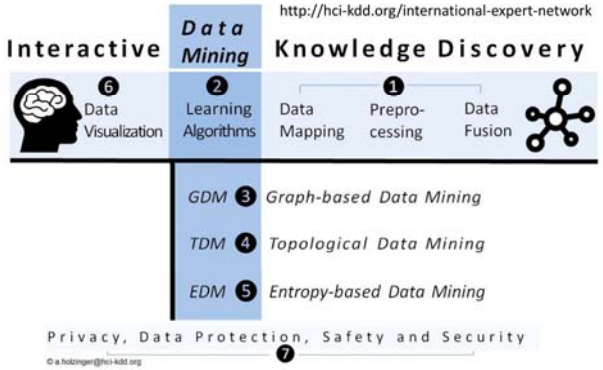
Andreas Holzinger
VO 709.049 Medical Informatics
18.01.2017 11:15-12:45

Lecture 10 Biomedical Information Systems and Knowledge Management

a.holzinger@tugraz.at
Tutor: markus.plass@student.tugraz.at
http://hci-kdd.org/biomedical-informatics-big-data

ML needs a concerted effort fostering integrated research



Holzinger, A. 2014. Trends in Interactive Knowledge Discovery for Personalized Medicine: **Cognitive Science meets Machine Learning.** IEEE Intelligent Informatics Bulletin, 15, (1), 6-14.

Keywords

- Bioinformatics workflows
- Clinical workflow & management systems
- Cloud computing in healthcare
- Communication standards
- Digital Imaging and Communication in Medicine (DICOM)
- Formal methods & workflow modeling
- Health Level 7 (HL7)
- Logical Observation Identifier Names and Codes (LOINC)
- Medical multimedia
- Mobile computing in medicine
- Personal Health Record (PHR)
- Picture Archiving and Communication System (PACS)
- Quality
- Software as a Service (SaaS)
- Systems architecture
- Unified Modeling Language (UML)

Advance Organizer (1/2)

- **Bioinformatics workflow management system** = designed specifically to compose and execute a series of computational and/or data manipulation steps and/or workflows in the domain of bioinformatics;
- **Business process re-engineering (BPR)** = analysis and design of workflows and processes within an organization (=hospital). According to Davenport (1990) a BP is a set of logically related tasks performed to achieve a defined outcome/result;
- **Clinical Pathway** = aka care map, a tool used to manage the quality in healthcare concerning the standardization of care processes and promote organized and efficient patient care based on EBM;
- **Digital Imaging and Communications in Medicine (DICOM)** = a standard for handling, storing, printing, and transmitting data in medical imaging (also file format definition and a network communications protocol using TCP/IP);
- **Evidence-based medicine (EBM)** = aiming at developing mathematical estimates of benefit and harm from population-based research and apply these in the clinical routine, claiming that best research evidence on medical interventions come from experiments (e.g. randomized controlled trials);
- **Health Level Seven (HL 7)** = a Standardization Organization accredited by the American National Standards Institute (ANSI) to push consensus-based standards representing healthcare stakeholders;

Advance Organizer (2/2)

- **Hospital Information System (HIS)** = integrated information system for (administrative, financial, clinical etc.) information management in a hospital;
- **Integrating Healthcare Enterprise (IHE)** = initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information (i.e. promotes the coordinated use of established standards such as DICOM and HL7);
- **National Electrical Manufacturers Association (NEMA)** = holds copyright of DICOM;
- **Paradigm** = according to Kuhn (1962) a shared view of a group of researchers, comprising 4 elements: concepts, theories, methods and instruments;
- **Picture Archiving and Communication System (PACS)** =system for handling images from various medical imaging instruments, including ultrasound (US), magnetic resonance (MR), positron emission tomography (PET), computed tomography (CT), endoscopy (ENDO), mammographs (MG), Digital radiography (DR), computed radiography (CR) ophthalmology, etc.;
- **Workflow** = consists of a sequence of connected steps, succeeding the flow paradigm, where each step follows the precedent;

