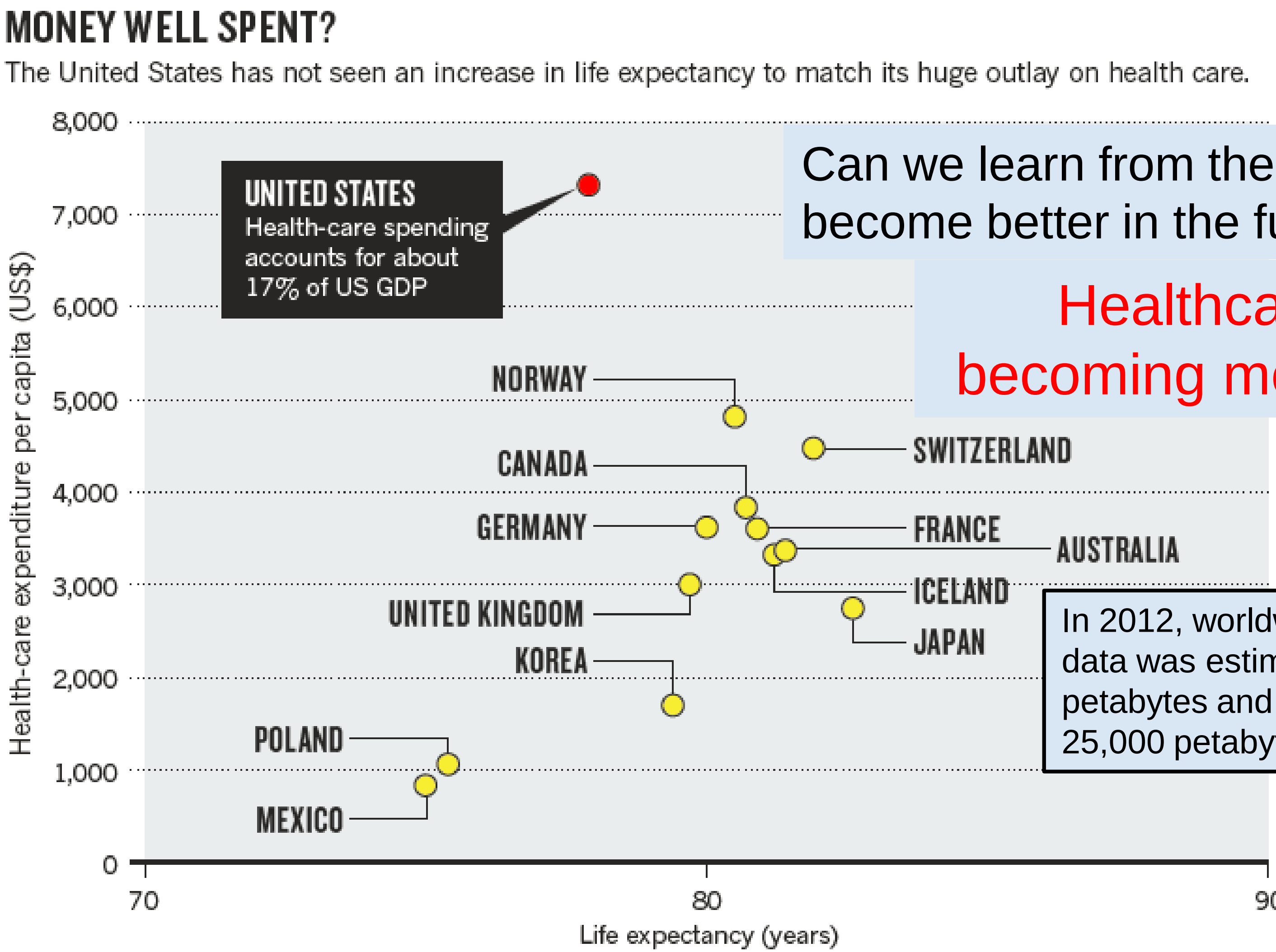
Lecture Notes
Big Data in Medical Informatics

Week 1:
Introduction to Biomedical Informatics: challenges and promises

Oya Beyan, Ph.D.
Prof. Dr. Stefan Decker

Motivation



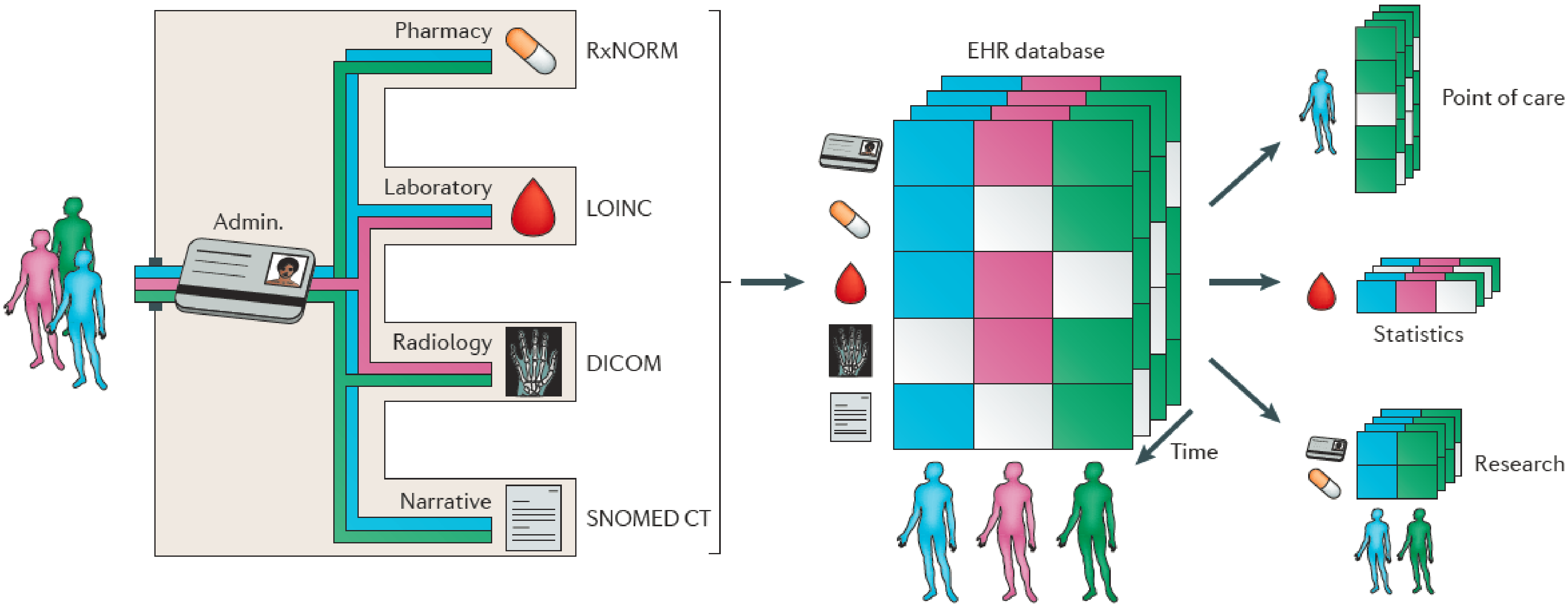
Can we learn from the past to become better in the future ??

Healthcare Data is becoming more complex !!

In 2012, worldwide digital healthcare data was estimated to be equal to 500 petabytes and is expected to reach 25,000 petabytes in 2020.

Hersh, W., Jacko, J. A., Greenes, R., Tan, J., Janies, D., Embi, P. J., & Payne, P. R. (2011). Health-care hit or miss? *Nature*, *470*(7334), 327.

Data Collection and Analysis



Effectively integrating and efficiently analyzing various forms of healthcare data over a period of time can answer many of the impending healthcare problems.

Jensen, Peter B., Lars J. Jensen, and Søren Brunak. "Mining electronic health records: towards better research applications and clinical care." *Nature Reviews Genetics* (2012).

Key Challenges

- Increasing large and complex data sets due to data intensive biomedicine
- Inreasing amounts of non-standardized and unstructured information
- Data quality, data integration and universal access
- Privacy, securtiy, safety and protection issues, fair use of data
- Time aspects in datasets

[Reference: Holzinger, 709.049]

Hospital IS

Doctors surrounded by information technology and final data outcome

G'sund Net. Ausgabe 45. März 2005

angelegt am 06.05.2006/20:26
geschr. von
gedruckt am 17.11.2006/08:24
Anfo: NCHIN

Radiologischer Befund

Kurzanamnese: St.p. SHT

Fragestellung: -

Untersuchung: **Thorax eine Ebene liegend**

SB

Bewegungsartefakte. Zustand nach Schädelhirntrauma.

Das Cor in der Größennorm, keine akuten Stauungszeichen.
Fragliches Infiltrat parahilär li. im UF, RW-Erguss li.

Zustand nach Anlage eines ET, die Spitze ca. 5cm cranial der Bifurkation, lieg. MS, orthotop positioniert. ZVK über re., die Spitze in Proj. auf die VCS. Kein Hinweis auf Pneumothorax. Der re. Rezessus frei.

Mit kollegialen Grüßen

**** Elektronische Freigabe durch am 09.05.2006 ****

[Reference: Holzinger, 709.049]

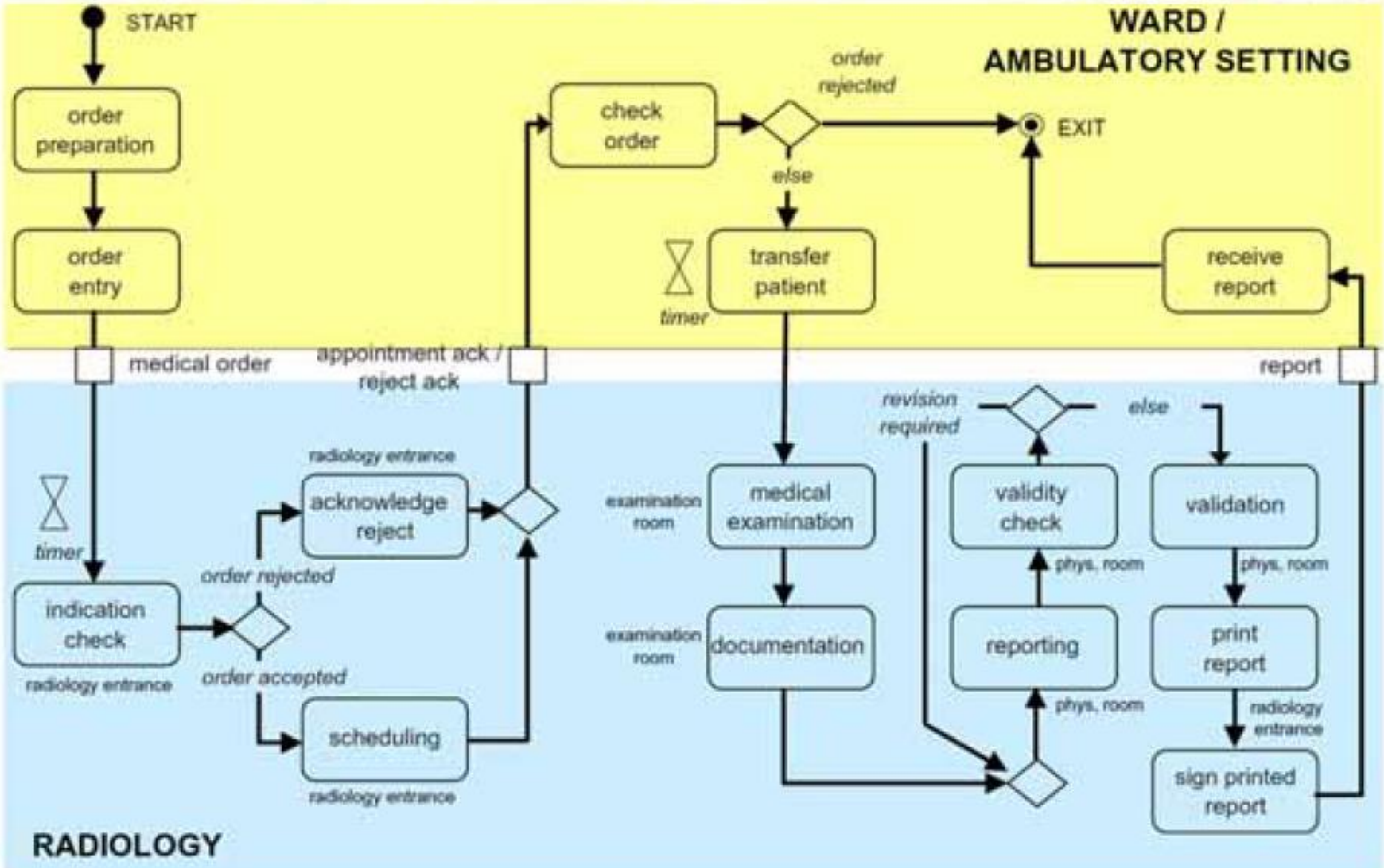
ns

A., Geierhofer, R. & Errath, M. 2007. Semantische Informationsextraktion in
chen Informationssystemen. *Informatik Spektrum, 30, (2), 69-78.*

Clinical workplace: a sample view

G'sund Net, Ausgabe 70, Juni 2011

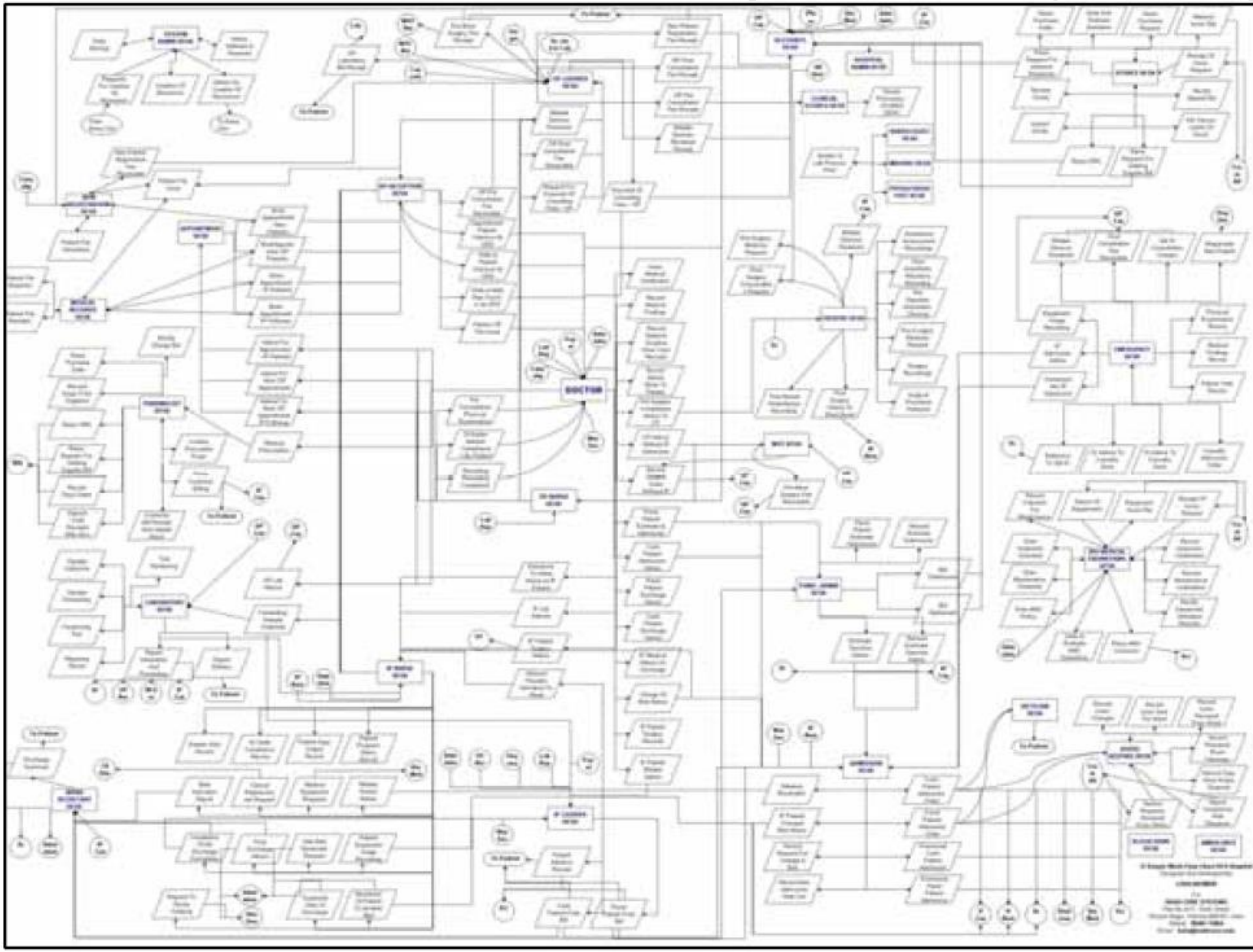
A workflow for medical order entry and result reporting



Lenz, R. & Reichert, M. (2007) IT support for healthcare processes-premises, challenges, perspectives. *Data & Knowledge Engineering, 61, 1, 39-58.*

More complicated workflow

[Reference: Holzinger, 709.049]



