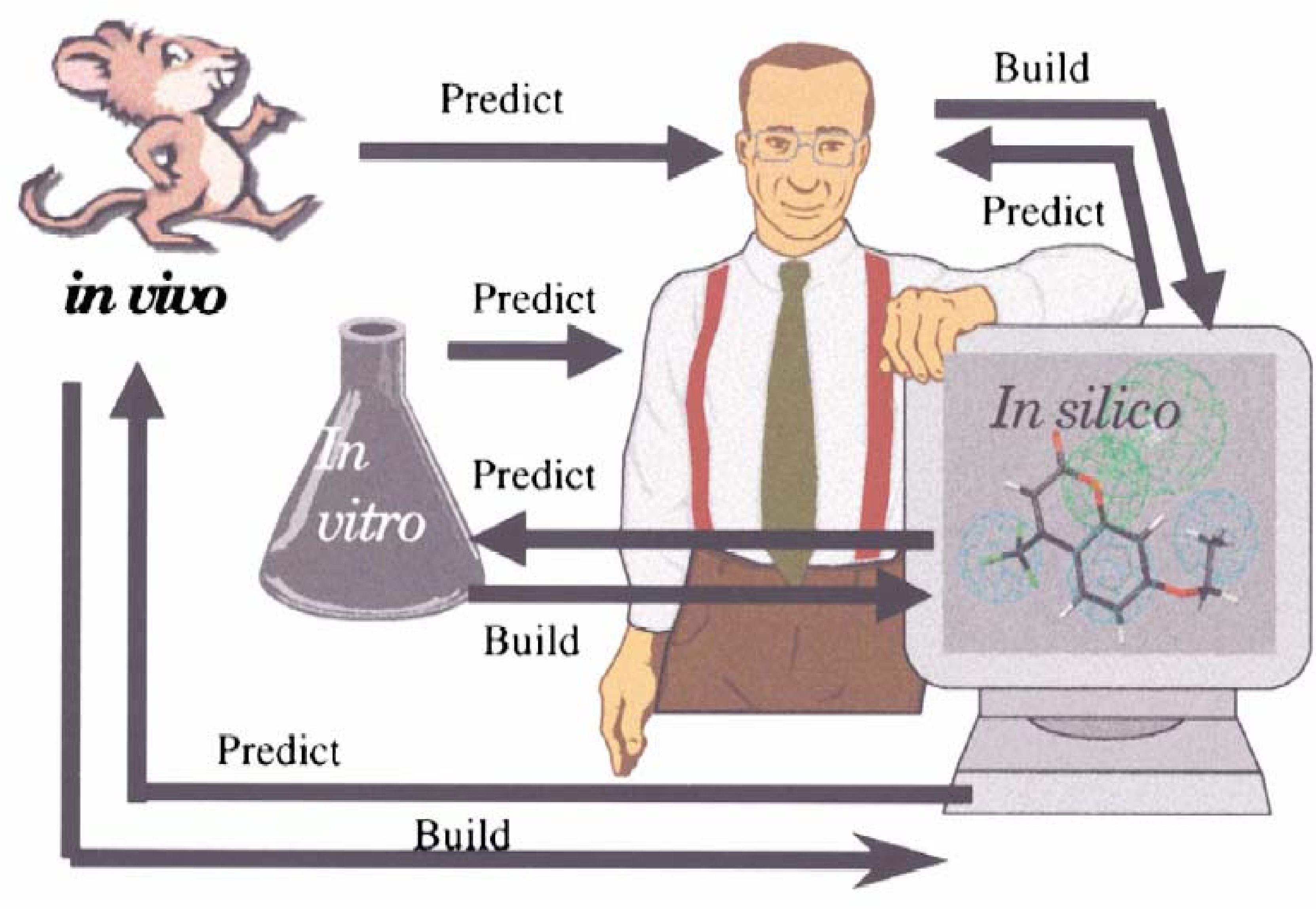
SYSTEMS BIOLOGY & SYSTEMS ENGINEERING





Hiroaki Kitano. Systems Biology: A Brief Overview. *Science,* vol 295. no. 5560, 1662-1664 (2002).
Ahmed K. Noor, Samuel L. Venneri, Donald B. Paul & Mark A. Hopkins. Structures Technology for Future Aerospace Systems. *Computers and Structures,* 74, 507-519 (2000).