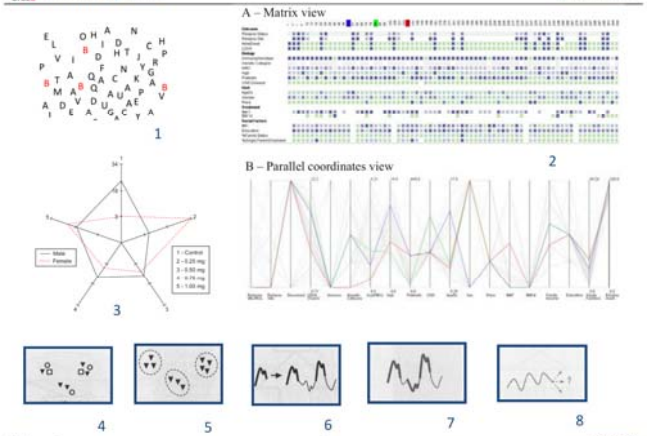
Learning Goals: At the end of this 10th lecture you ...

- have an overview about workflows and workflow modeling in health care;
- got an overview of typical architectures of hospital information systems for patient records as already discussed in lecture 4;
- have understood the principles of Picture Archiving and Communication PACS-Systems;
- know how important multimedia for medicine is;
- have a basic understanding of DICOM and HL 7;
- are aware of the constraints of open source software in the medical domain;
- have got an idea of possible future systems solutions;

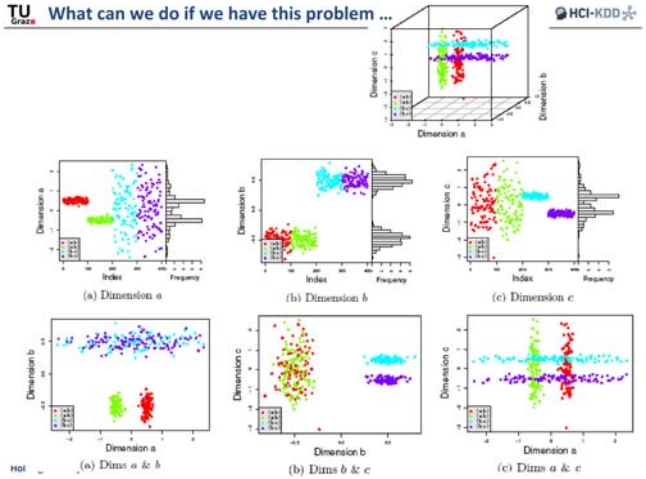
Agenda for today

- 00 Reflection – follow-up from last lecture
- 01 Information Systems Challenges
- 02 Workflows – Clinical Pathways
- 03 Hospital Information Systems (HIS)
- 04 Multimedia: PACS
- 05 Data Standards for Communication
- 06 Open …
- 07 Conclusion & Future Outlook

Warm-up Quiz



What can we do if we have this problem ...



01 Information Systems Challenges

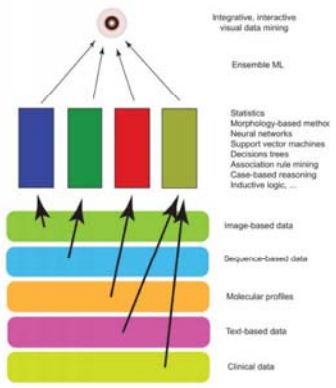
Slide 10-1 Key Challenges

- Lack of **Integrated** Approaches
- Clinical Workplace **efficiency**
- Cloud Computing (**Privacy,** Security, Safety, Data Protection…)
- Useful, Useable **Service oriented** computing (as electricity is already!)

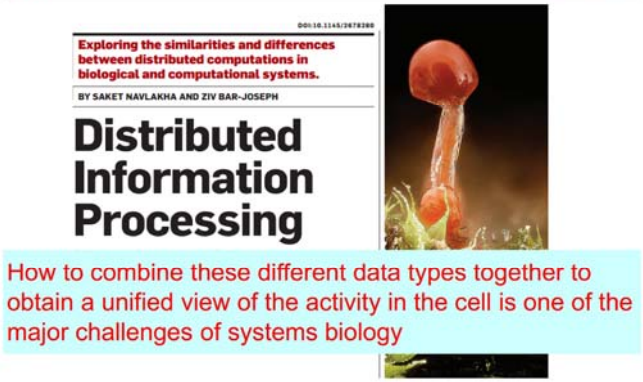
Holzinger, A. & Jurisica, I. 2014. Knowledge Discovery and Data Mining in Biomedical Informatics: The future is in Integrative, Interactive Machine Learning Solutions In: Lecture Notes in Computer Science LNCS 8401. Heidelberg, Berlin: Springer, pp. 1-18, doi:10.1007/978-3-662-43968-5_1.