Classical Diagnostics-Therapeutic Cycle

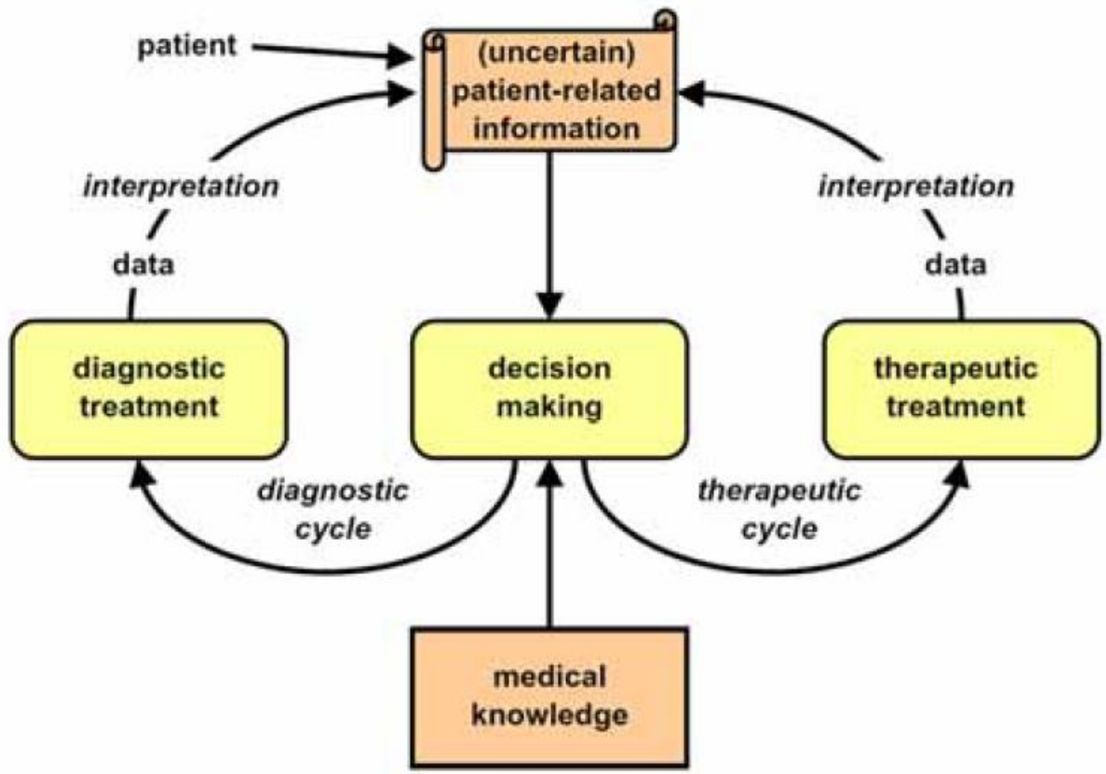
Two bases of decision making:
data and knowledge

Challanges for data:

- Deal with uncertain information
- Partial observation - limited data source integration

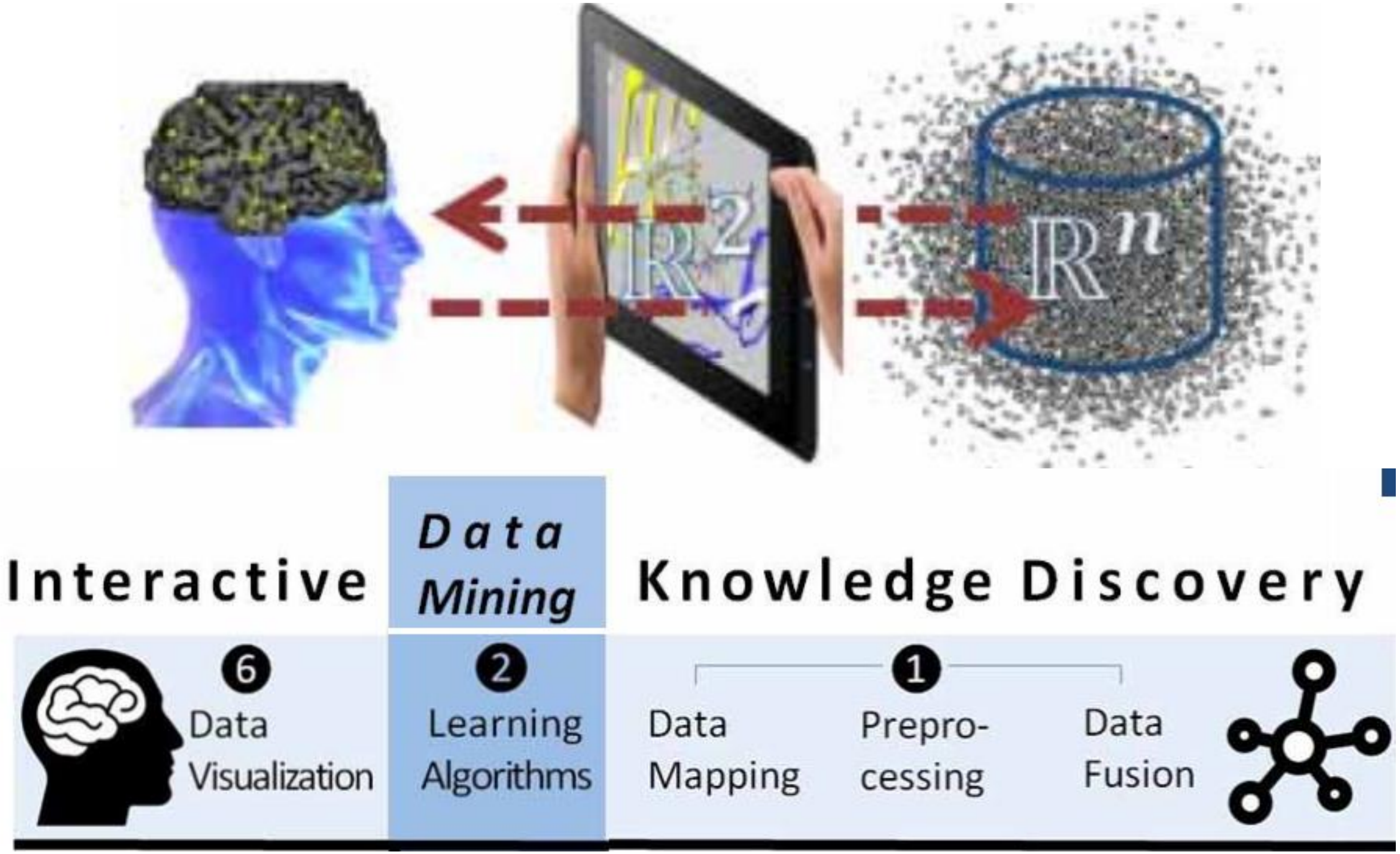
Challanges for Knowledge:

- Access to evidence
- Find right evidence for right person



Lenz, R. & Reichert, M. 2007. IT support for healthcare processes-premises, challenges, perspectives. Data & Knowledge Engineering, 61, (1), 39-58.

Human Computer Interaction



[Reference: Holzinger, 709.049]

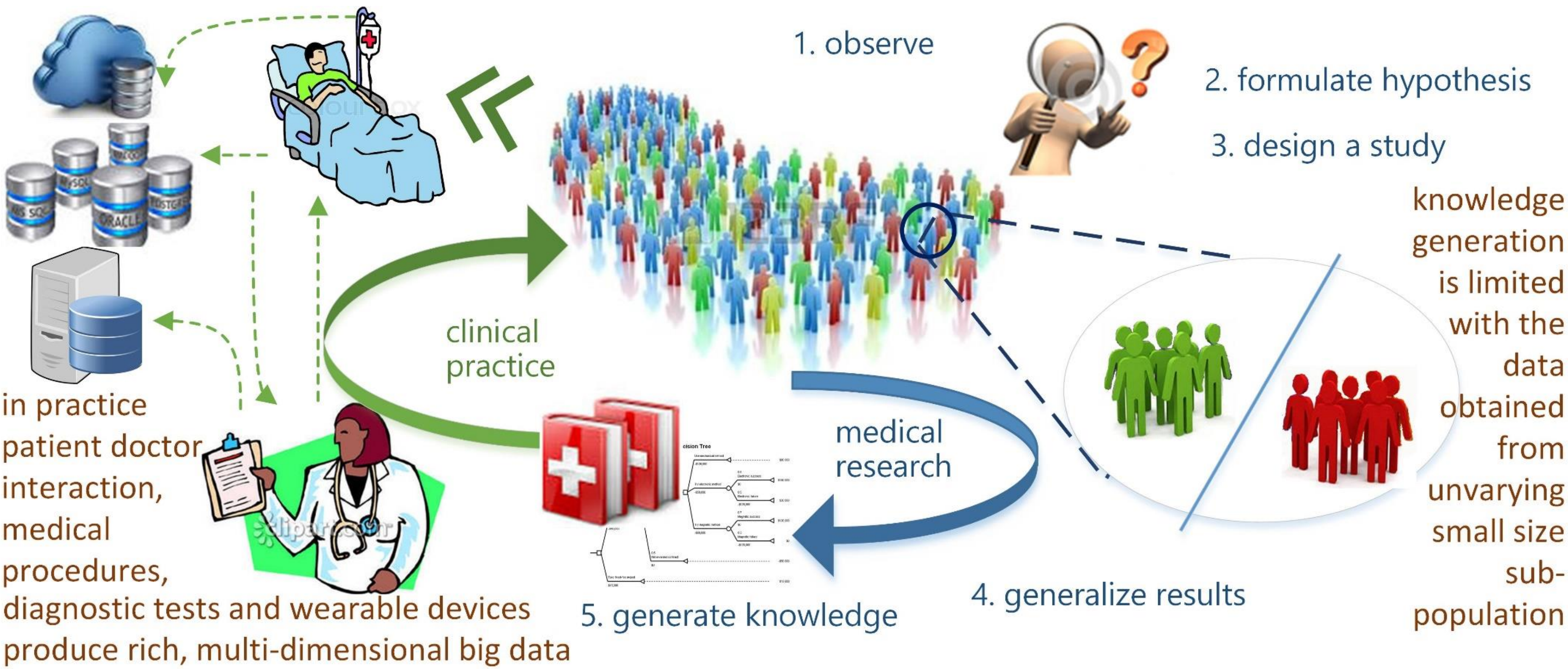
Healthcare Analytics in the Electronic Era

- Old way: **Data are expensive and small**
  - Input data are from clinical trials, which is small and costly
  - Modeling effort is small since the data is limited

- EHR era: **Data are cheap and large**
  - Broader patient population
  - Noisy data
  - Heterogeneous data
  - Diverse scale
  - Longitudinal records

RWTH Informatik 5 | Ahornstr. 55 D-52056 Aachen Germany |
Tel +49/241/8021501 | Fax +49/241/8022321 | http://dbis.rwth-aachen.de/cms

The Gap between knowledge generation and practice



RWTH Informatik 5 | Ahornstr. 55 D-52056 Aachen Germany |
Tel +49/241/8021501 | Fax +49/241/8022321 | http://dbis.rwth-aachen.de/cms

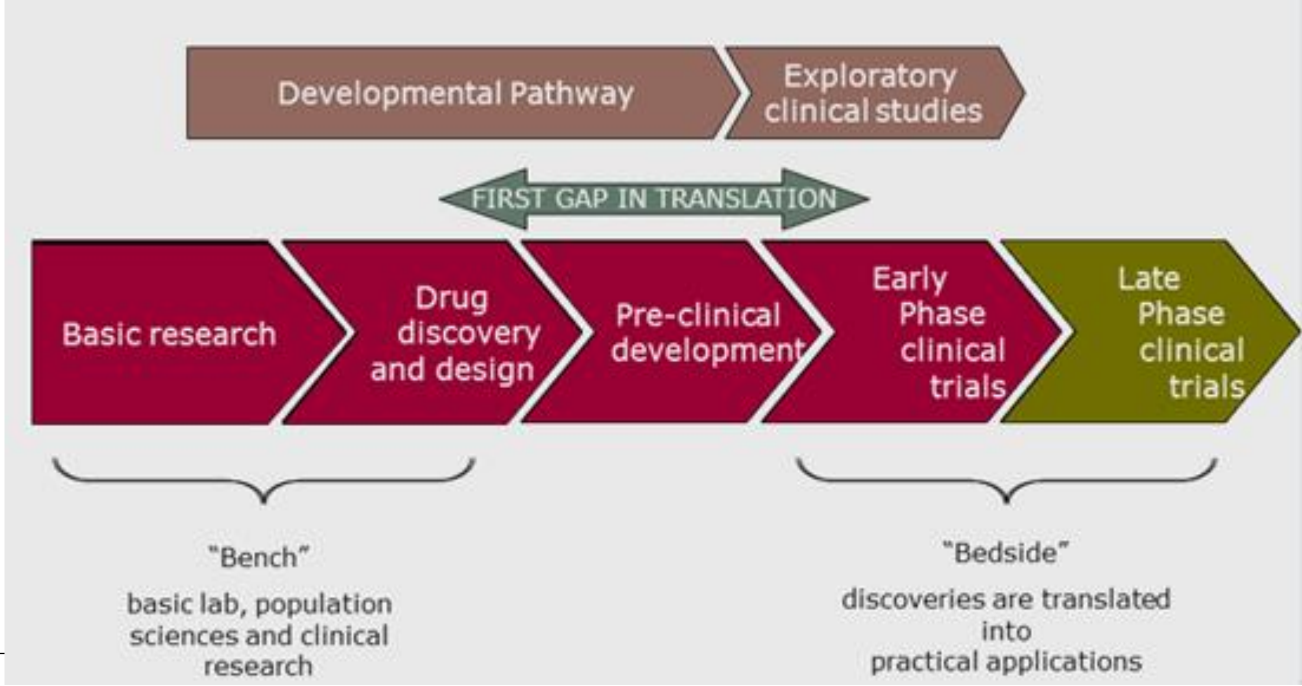
The Gap between knowledge generation and practice

- Two seperated islands: medical science and clinical practice
- **How to bridge the gap?**

  Translational Medicine

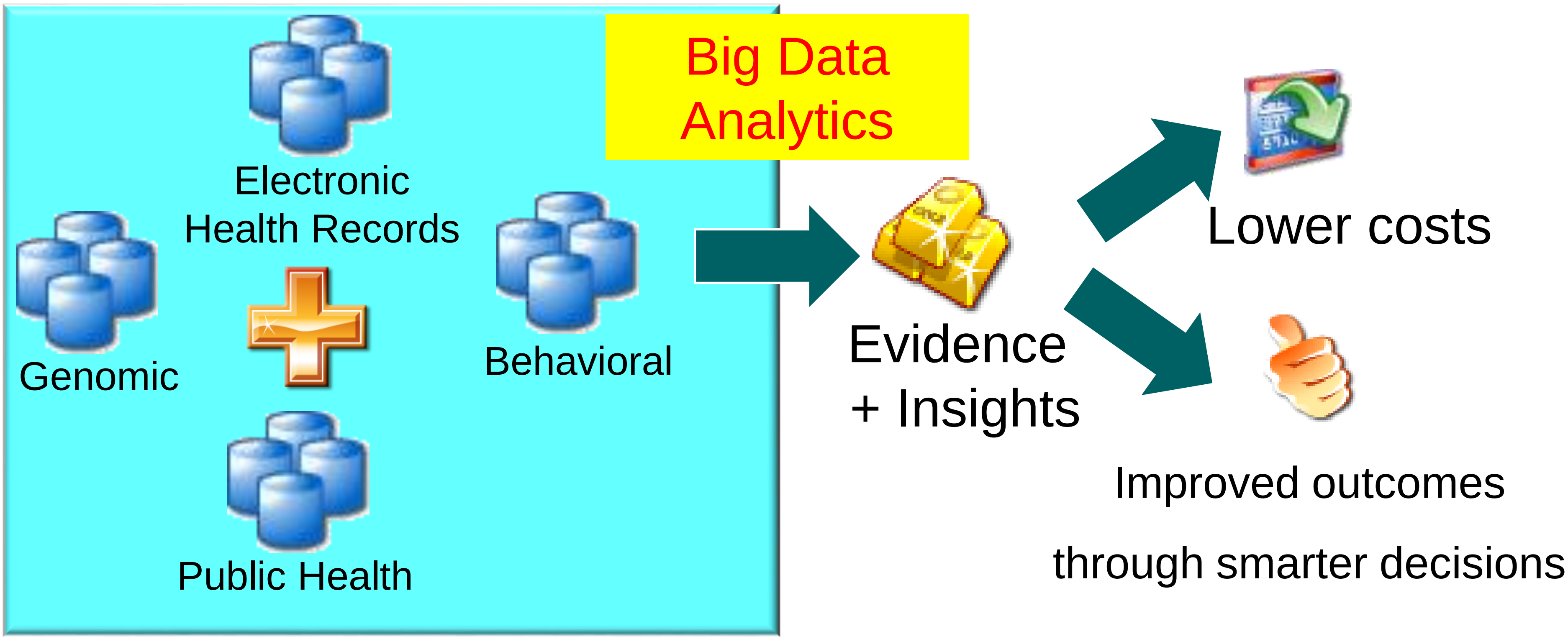
  “Research [that] transforms scientific discoveries arising from laboratory, clinical or population studies into new clinical tools and applications that improve human health by reducing disease incidence, morbidity and mortality.”

  – Translational research fosters the **multidirectional integration** of basic research, patient-oriented research, and population-based research, with the long-term aim of improving the health of the public



RWTH Informatik 5 | Ahornstr. 55 D-52056 Aachen Germany |
Tel +49/241/8021501 | Fax +49/241/8022321 | http://dbis.rwth-aachen.de/cms

Overall Goals of Big Data Analytics in Healthcare



GOAL: Provide Personalized care through right intervention to the right patient at the right time.