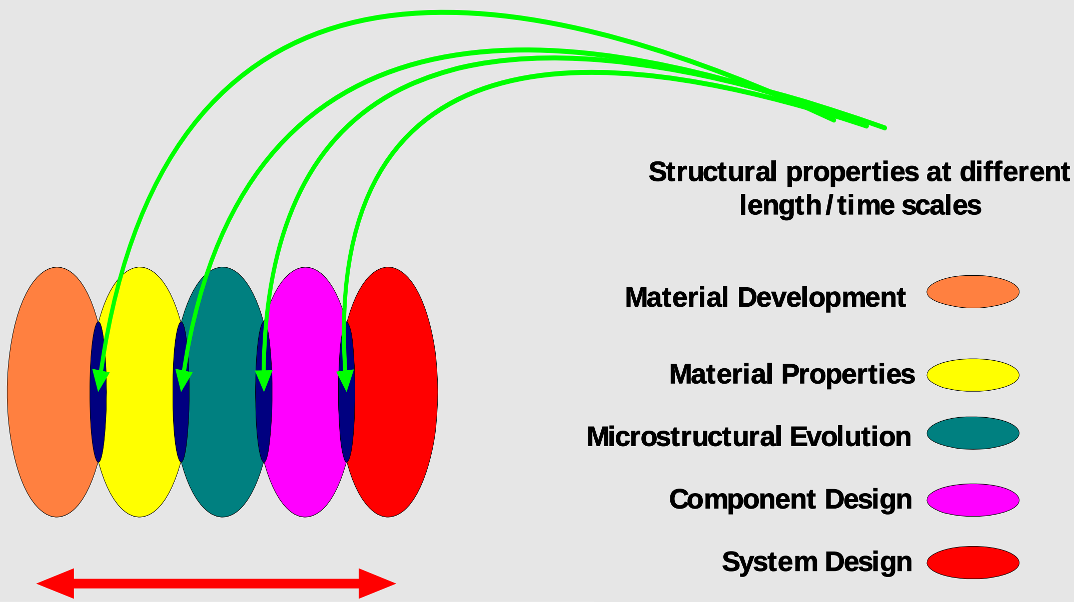
In-silico MATERIALS SCIENCE: A Biological Analogue



Sean Ekins and Steven A. Wrighton. Application of *in silico* approaches to predicting drug–drug interactions. *Journal of Pharmacological and Toxicological Methods,* Vol 45, Issue 1, 65-69 (2001).

TMS
PROMOTING THE GLOBAL SCIENCE AND ENGINEERING PROFESSIONS CONCERNED WITH MINERALS, METALS, AND MATERIALS

INFORMATICS: Tools for Materials Discovery



Data "fusion" through Informatics

Materials Science platform:

- Material theory
- Materials processing
- Property measurement
- Computational techniques
- Materials characterization
- Discrete / atomistic modeling
- Continuum modeling

IT platform :

- Ultra large scale databases of appropriate information and data (data warehousing)
- Scientific visualization
- Data mining
- Remote access / distributed computing

- Data acquisition: The barriers to data acquisition include finding sources of data, locating data within the sources, extracting sufficient metadata to make the property data meaningful, harmonizing the information from different sources, and collecting the results in a fully documented, searchable, and accessible data system. The group identified several means by which the materials community as a whole could contribute to improving data access and acquisition.
- Data standards: Specific needs included a listing and taxonomy of the elements of an information framework to describe materials as well as both de-facto and formalized standards for data formats.
- Dissemination/access: A requirement to define access policies was identified. It was envisaged that different tiers will probably be required to control access to the data depending on the status of who is trying to access it.
- Analysis tools: Knowledge extraction from the fused data sets through the detection of knowledge gaps, identification of anomalies, model development, pattern recognition, and mapping (forward and inverse) techniques is needed. These capabilities need to handle the various forms of information, conform to the likely data standards, and must be joined by a system architecture, as needed by the combinatorial community.
- Visibility/awareness: The practitioners of materials science and related fields need to become educated about informatics approaches, even to the point of developing materials informatics as a recognized new sub-discipline within materials science. The group identified a number of activities to raise awareness and visibility of materials informatics broadly and to foster a materials informatics community.
- Resources: As with any endeavor of this nature, financial and management support greatly help further progress on resolving the issues and needs described. There may also be a need to develop standard reference materials that can be used to support data validation.