Unsolved: Data Integration, Data Fusion, Data Mapping

Goal:
Unified View,
one single workflow,
Workflow continuity,
no media gaps
(DE: Medienbruch)



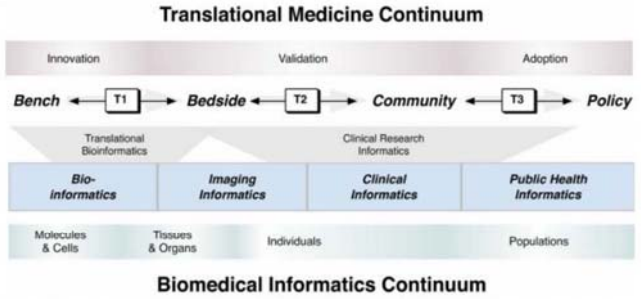
Holzinger, A. & Jurisica, I. 2014. Knowledge Discovery and Data Mining in Biomedical Informatics: The future is in Integrative, Interactive Machine Learning Solutions In: Lecture Notes in Computer Science LNCS 8401. Heidelberg, Berlin: Springer, pp. 1-18, doi:10.1007/978-3-662-43968-5_1.



How to combine these different data types together to obtain a unified view of the activity in the cell is one of the major challenges of systems biology

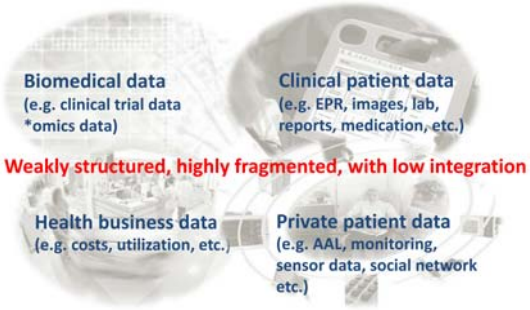
Navlakha, S. & Bar-Joseph, Z. 2014. Distributed information processing in biological and computational systems. *Commun. ACM,* 58, (1), 94-102, doi:10.1145/2678280.

Translational Health Informatics Continuum



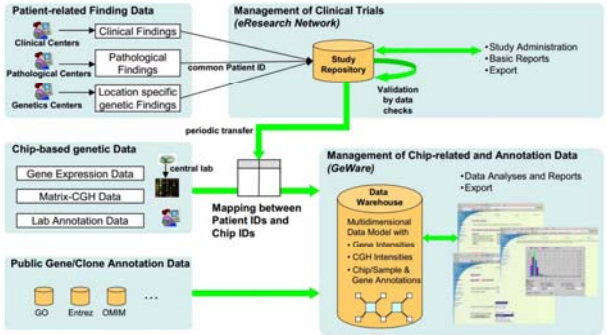
Sarkar, I. 2010. Biomedical informatics and translational medicine. Journal of Translational Medicine, 8, (1), 2-12.

Not only a problem at cell level ...



Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. (2011) *Big data: The next frontier for innovation, competition, and productivity. Washington (DC), McKinsey Global Institute.*

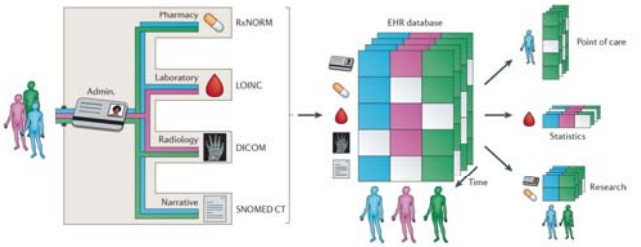
Example Data Integration Architecture



Kirsten, T., Lange, J. & Rahm, E. 2006. An integrated platform for analyzing molecular-biological data within clinical studies. Current Trends in Database Technology–EDBT 2006. Heidelberg: Springer, pp. 399-410, doi:10.1007/11896548_31.