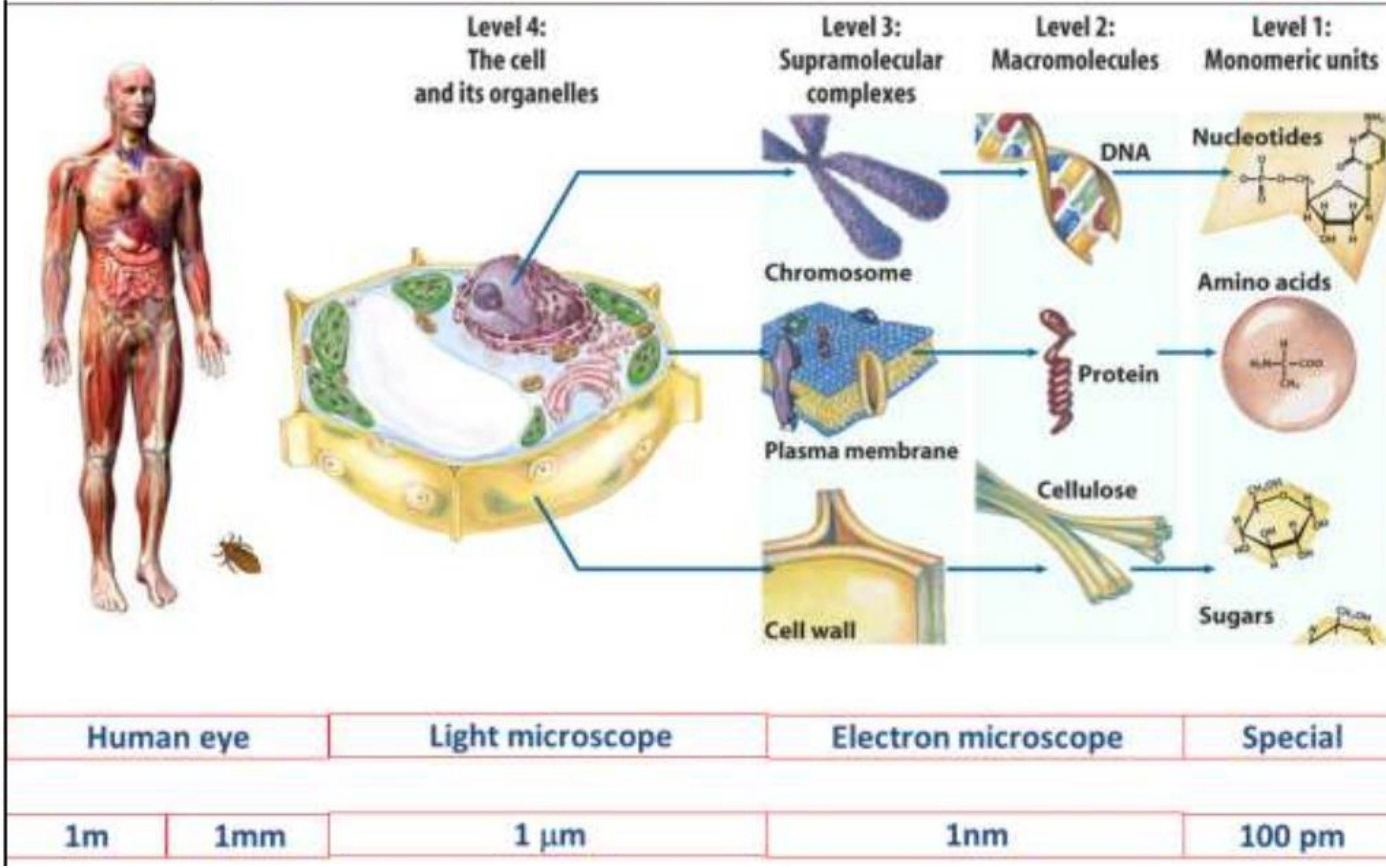
Chandan K. Reddy and Charu C. Aggarwal (editors), "Healthcare Data Analytics", Chapman and Hall/CRC Press, ISBN: 978-1482232110, 2015.

Main Problems of the Health Data

- Heterogeneity : different data sources
- Complexity
  - High Dimensionally
  - Uncertainly
  - Non Standardization
  - Weakly structured data

Building Blocks of Life

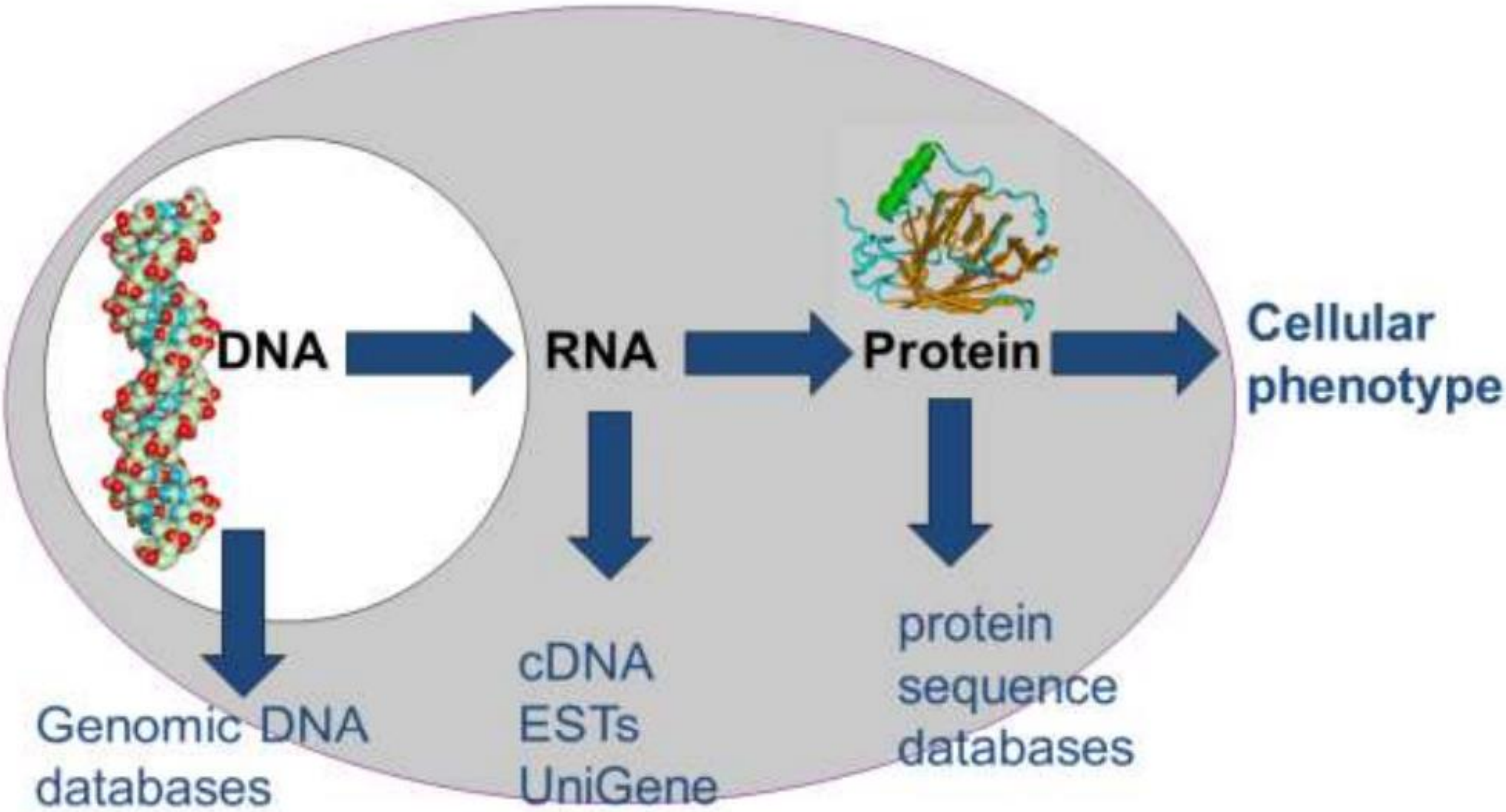
[Reference: Holzinger, 444.152]



RWTH Informatik 5 | Ahornstr. 55 D-52056 Aachen Germany |
Tel +49/241/8021501 | Fax +49/241/8022321 | http://dbis.rwth-aachen.de/cms

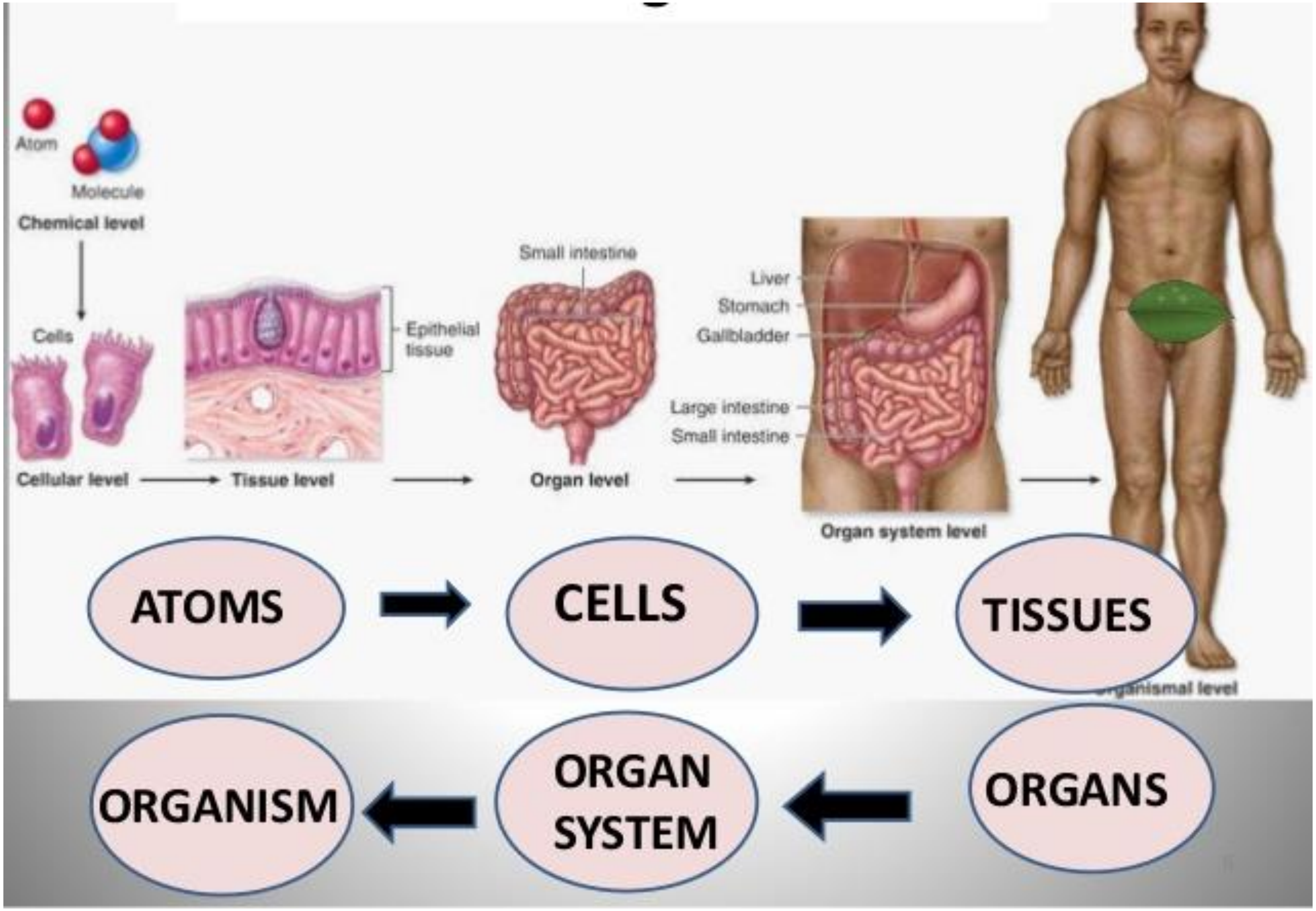
The Dogma of Molecular Biology

[Reference: Holzinger, 444.152]



Crick, F. 1970. Central Dogma of Molecular Biology. *Nature, 227, (5258), 561-563.*

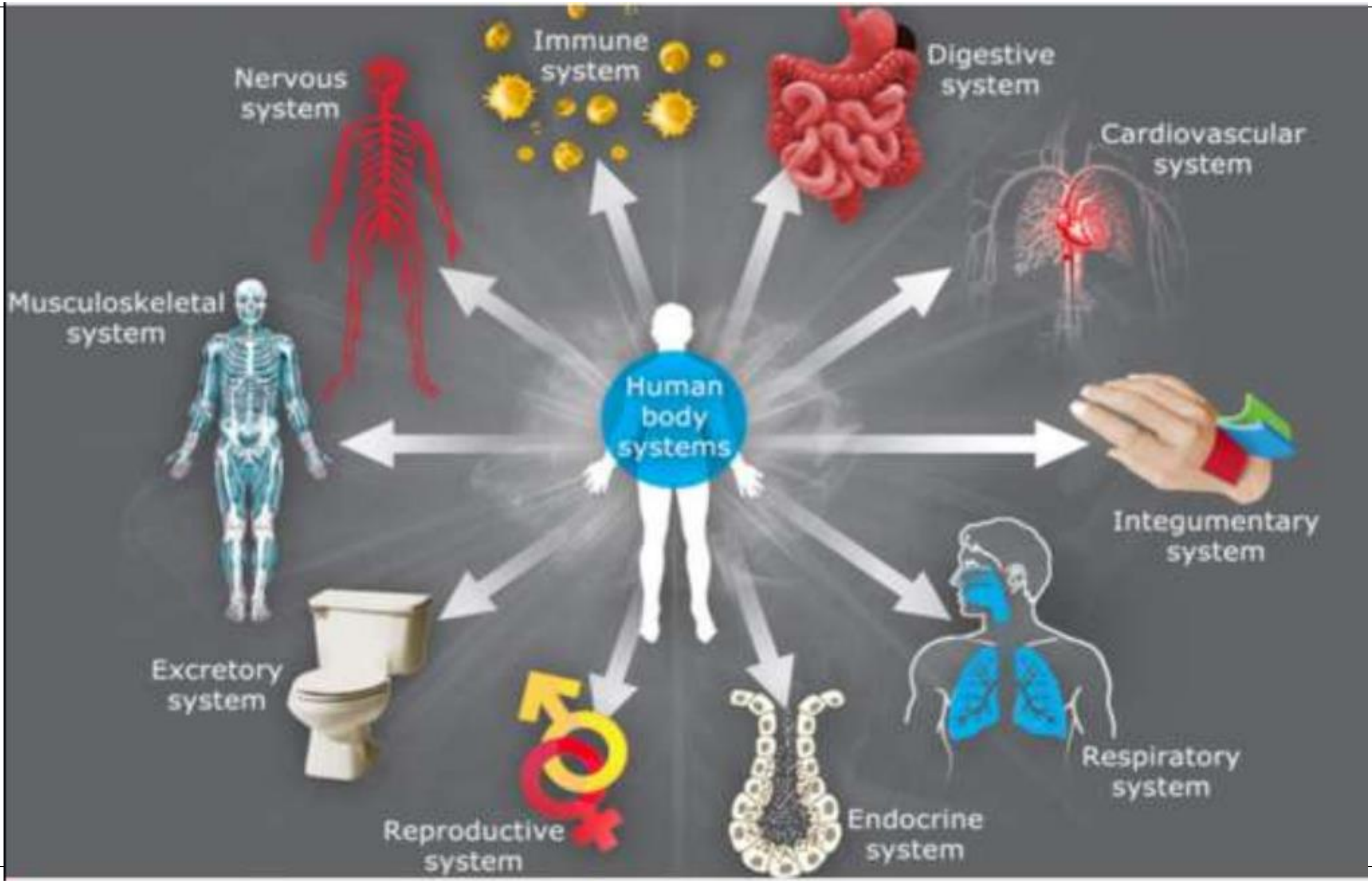
Levels of Organization



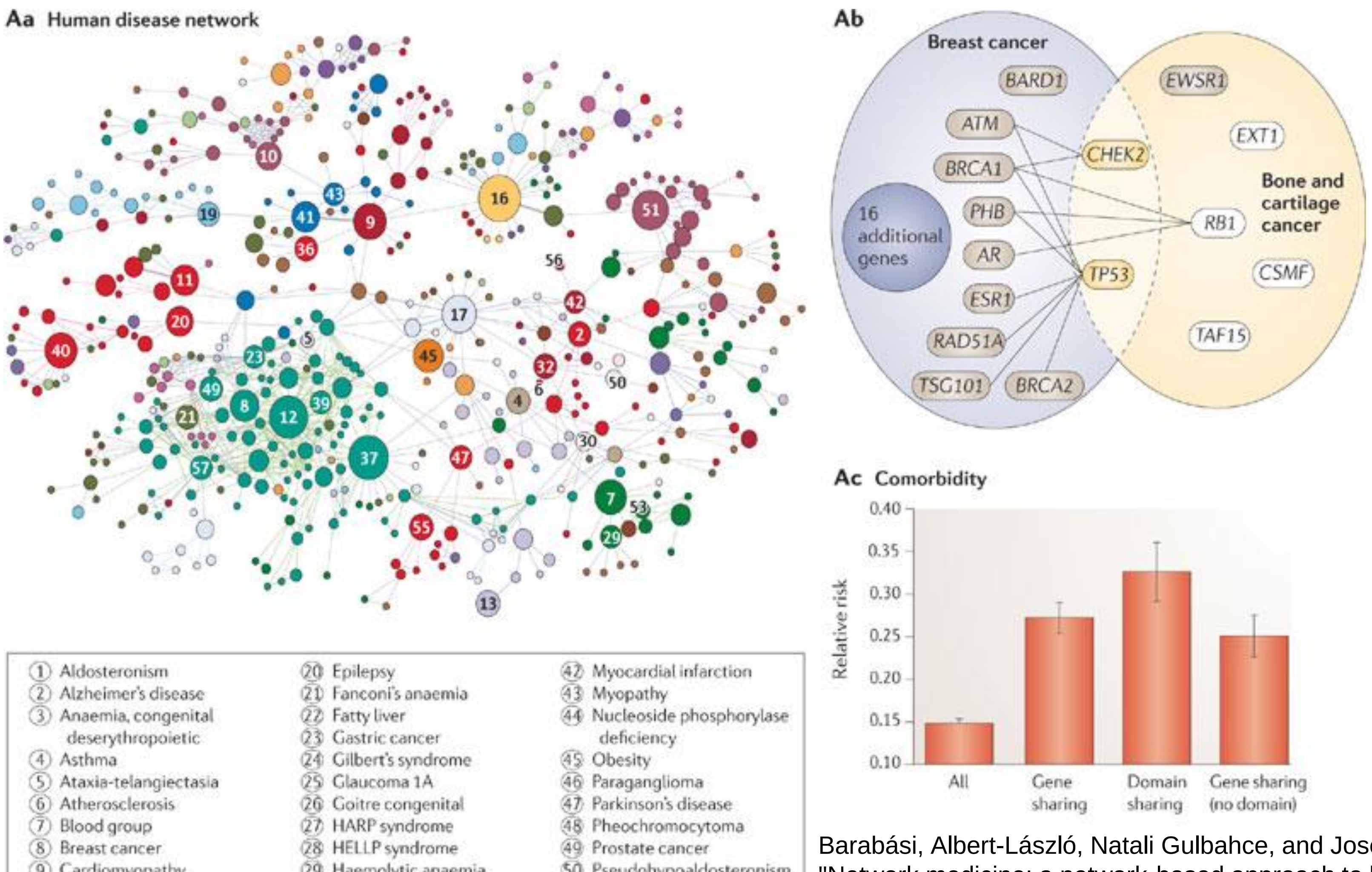
RWTH Informatik 5 | Ahornstr. 55 D-52056 Aachen Germany |
Tel +49/241/8021501 | Fax +49/241/8022321 | http://dbis.rwth-aachen.de/cms

Human Organ Systems

[Reference: Holzinger, 444.152]

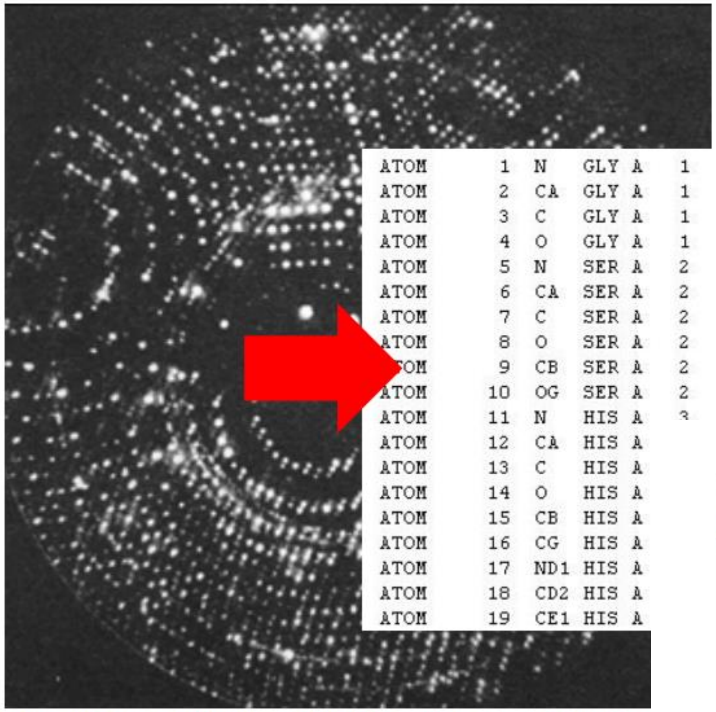


Complex intracellular and intercellular network that links tissue and organs



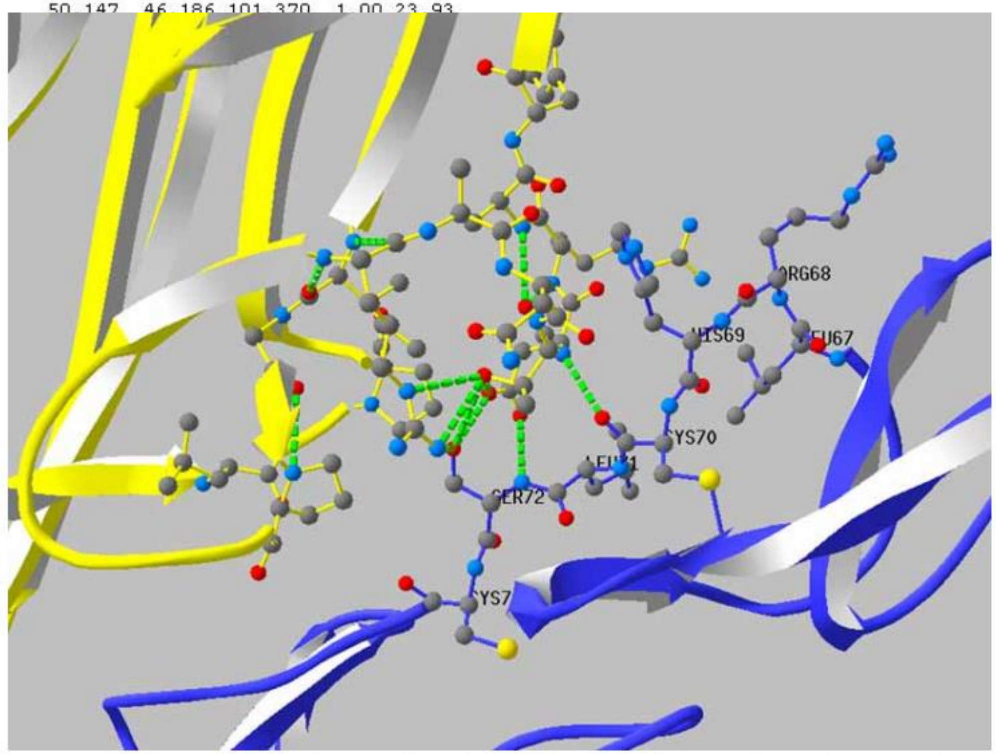
Barabási, Albert-László, Natali Gulbahce, and Joseph Loscalzo. "Network medicine: a network-based approach to human disease." Nature Reviews Genetics 12.1 (2011): 56-68.

Data Visualization



ATOM	1	N	GLY	A	1	44.842	51.034	101.284	0.01	27.20
ATOM	2	CA	GLY	A	1	45.640	50.230	100.389	0.01	26.99
ATOM	3	C	GLY	A	1	46.692	49.648	101.308	0.01	26.80
ATOM	4	O	GLY	A	1	46.895	50.222	102.381	0.01	26.91
ATOM	5	N	SER	A	2	47.283	48.516	100.951	1.00	26.26
ATOM	6	CA	SER	A	2	48.277	47.866	101.761	1.00	26.17
ATOM	7	C	SER	A	2	49.212	47.031	100.845	1.00	24.21
ATOM	8	O	SER	A	2	49.060	47.195	99.630	1.00	19.77
ATOM	9	CB	SER	A	2	47.438	47.091	102.800	1.00	26.31
ATOM	10	OG	SER	A	2	46.276	46.356	102.404	1.00	27.99
ATOM	11	N	HIS	A	3	50.147	46.186	101.370	1.00	23.93
ATOM	12	CA	HIS	A						
ATOM	13	C	HIS	A						
ATOM	14	O	HIS	A						
ATOM	15	CB	HIS	A						
ATOM	16	CG	HIS	A						
ATOM	17	ND1	HIS	A						
ATOM	18	CD2	HIS	A						
ATOM	19	CE1	HIS	A						

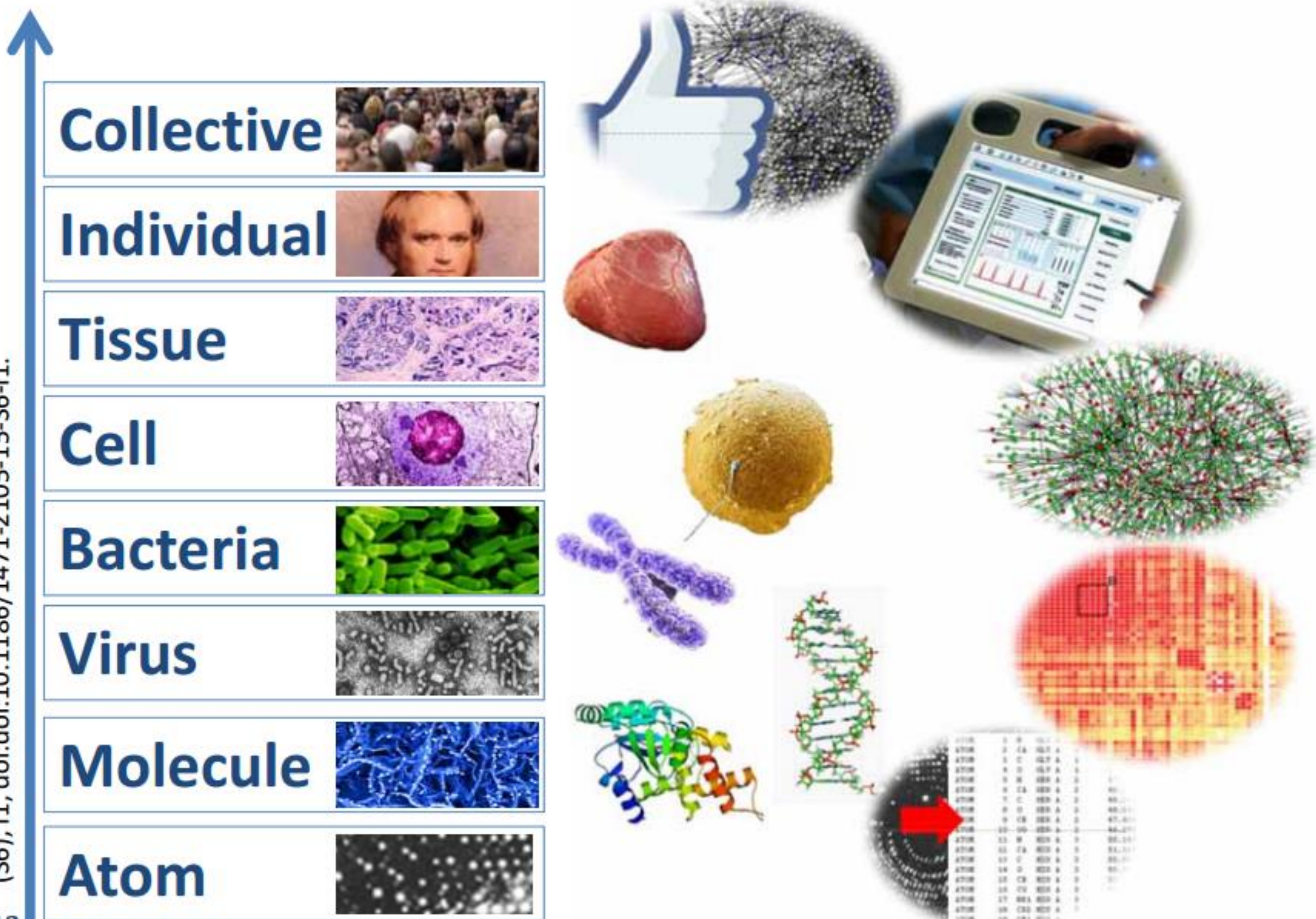
Wiltgen, M. & Holzinger, A. (2005) Visualization in Bioinfor
and Biological Annotations. In: *Central European Multimec*
Technical University (CTU), 69-74



Wiltgen, M., Holzinger, A. & Tilz, G. P. (2007) Interactive Analysis and Visualization of Macromolecular Interfaces Between Proteins. In: *Lecture Notes in Computer Science (LNCS 4799). Berlin, Heidelberg, New York, Springer, 199-212.*

[Reference: Holzinger, 709.049]

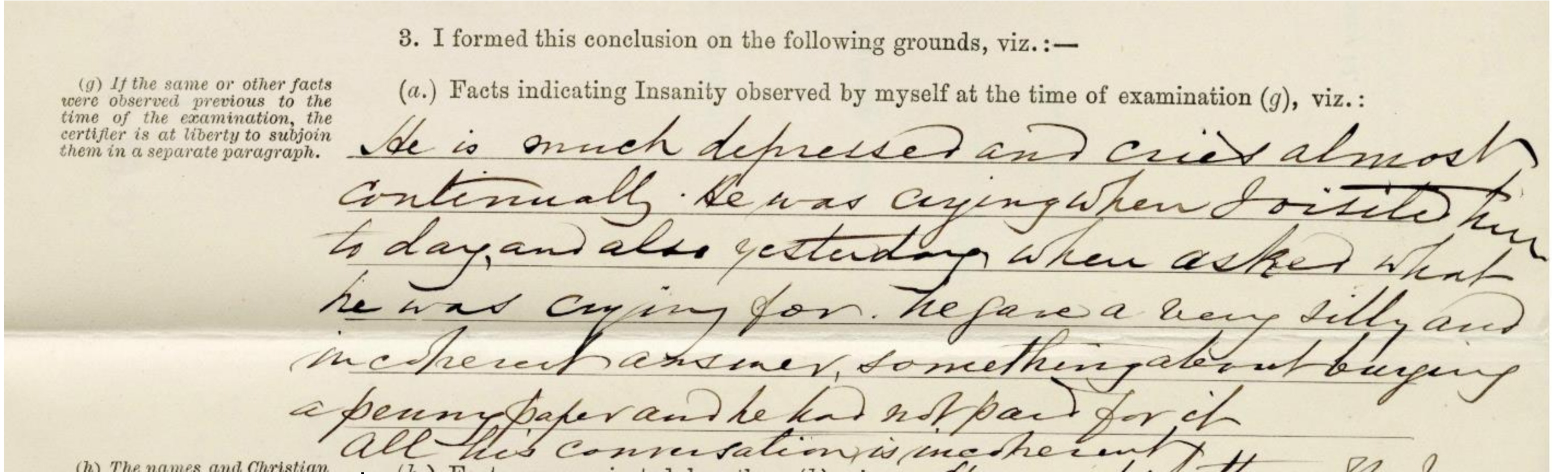
Levels of Abstraction



Holzinger, A., Dehmer, M. & Jurisica, I. 2014. Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions. BMC Bioinformatics, 15, (S6), I1, doi:doi:10.1186/1471-2105-15-S6-I1.

Data Management in the Clinic

- Why the paper-based medical record is not adequate for meeting the needs of modern medicine?



3. I formed this conclusion on the following grounds, viz.:—

(g) If the same or other facts were observed previous to the time of the examination, the certifier is at liberty to subjoin them in a separate paragraph.

(a.) Facts indicating Insanity observed by myself at the time of examination (g), viz.:

He is much depressed and cries almost continually. He was crying when I visited him to day and also yesterday, when asked what he was crying for. he gave a very silly and incoherent answer, something about buying a penny paper and he had not paid for it. All this conversation is incoherent

(h) The names and Christian

- It arose in the 19th century as a highly personalized “lab notebook” that clinicians could use to record their observations and plans so that they could be reminded of pertinent details when they next saw the same patient.
  - No bureaucratic requirements,
  - no assumptions that the record would be used to support communication among varied providers of care,
  - few data or test results to fill up the record’s pages,
  - Clinician centric: no other parties are involved

RWTH Informatik 5 | Ahornstr. 55 D-52056 Aachen Germany |
Tel +49/241/8021501 | Fax +49/241/8022321 | http://dbis.rwth-aachen.de/cms

Data Management in the Clinic

Todays problem: complexity

- analyze the ***processes*** associated with the **creation and use of records** rather
- **Do not think** the record as an **object** that can be moved around as needed within the institution.

- The key idea: **evolving clinical workstation**

Data Management in the Clinic

Medical record in a new incarnation:

- *electronic,*
- *accessible,*
- *confidential,*
- *secure,*
- *acceptable to clinicians and patients, and*
- *integrated with other types of non patient-specific information.*

[redacted]
Diagnosis: Benign Neoplasm of P. Fn. tdy

CARDIO-NUCLEAR DIAGNOSTIC CENTER
X-RAY • CT SCAN • ULTRASOUND • BONE DENSITY•MRI

(e)

Patient History Questionnaire (MRI)

Patient Name: [redacted]
Procedure: MRI BRAIN C E S

Reason for Procedure:

Please circle any of the following symptoms that you are experiencing:
Headache Weakness Seizures Vision Problems Fever Blackouts Difficulty Thinking Difficulty Remembering
Difficulty Calculating Nausea Hearing Loss Ringing In Ears Dizziness Speech Problems Walking Problems
Pain Numbness Bowel Changes Bladder Changes Clicking Noises Limited Range Of Motion Swelling
Joint Locking Joint Giving Way

When did your symptoms start? Don't remember Date of Injury / Accident
Where do you have pain? Back & knees
How were you injured? long time ago

Medical History:

Please circle any of the following you have or have had:
Cancer Heart Disease Kidney/Renal Disease Multiple Myeloma Hypertension Sickle Cell Anemia Diabetes Stroke
TIA Asthma Emphysema Bronchitis Tumor, Lump, Mass Radiation Therapy Fracture / Broken Bone Arthritis

Yes No Have you had any tests (US, CT, X-ray, etc.) performed for the symptoms you are experiencing? If yes, please list when and where you had the test. Here every 3 months

Yes No Have you had surgery on your eyes or ears?
Yes No Have you had surgery on your brain? Pitutary removal
Yes No Have you had surgery on your spine?
Yes No Have you had any cardiac surgery?
Yes No Have you had any vascular surgery?
Yes No Have you had any joint replacement surgery?
Yes No Have you had any abdominal surgery?
Yes No Have you had any pelvic surgery?
Yes No Have you had any other surgery? gall bladder
Yes No Do you have any allergies (e.g., medications, latex, food, etc.)? If yes, please list all allergies.
Ultram, Penicillin, Ampicillin, Amphicillin, cephalosporin
Phenegan

I hereby certify that the above information is true and correct to the best of my knowledge.

[redacted] [redacted] 9-26-11
Patient or Legal Representative Signature Print Name and Authority (if legal representative) Date

Technologists Notes: follow up

Data Management in the Clinic

- Why it is very difficult to automate medical records?
- Why applying a computer to a physical problem in a medical context is no different from doing so in a physics laboratory or for an engineering application?

problem of abstraction level

- The use of computers in
  - **low-level processes**: e.g. physics or chemistry – the solvent properties of water
  - **higher-level processes**: carried out in more complex objects such as organisms,
    - record the properties or behavior of human beings, we are using the descriptions of very high-level objects

- Common solution **high abstraction** : events or processes have been reduced to low-level objects, e.g. ATM machine
- In biomedicine, **abstractions carried to high degree would be worthless** from either a clinical or research perspective.

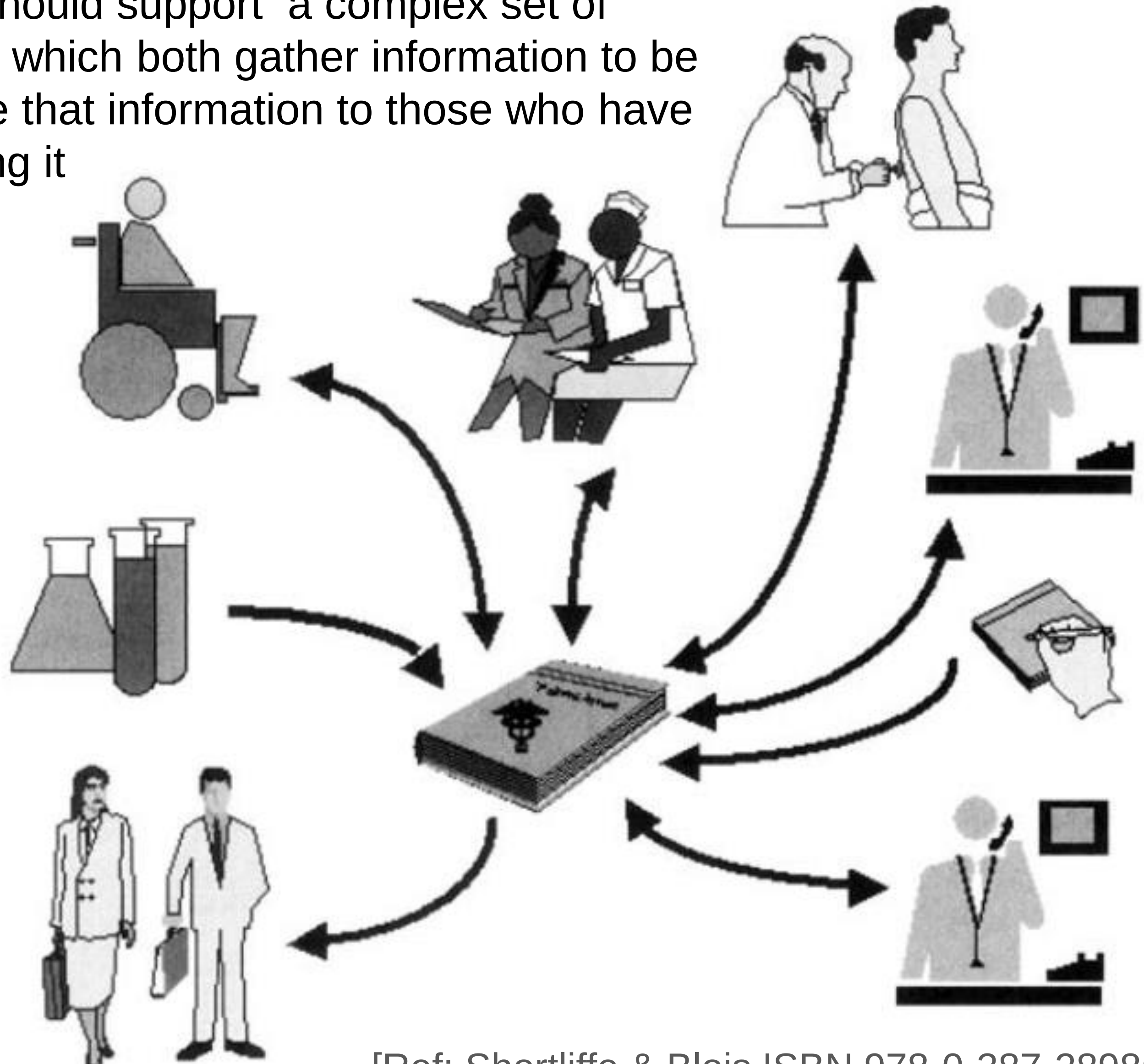
Patient records as a process

Integrated clinical workstations : **single-entry points** into a medical world

- computational tools assist clinical matters
  - reporting results of tests, allowing direct entry of orders by clinicians, facilitating access to transcribed reports,
- supporting telemedicine applications
- decision-support functions
- supporting administrative and financial topics
  - tracking of patients within the hospital, managing materials and inventory, supporting personnel functions, and managing the payroll
- research
  - analyzing the outcomes associated with treatments and procedures,
  - performing quality assurance,
  - supporting clinical trials,
- scholarly information
  - accessing digital libraries,
  - supporting bibliographic search,
  - providing access to drug information databases
- office automation (e.g., providing access to spreadsheets, word processors).

Patient records as a process

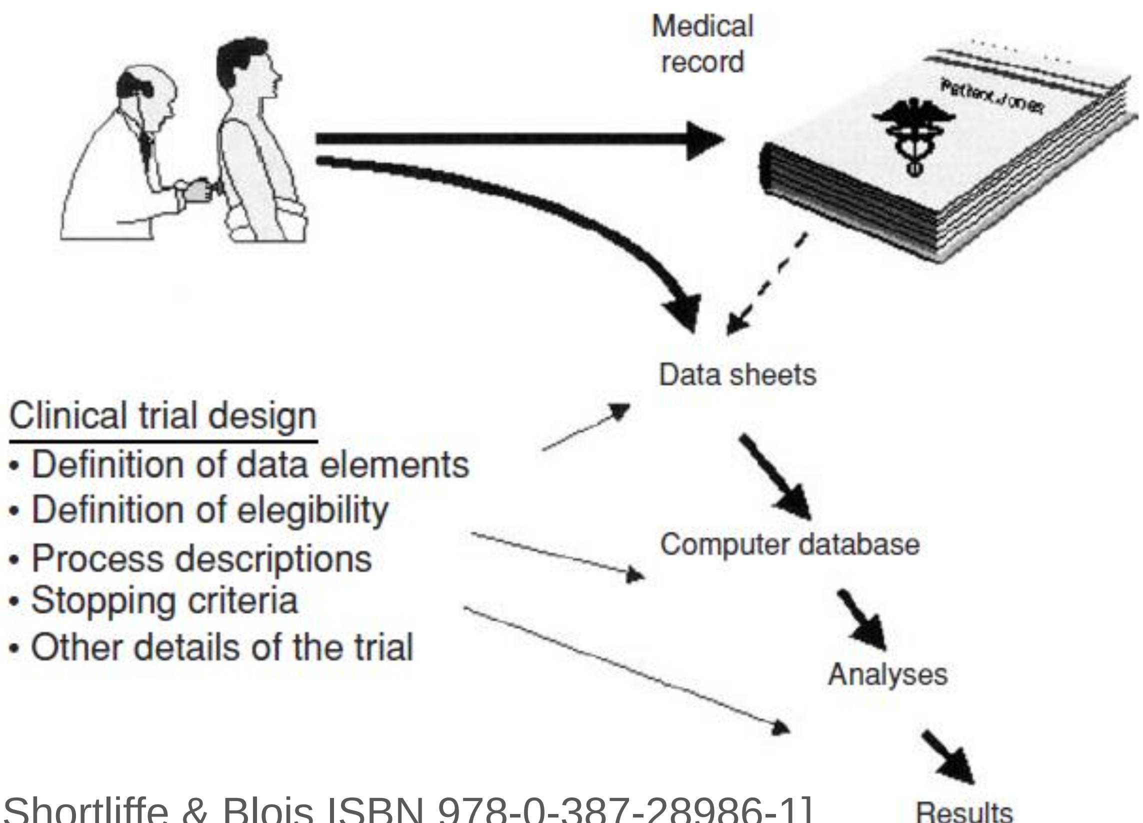
Todays medical records should support a complex set of organizational processes, which both gather information to be shared and then distribute that information to those who have valid reasons for accessing it



[Ref: Shortliffe & Blois ISBN 978-0-387-28986-1]

The Medical Record and Clinical Trials

- **clinical trials**: experiments in which data from specific patient interactions are pooled and analyzed in order to learn about the safety and efficacy of new treatments or tests and to gain insight into disease processes that are not otherwise well understood



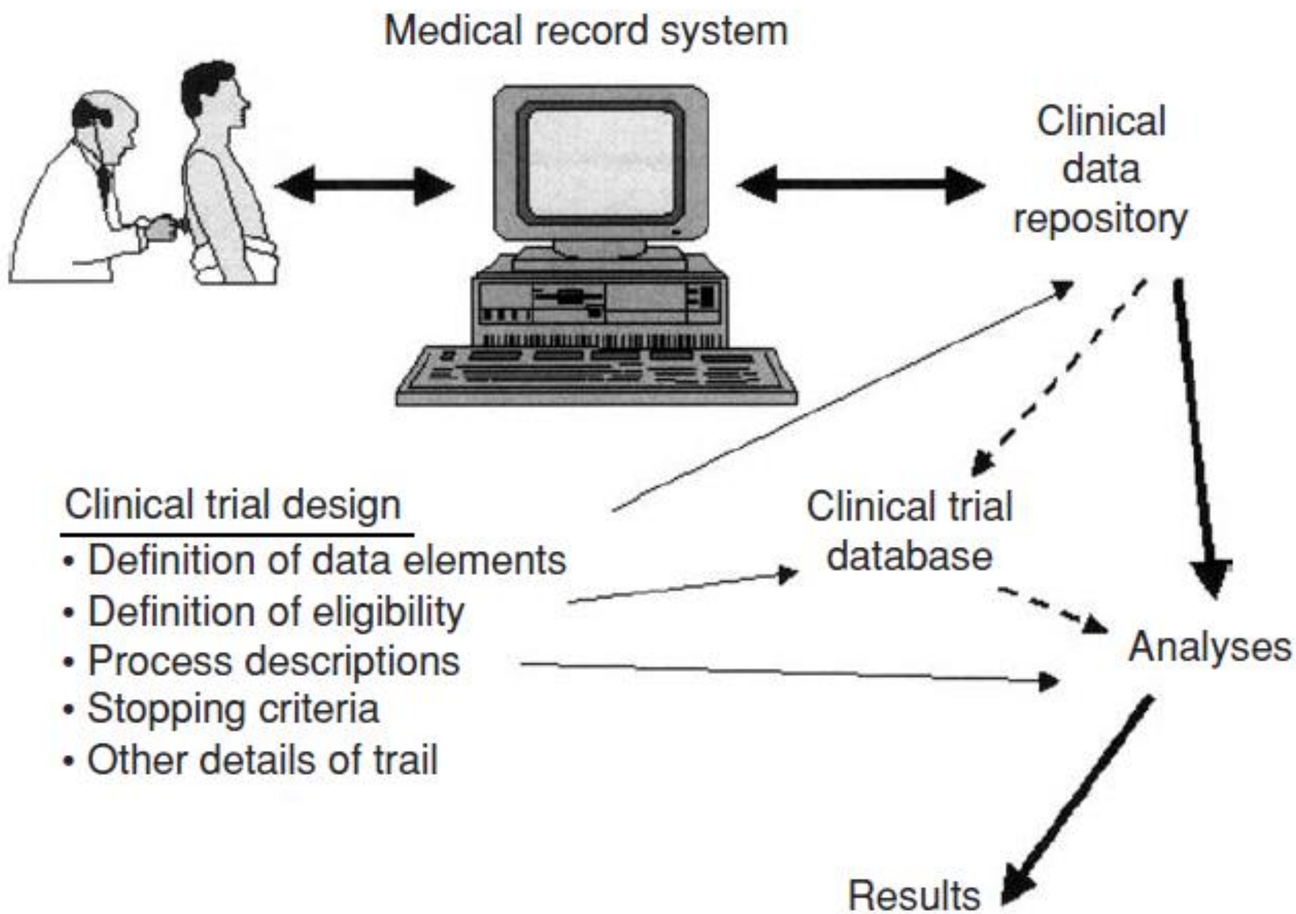
[Ref: Shortliffe & Blois ISBN 978-0-387-28986-1]

The Medical Record and Clinical Trials

- clumsy methods for data acquisition
  - manual capture of information onto datasheets
  - later transcribed into computer databases for statistical analysis
- Problem:
  - labor-intensive,
  - fraught with opportunities for error, and
  - adds to the high costs associated with randomized prospective research protocols.

The Medical Record and Clinical Trials

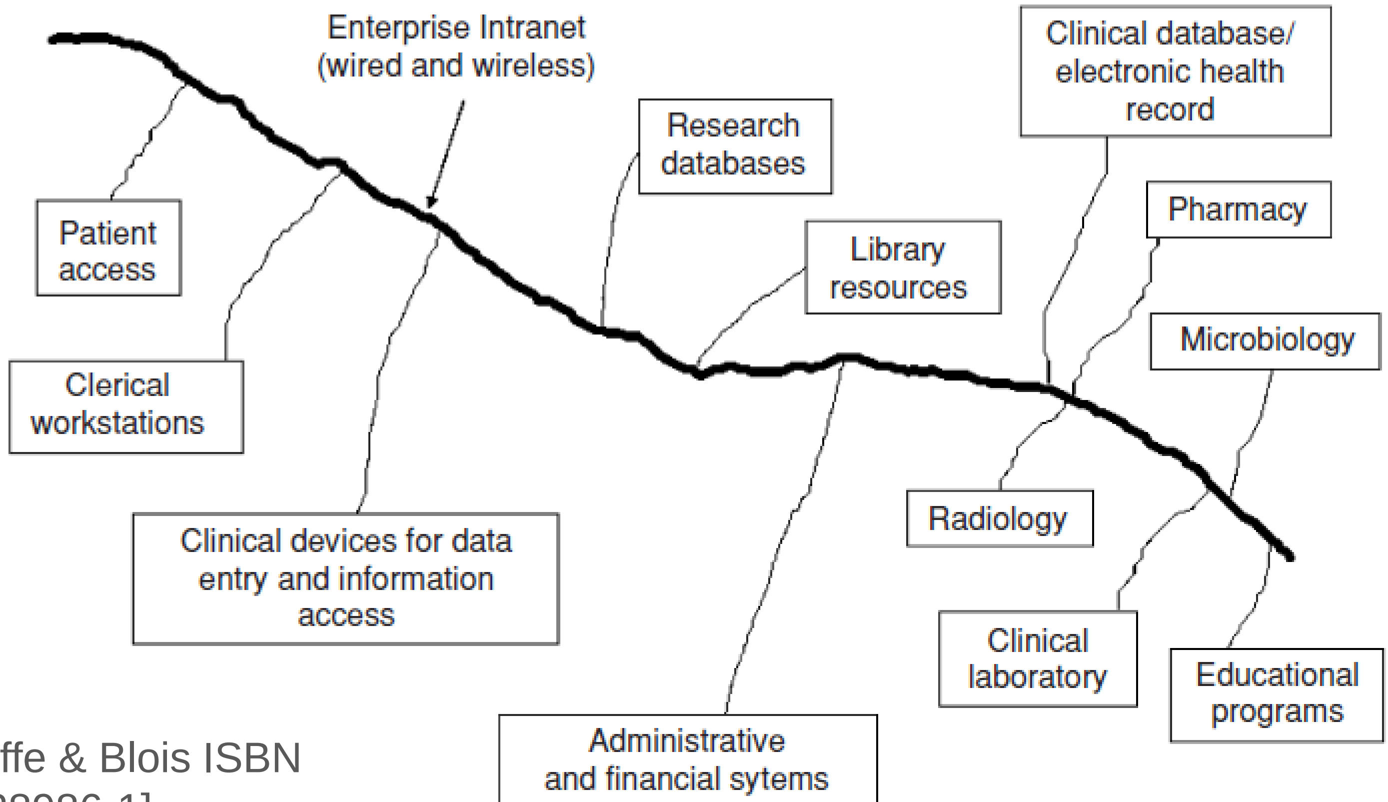
- The data needed for a study can be derived directly from the EHR, thus making research data collection a by-product of routine clinical record keeping



[Ref: Shortliffe & Blois ISBN 978-0-387-28986-1]

Integrating the Patient Record with Other Information Resources

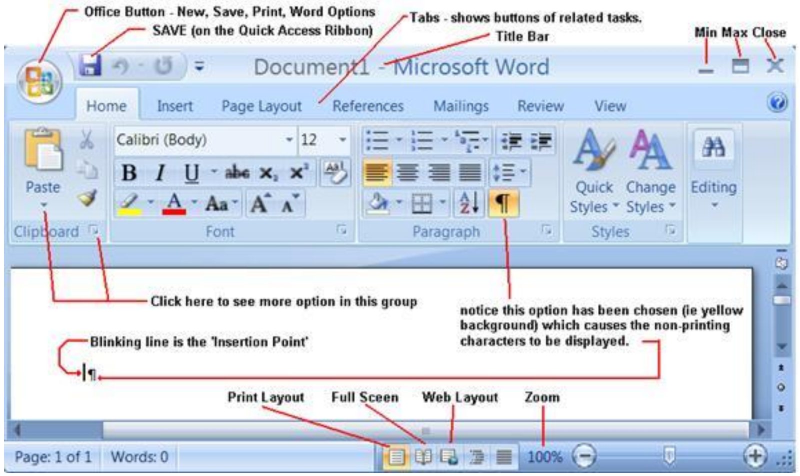
- To achieve effective use of EHR offer a critical mass of functionality that makes the system both smoothly integrated and useful for essentially every patient encounter.



[Ref: Shortliffe & Blois ISBN 978-0-387-28986-1]

Rethinking Common Assumptions

to create an electronic version of an object or process from the physical world.



The computer can thus facilitate paradigm shifts in how we think about such familiar concepts.

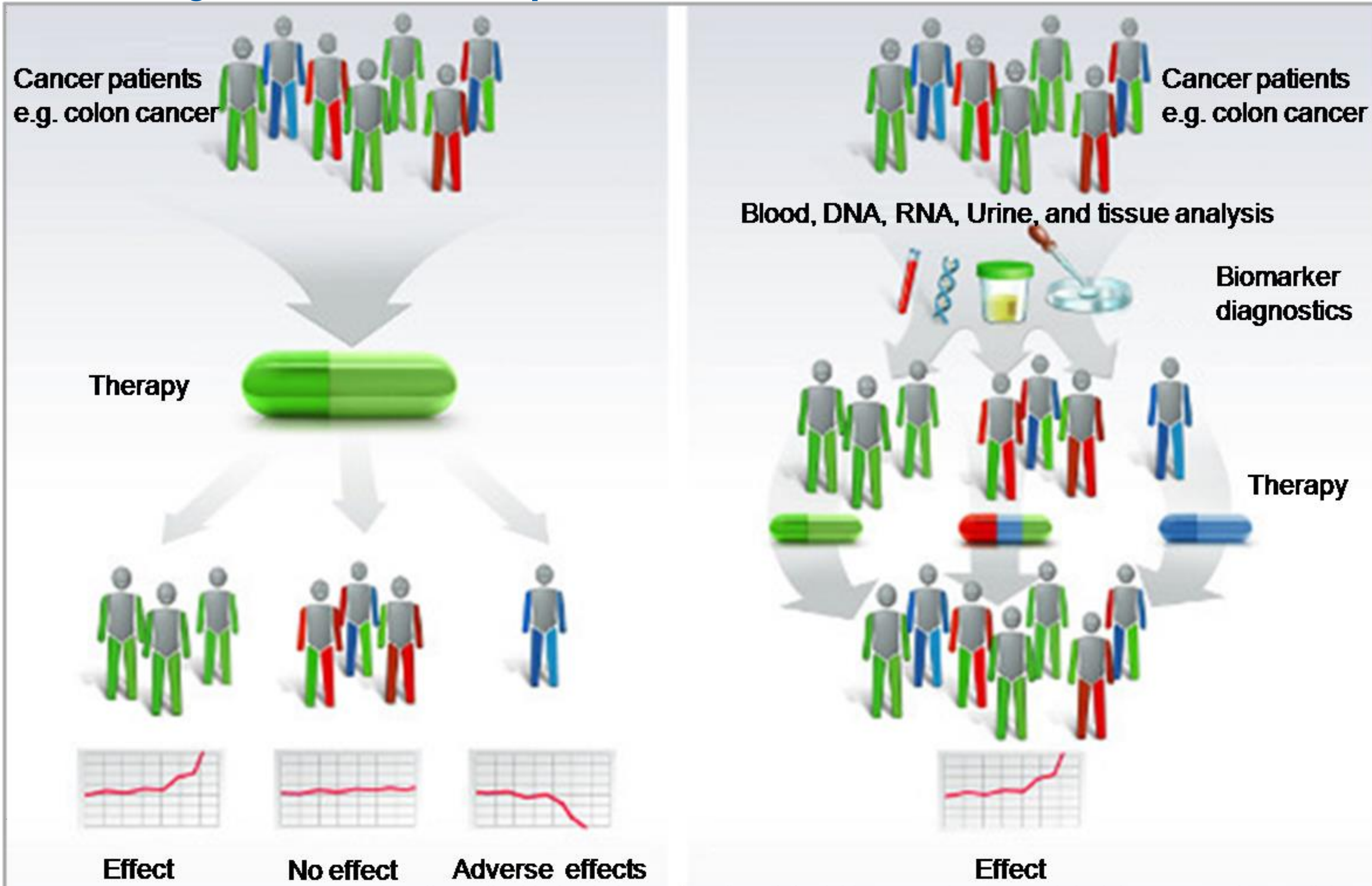
Rethinking Common Assumptions

new transformative paradigm called BIG DATA that is revolutionizing medicine

We are moving towards hypothesis based knowledge generation to unbiased, data driven analysis to discover new insights for

- understanding disease progress
- predicting treatment outcome and prognosis
- identifying risks

Rethinking Common Assumptions

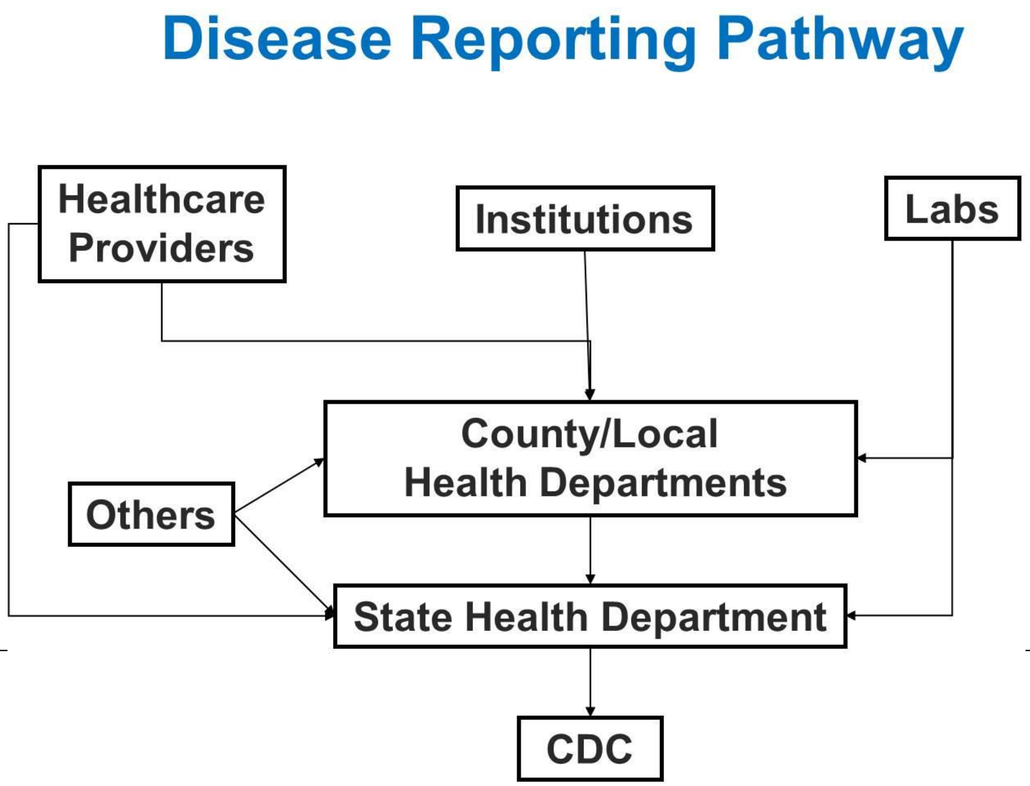


http://www.ibmbigdatahub.com/blog/bench-bedside-leveraging-research-insights-personalize-medicine

A Model of Integrated Disease Surveillance

Disease surveillance is an epidemiological practice by which the spread of disease is monitored in order to establish patterns of progression.

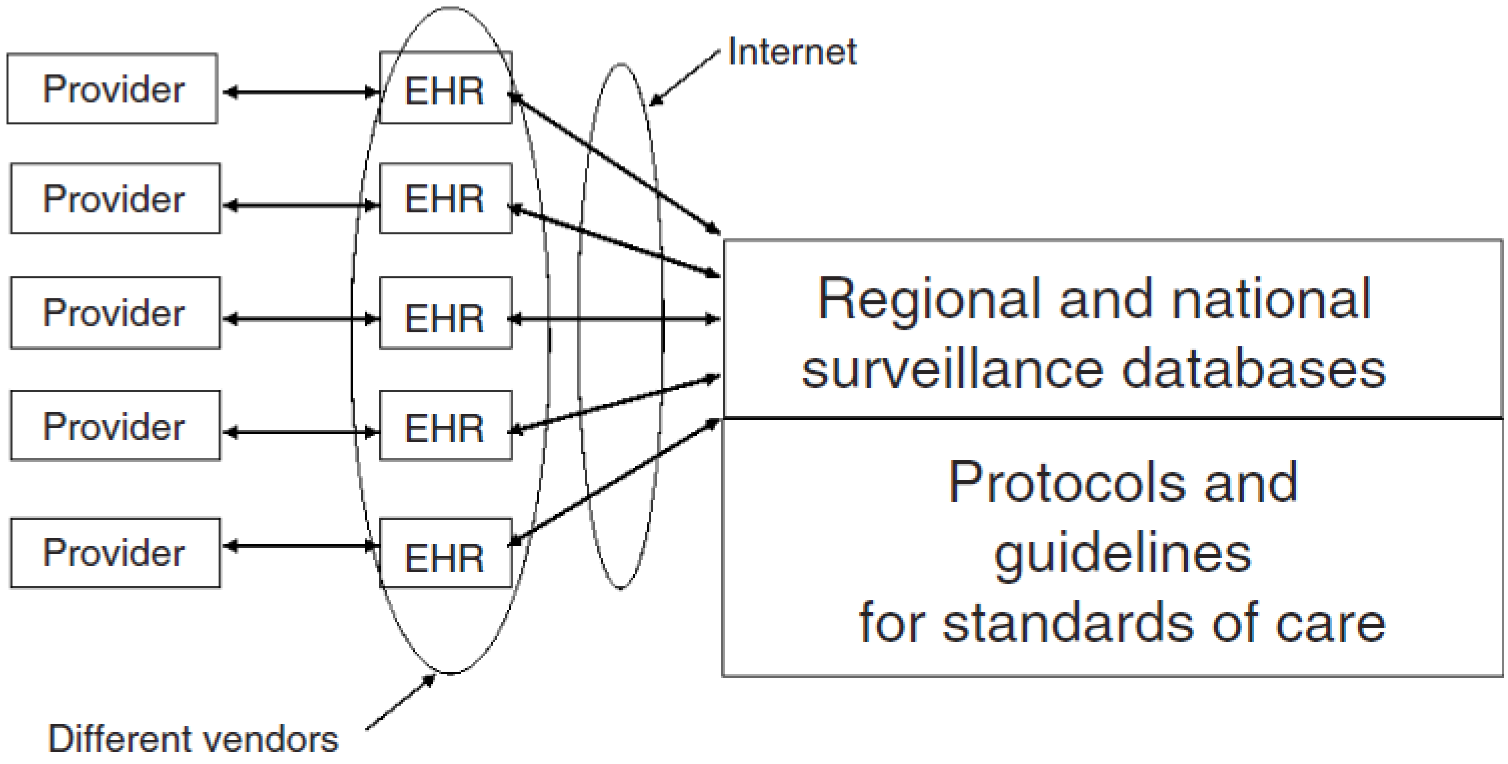
- The main role of disease surveillance is to predict, observe, and minimize the harm caused by outbreak, epidemic, and pandemic situations, as well as increase knowledge about which factors contribute to such circumstances.
- A key part of modern disease surveillance is the practice of disease case reporting to Centers for Disease Control and Prevention.



the nation's networking infrastructure could play in integrating clinical data and enhancing care delivery

A Model of Integrated Disease Surveillance

- A future vision of surveillance databases, in which clinical data are pooled in regional and national repositories



A Model of Integrated Disease Surveillance

The practical issues for realizing integrated clinical data spaces :

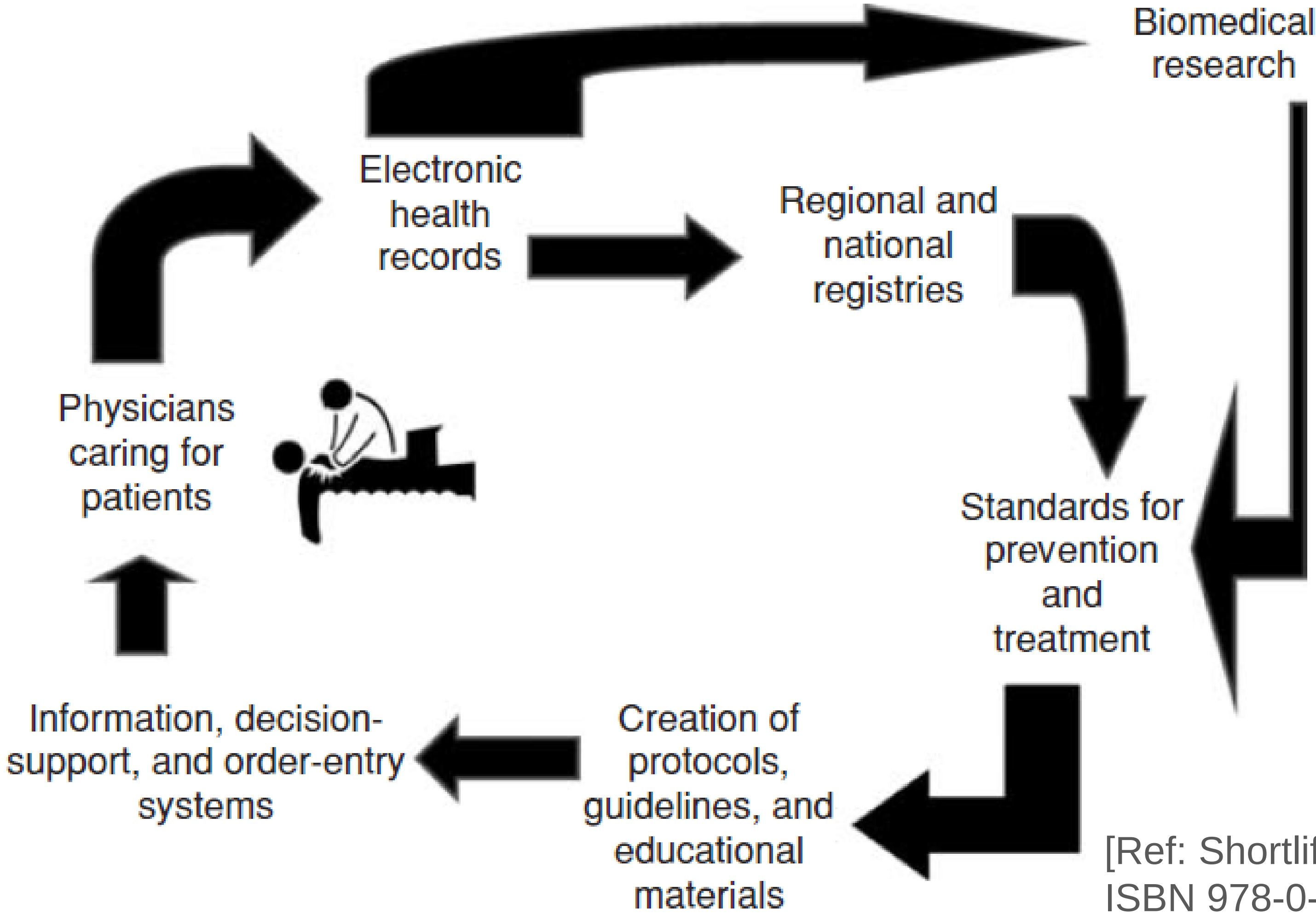
- *Encryption of data*
- *Standards for data transmission and sharing*
- *Standards for data definitions*
- *Quality control and error checking*

A Model of Integrated Disease Surveillance

Benefits of Integrated information:

- Recommended steps for health promotion and disease prevention
- Detection of syndromes or problems, either in their community or more widely
- Trends and patterns of public health importance
- Clinical guidelines, adapted for execution and integration into patient-specific decision support rather than simply provided as text documents
- Opportunities for distributed (community-based) clinical research, whereby patients are enrolled in clinical trials and protocol guidelines are in turn integrated with the clinicians' EHR to support protocol-compliant management of enrolled patients

The Cycle of Information Flow in Clinical Care

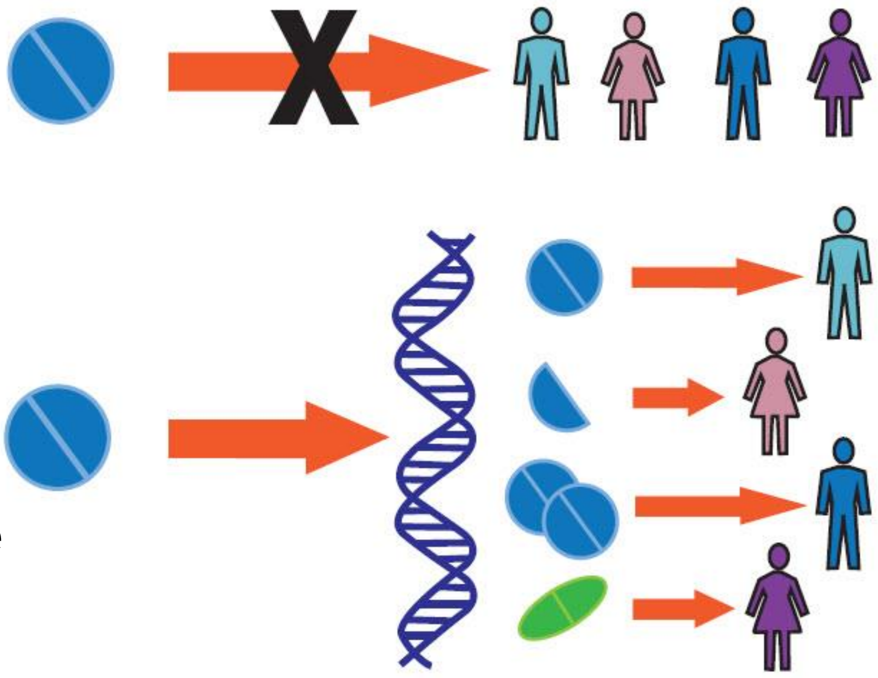


[Ref: Shortliffe & Blois ISBN 978-0-387-28986-1]

What is Personalized Medicine ?

Definition

- Personalized medicine is a medical model that separates patients into different groups –
  with <u>medical decisions</u>, <u>practices</u>, <u>interventions</u> and/or <u>products</u> being tailored to the individual patient based on their predicted response or risk of disease.
- Synonyms:
- precision medicine, stratified medicine
- P4 medicine
  predictive, preventive, personalized, and participatory in healthcare

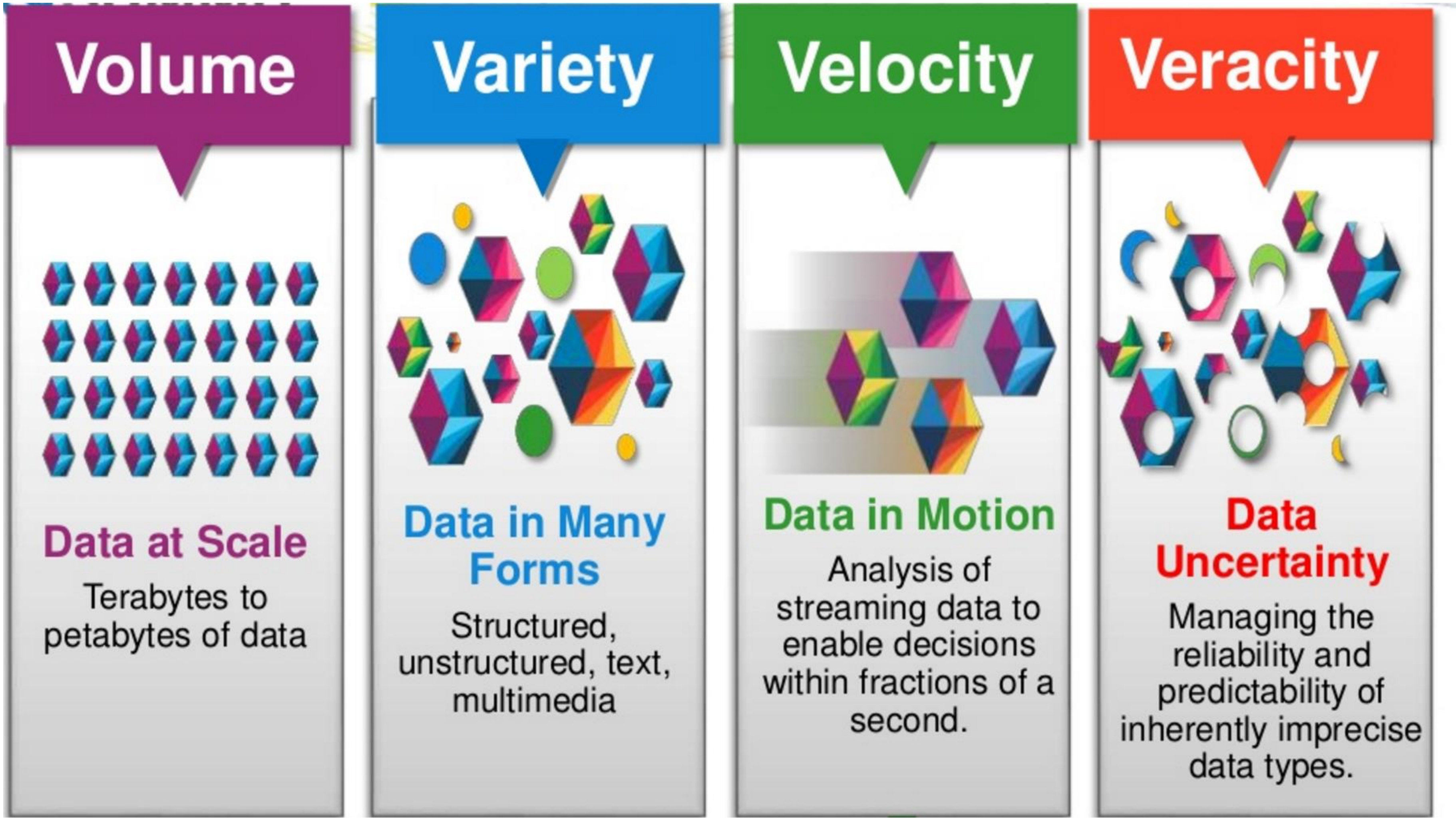


Source: https://iipm.medicinedept.iu.edu/

What is Personalized Medicine ?

- Make better decisions on **risk prediction** and focus on **prevention**, rather than disease management.
- Detect disease at an **earlier stage**, when it is easier and less expensive to treat effectively
- Stratify patients into groups that enable the selection of **optimal therapy**
- **Reduce adverse drug** reactions by more effective early assessment of individual drug responses
- Improve the selection of **new biochemical targets** for drug discovery
- Reduce the time, cost, and **failure rate of clinical trials** for new therapies
- Shift the emphasis in medicine from **reaction to prevention** and from **disease to wellness**

Big Data

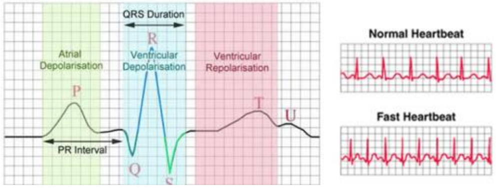


Source: Big Data: Transform Industries, Institutions, and Societies with Watson, Paul Yip

Where the data comes from ?

Clinical data: Records, imaging, screening and diagnostic testing...

- Electronic health records
- Administrative data
- Claims data
- Disease registries
- Health surveys
- Clinical trials data

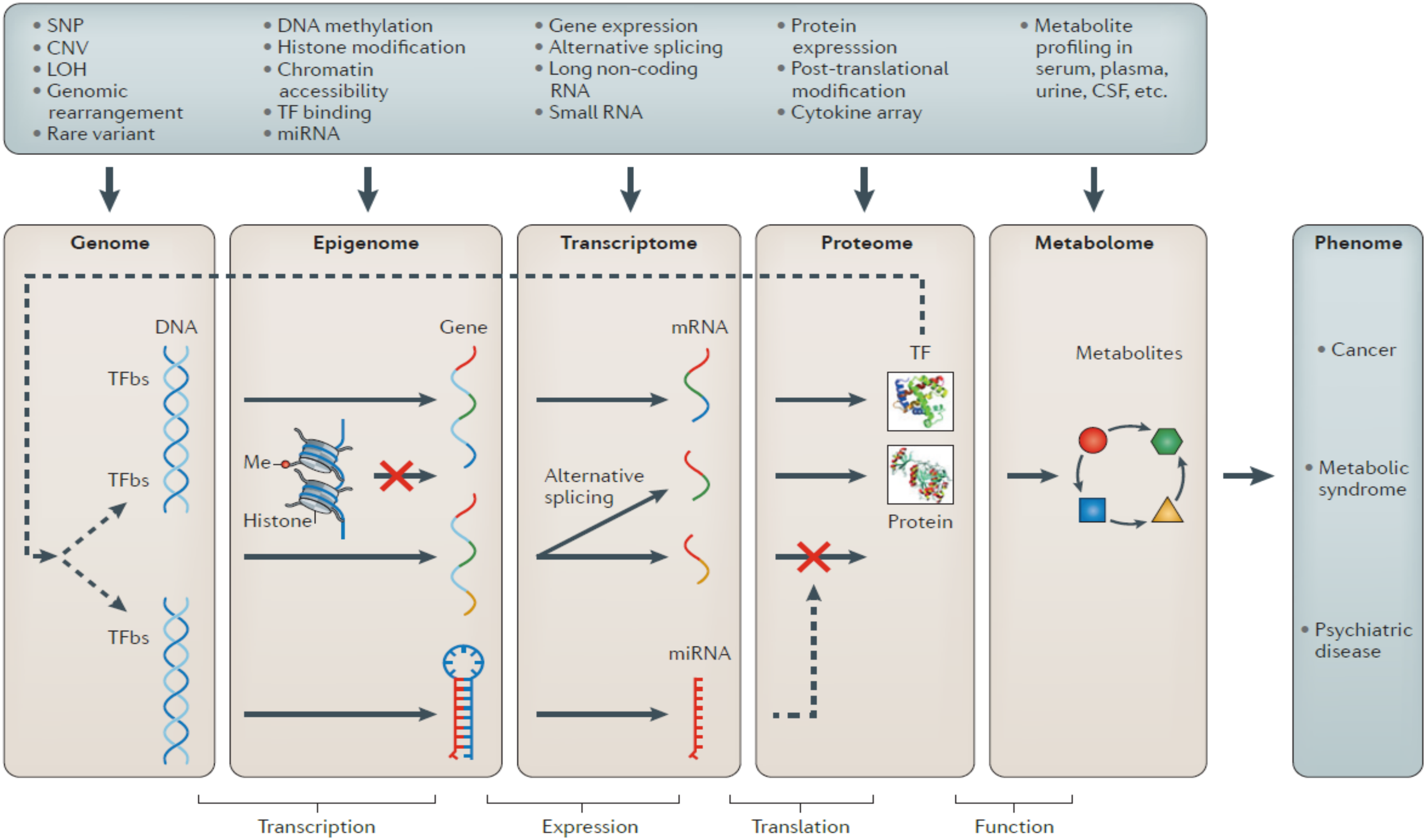


- ECG
- biopsy
- prostate-specific antigen (PSA) test
- ...

- X-ray radiography
- magnetic resonance imaging
- medical ultrasonography
- endoscopy,
- thermography
- nuclear medicine functional imaging techniques (PET , SPECT)

Where the data comes from ?

Genomic Data Landscape



Source: Ritchie, Marylyn D., et al. "Methods of integrating data to uncover genotype-phenotype interactions." Nature Reviews Genetics 16.2 (2015): 85-97.

Where the data comes from ?

Sensors and Smart Phones

- wearable body sensors
- internet of things
- Smart phones
- Robots

RWTH Informatik 5 | Ahornstr. 55 D-52056 Aachen Germany |
Tel +49/241/8021501 | Fax +49/241/8022321 | http://dbis.rwth-aachen.de/cms

How big is the data?

- The proliferation of smart devices and exponentially decreasing cost to sequence the human genome along with growth of electronic communication via social media is generating an explosion of health related data.

Some estimates just for genome data

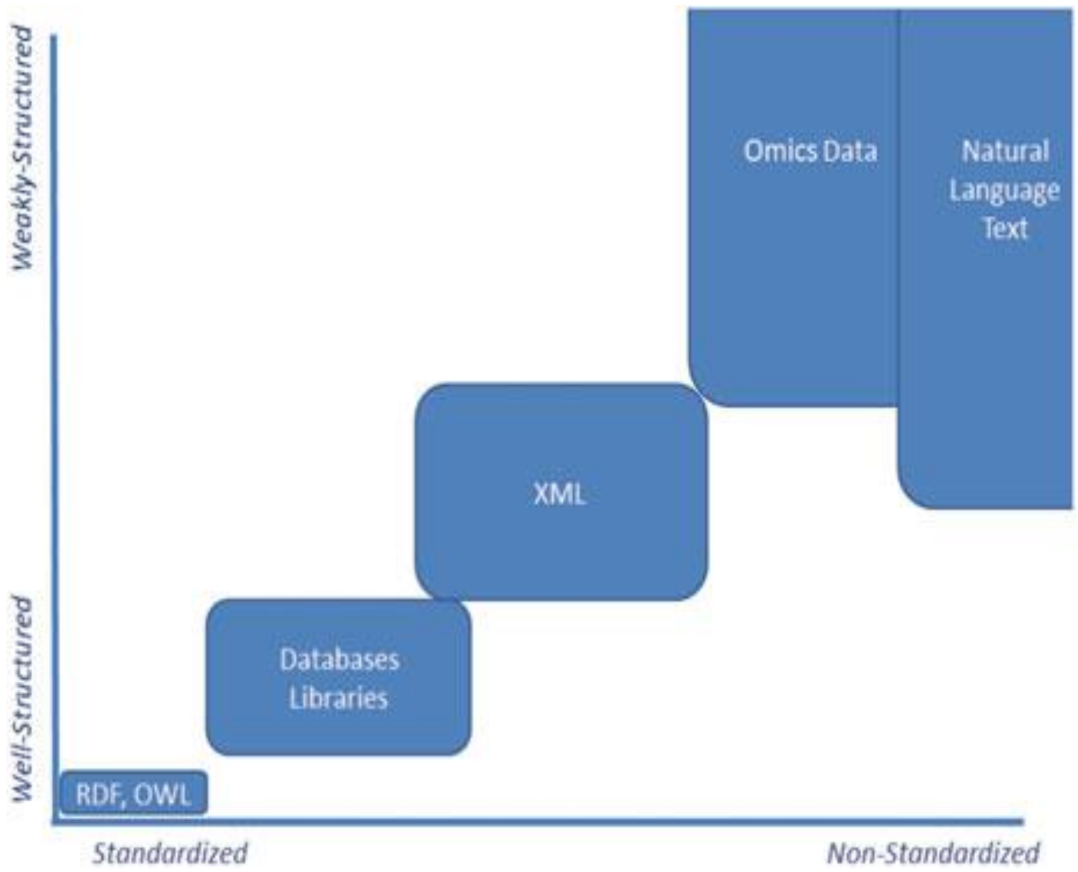
- The human genome contains 2.9 billion base pairs. So if you represented each base pair as a byte then it would take 2.9 billion bytes or **2.9 GB**.
- By 2025, between **100 million and 2 billion** human genomes could have been sequenced,
- the number of data that must be stored for a single genome are **30 times** larger than the size of the genome itself (to correct errors incurred during sequencing and preliminary analysis)
- the data-storage demands for this alone could run to as much as **2–40 exabytes** (1 exabyte is 10^{18} bytes)
- YouTube’s projected annual storage needs of 1–2 exabytes of video by 2025 and Twitter’s projected 1–17 petabytes per year

Source: Hayden, Erika Check. "Genome researchers raise alarm over big data." Nature (2015).

Challanges

Understanding and Using

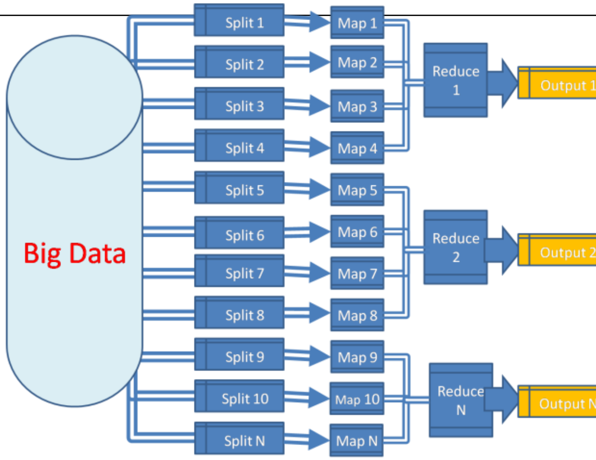
- Increasingly large and complex datasets
- non-standardized and unstructured information
- Semi-structured, weakly structured data
- Data quality, data integration, universal access.
- Privacy, security, safety, and data protection issues

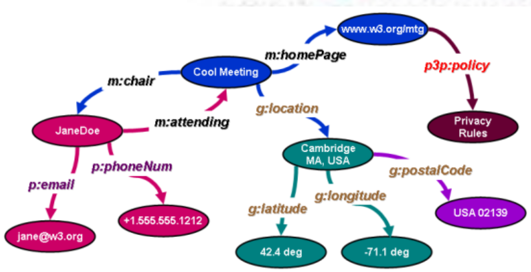


Source Holzinger, Andreas. *Biomedical informatics*. Books on Demand, 2012.

Big Data Technologies

- efficient and effective storage techniques: data compression and storage virtualization
- distributed data storage and analysis environments for processing vast amounts of data (multi-terabyte data-sets)
- programming methods that allows for massive job execution scalability against thousands of servers or clusters of servers
- Schema-less databases, or NoSQL databases for the storage and retrieval of large volumes of unstructured, semi-structured, or structured data.
- high-performance machine learning and data analytics platform focused specifically on handling big data
- technologies to make big data smarter





How big data will change the medicine ?

Today

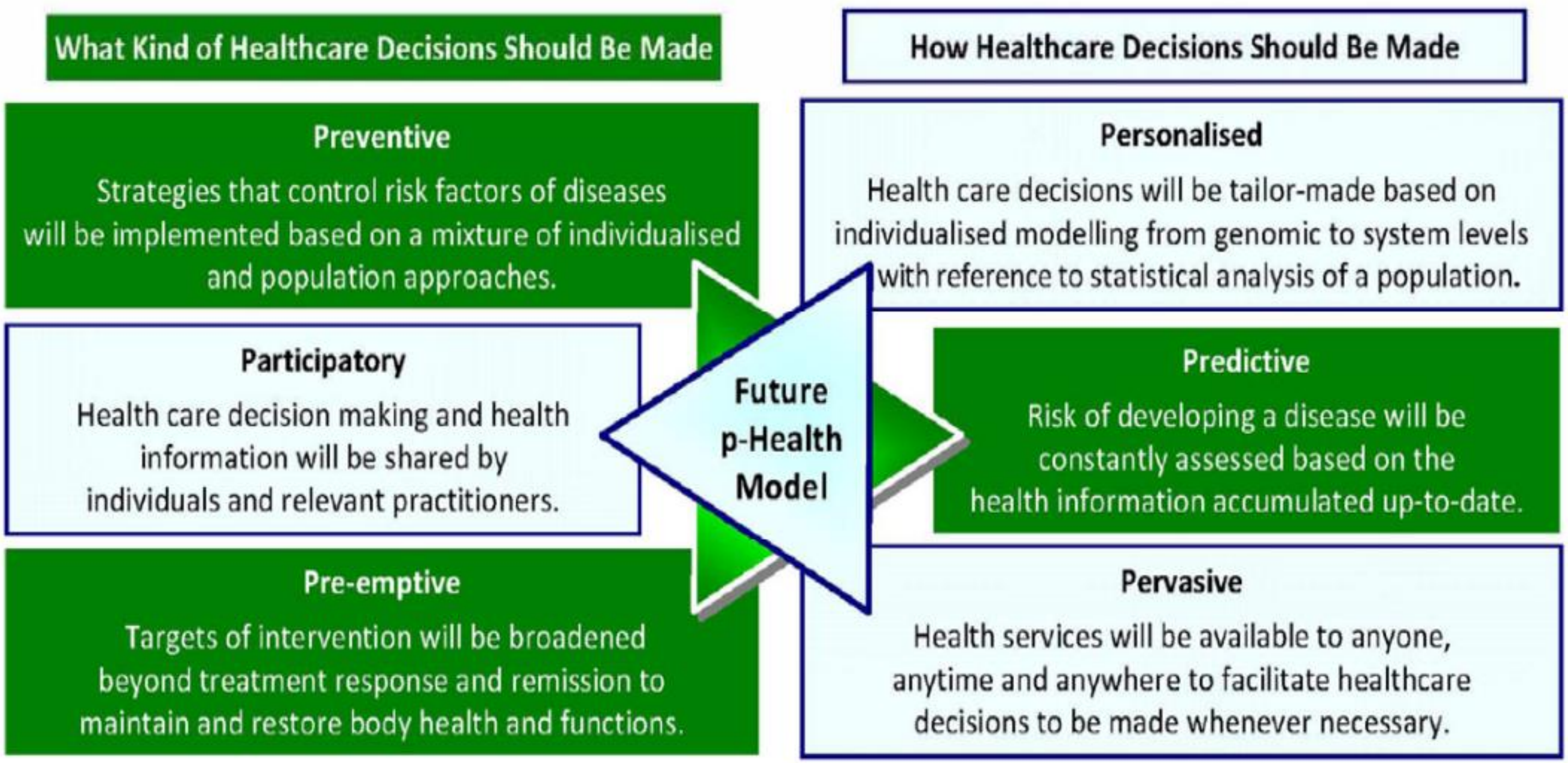
A patient has late-stage non-small cell lung cancer (NSCLC). She has gone through a number of treatments, but none were able to arrest the cancer's spread. The mother of four has progressive disease and precious little time to waste on treatments that do not work. Her physician read reports of a newly approved drug called Xalkori® (crizotinib) that might offer hope. However, only about five percent of NSCLC patients whose tumors have the anaplastic lymphoma kinase (ALK) gene rearrangement can potentially benefit. A newly approved diagnostic test determines that the patient has the gene rearrangement and that the drug is a treatment option for her. After starting to take Xalkori®, the tumors begin to respond.

Source: http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/the_case_for_pm1.pdf

The Future

Imagine a physician sitting down with his laptop and a morning cup of coffee. On a website that he uses to help manage his practice, an alert pops up. It tells him that a series of studies have demonstrated a connection between multiple rare mutations found in 10 percent of people and the likelihood that they might convert to type 2 diabetes. Nearly all of his patients have had their entire genome sequenced and entered into their electronic medical record – a process that takes only a week, costs a few hundred dollars, and is reimbursed by insurance companies because of the many benefits it provides to lifelong health management. He conducts a quick search of his 2,000 patient database and finds about 80 who are at risk. To half of those patients, he sends a strong reminder and advice on diet and lifestyle choices they can take to avoid the disease. To the other half, whose medical records reveal pre-diabetic symptoms, he sets up appointments to consider more proactive treatment with drugs that can prevent the onset of disease.

4 Ps of Personalised Medicine



Zhang, Y. T. & Poon, C. C. Y. (2010) Editorial Note on Bio, Medical, and Health Informatics. *Information Technology in Biomedicine, IEEE Transactions on, 14, 3, 543-545.*

Terminology: Medical /biomedical Informatics

- 1970+ Begin of Medical Informatics
- Focus on data acquisition, storage, accounting (typ. “EDV”)
- The term was first used in 1968 and the first course was set up 1978
- 1985+ Health Telematics
- Health care networks, Telemedicine, CPOE-Systems etc.
- 1995+ Web Era
- Web based applications, Services, EPR, etc.
- 2005+ Ambient Era
- Pervasive & Ubiquitous Computing
- 2010+ Quality Era – Biomedical Informatics
- Information Quality, Patient empowerment, individual
- molecular medicine

Terminology: Medical /biomedical Informatics

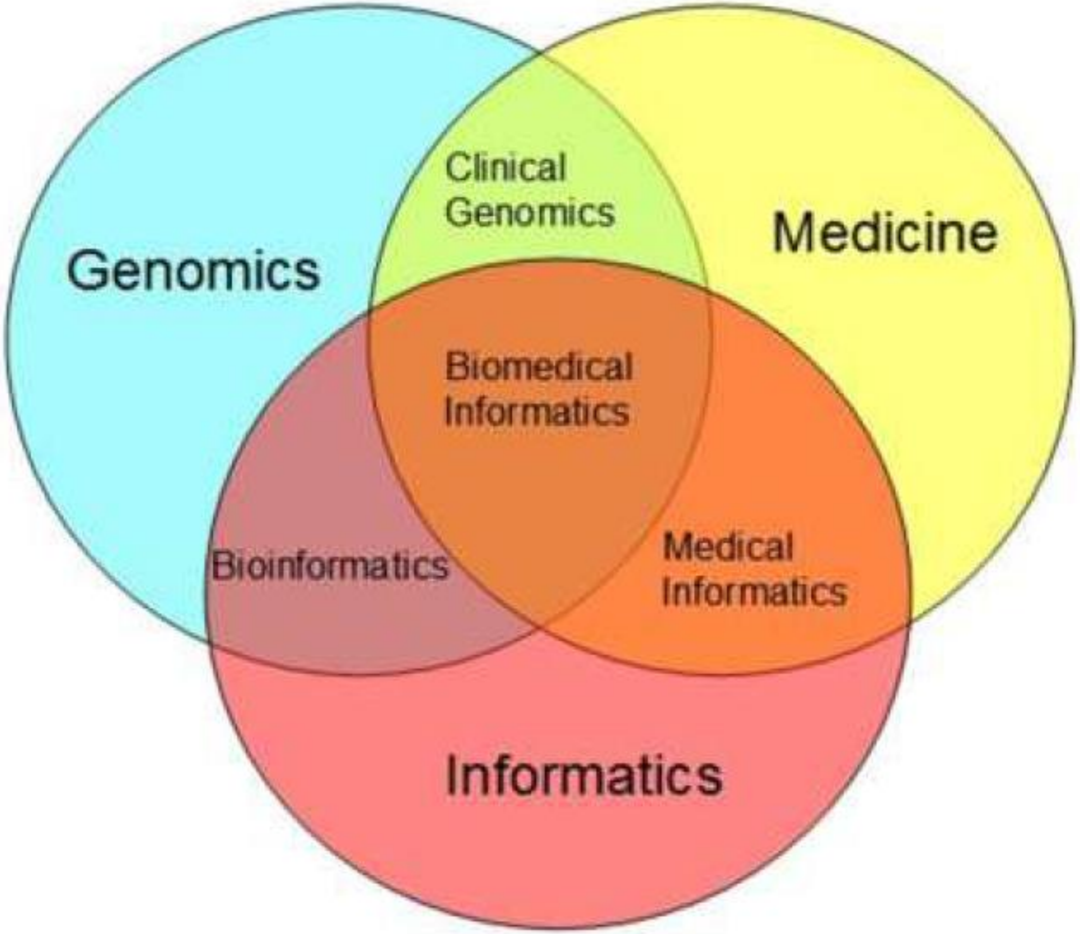
- The term **information science** originated in the field of library science and is used to refer, somewhat generally, to the broad range of issues related to the management of both paper-based and electronically stored information.
- The terms **biomedical computing** or **biocomputation** are non descriptive and neutral, implying only that computers are employed for some purpose in biology or medicine
- **medical informatics**, which is broader than **medical computing** (it includes such topics as medical statistics, record keeping, and the study of the nature of medical information itself) and deemphasizes the computer while focusing instead on the nature of the field to which computations are applied
- **biomedical informatics** It is becoming the most widely accepted term and should be viewed as encompassing broadly all areas of application in health, clinical practice, and biomedical research.

Terminology: Medical /biomedical Informatics

Medical informatics is the rapidly developing scientific field that deals with resources, devices and formalized methods for optimizing the storage, retrieval and management of biomedical information for problem solving and decision making.

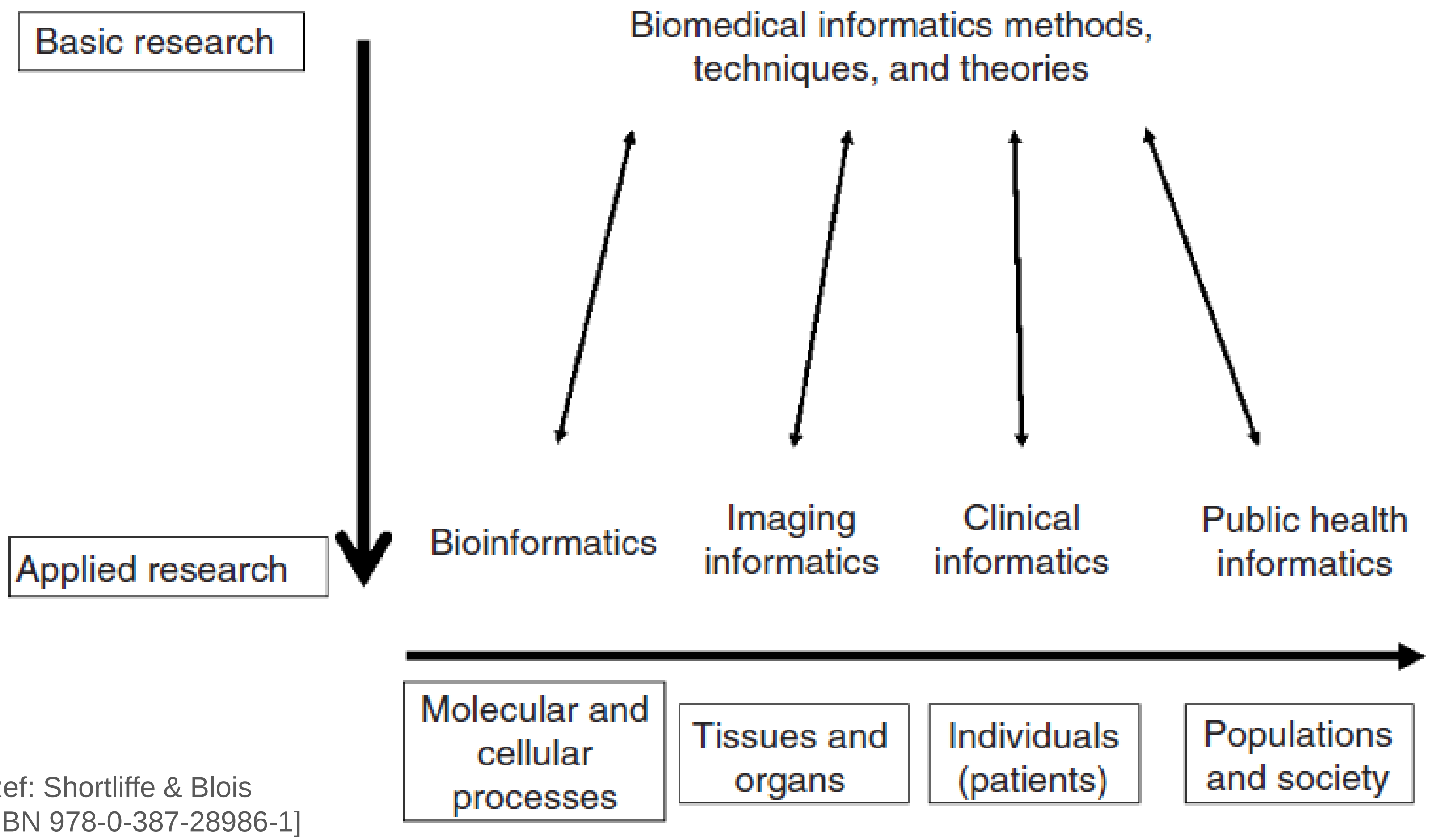
Medical Informatics is the branch of science concerned with the use of computers and communication technology to acquire, store, analyze, communicate, and display medical information and knowledge to facilitate understanding and improve the accuracy, timeliness, and reliability of decision-making.
Warner, Sorenson and Bouhaddou, Knowledge Engineering in Health Informatics, 1997

Terminology: Medical /biomedical Informatics



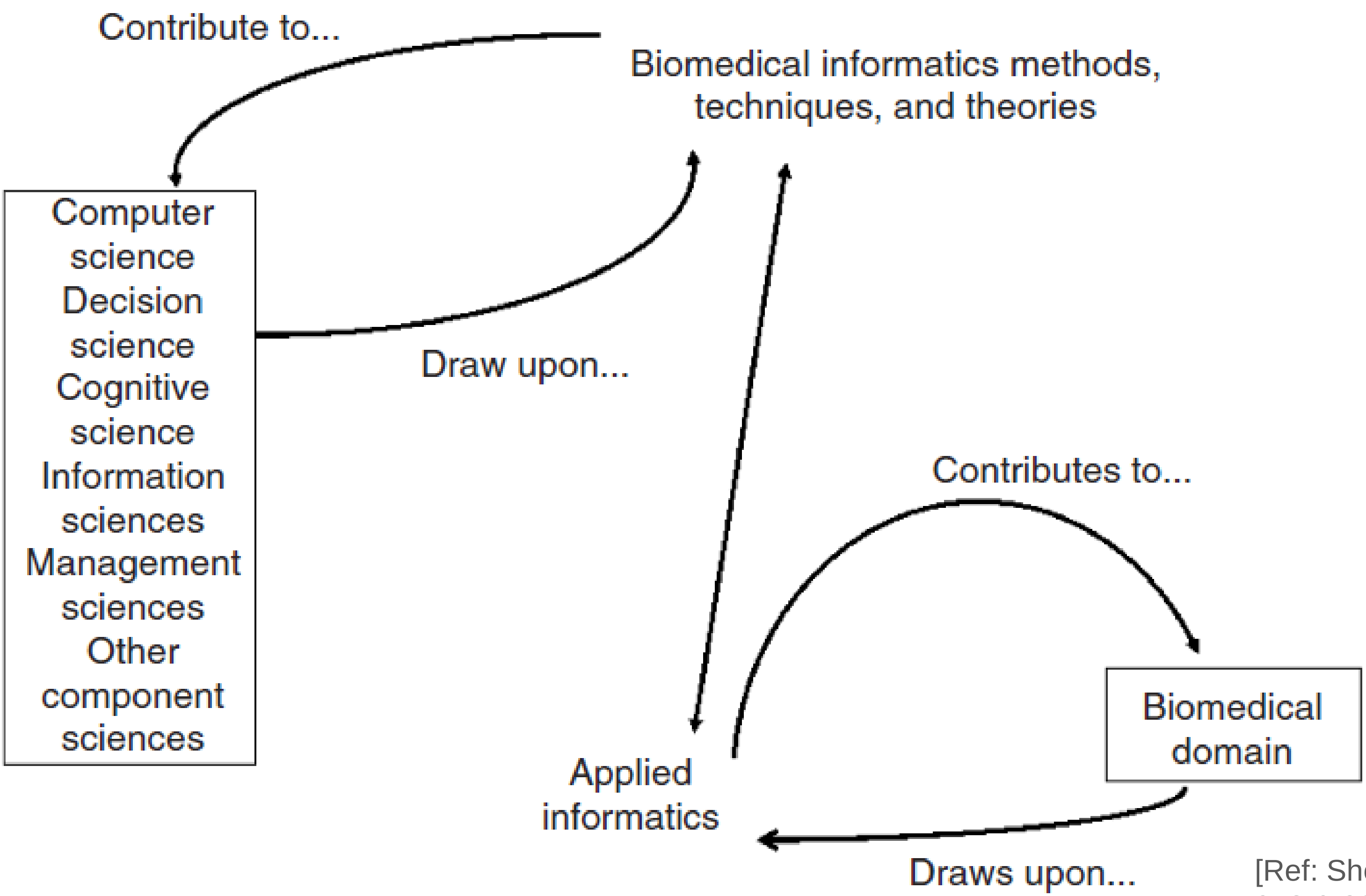
http://www.bioinformaticslaboratory.nl/twiki/bin/view/BioLab/EducationMIK1-2

Biomedical informatics as basic and applied science



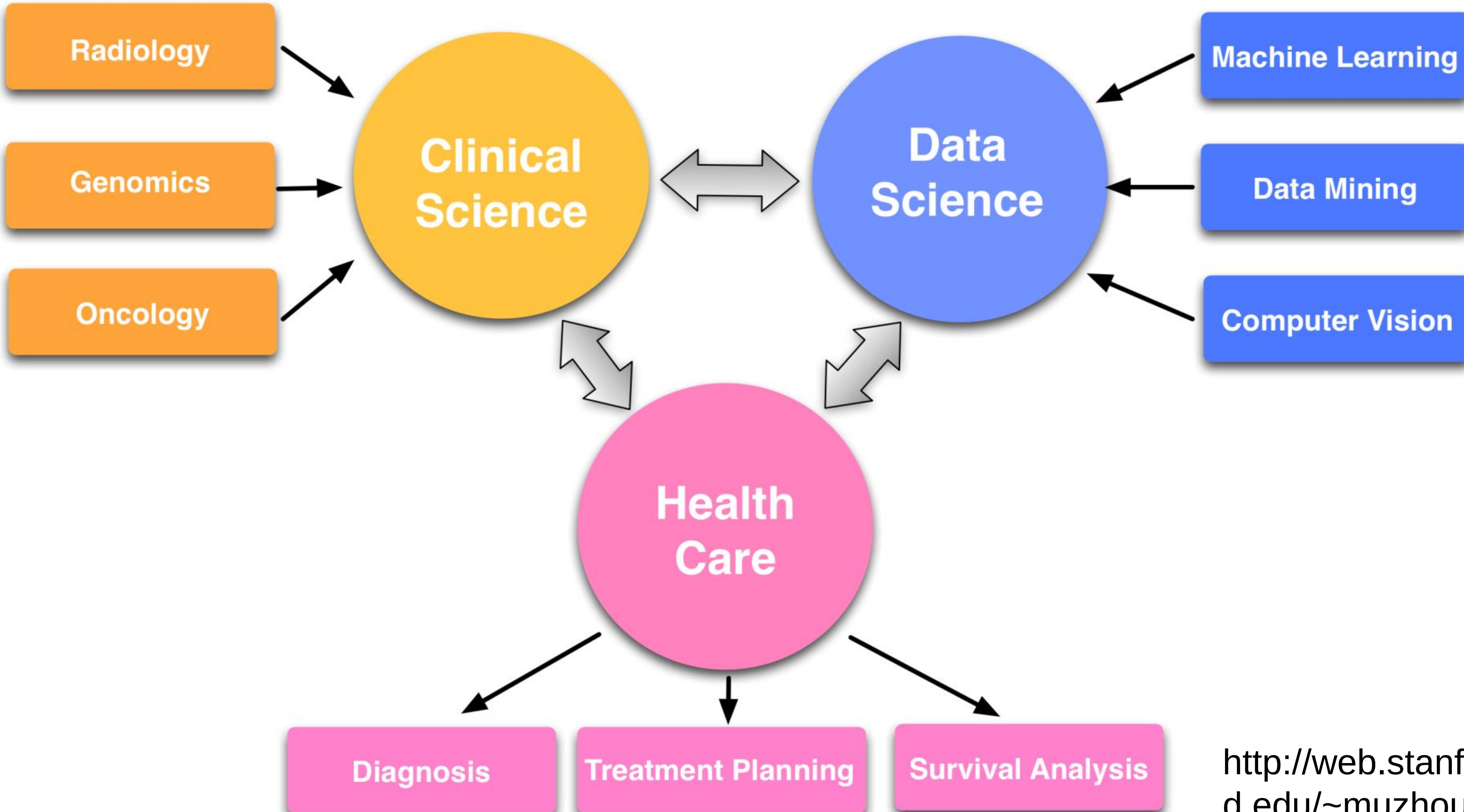
[Ref: Shortliffe & Blois
ISBN 978-0-387-28986-1]

Component sciences in biomedical informatics



[Ref: Shortliffe & Blois ISBN 978-0-387-28986-1]

Big Data in Medical Informatics



http://web.stanford.edu/~muzhou1/

Topics

- Understanding the Biomedical Data
- Structured, unstructured, and semi structured data
- Methods of structuring data - nomenclatures, thesauruses ..
- data acquisition methods

- Clinical data sources
- Electronic health records
- Types of hospital information systems
- Standards for coding the clinical data (UMLS, ICD, SNOMED, DICOM, LOINC)

- Data Interoperability
- types of interoperability (semantic, syntactic)
- interoperability standards (HL7, CDA, FHIR..)
- clinical data exchange protocols

Topics

- Semantic Web Technologies in Biomedicine
- Main concepts of the Semantic Web Technologies
- Biomedical ontologies and examples
- Introduction to protégé tool
- Basic concepts of RDF and Linked Data
- Biomedical data repositories and data sharing
- Applications of semantic web technologies in health care

- Methods and tools for biomedical data analysis
- basic biomedical data analysis workflow
- main methods for biomedical data analysis
- evaluation and validation
- tools for data analysis
- Introduction to the R programming language

Topics

- Predictive Analytics and Decision Making Systems
- review of clinical prediction models
  (regression, Bayesian models, neural networks, reinforcement learning)
- examples of predictive health care applications from real life
-
- Genomic Data Analysis
- Genomic data generation and data types
- methods and standards genomic data analysis
- examples of genomic studies for personalized medicine
- open biomedical data repositories
- tools for genomic data science
- introduction to Galaxy platform

Text Books

- Healthcare Data Analytics. Chandan K. Reddy and Charu C. Aggarwal. Boca Raton, FL: Chapman & Hall/CRC Press (2015) 724 pp.
- Interactive Knowledge Discovery and Data Mining in Biomedical Informatics Editors: Holzinger, Andreas, Jurisica, Igor (Eds.) (2014)
- Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics) 4th ed. 2014 Edition by Edward H. Shortliffe (Editor), James J. Cimino (Editor)

Course organization

Location:
- Chair Informatik 5 , Ahornstraße 55, Building extension E2, 2nd floor

Contact via email: beyan@dbis.rwth-aachen.de

All materials and information about lecture will be provided via L2P

12 Lectures and 6 recitations

Exam eligibility: Collecting 50% of points from exercises.