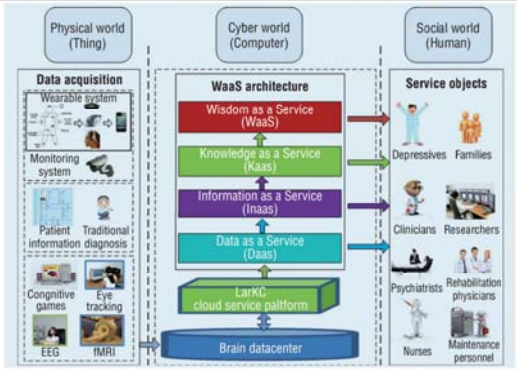
Example: Data Integration via health card

RxNorm is a US-specific terminology that contains all medications available on the US market, part of UMLS.



Jensen, P. B., Jensen, L. J. & Brunak, S. 2012. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics, 13, (6), 395-405.

Service-oriented computing example Zhong et al. (2015)



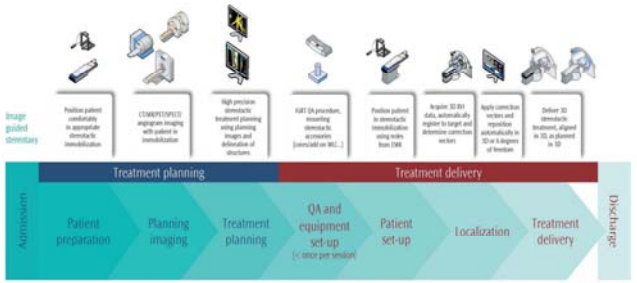
Zhong, N., Yau, S. S., Ma, J., Shimojo, S., Just, M., Hu, B., Wang, G., Oiwa, K. & Anzai, Y. 2015. Brain Informatics-Based Big Data and the Wisdom Web of Things. *Intelligent Systems, IEEE,* 30, (5), 2-7.

02 Clinical Pathways and Workflows

Example

http://www3.gehealthcare.co.uk/~/media/images/specialty/cardiology/acute-coronary-syndrome-pathway.jpg

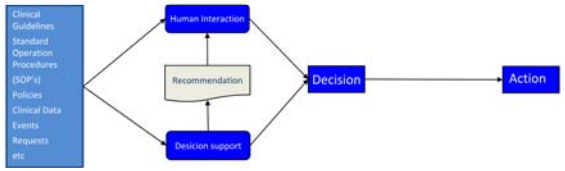
Slide 10-2 Typical Workflow in a hospital: Example Radiology



SRT = stereotactic radiation therapy
SRS = stereotactic radiosurgery

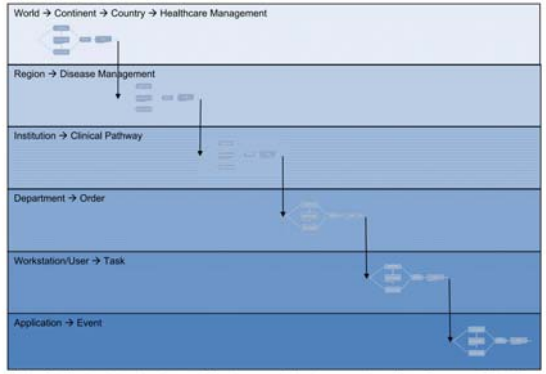
http://www.elekta.com/

Slide 10-3 Workflow > Interaction > Decision > Action



Colaert (2010): The Pain Points in Health Care and the Semantic Web, Advanced Clinical Application Research Group, Agfa HealthCare

Slide 10-4 Various Levels of Decision Support



Colaert (2010): The Pain Points in Health Care and the Semantic Web, Advanced Clinical Application Research Group, Agfa HealthCare

Slide 10-5 Workflow modeling in a nutshell

- Workflow modeling is the process of simplifying the real-world;
- Modeling is based on facts gathered during observations and we need to accept that this representation can never be perfect;
- Expectations from a model should be limited to the intentions with which it is designed for, be it problem solving or understanding of system intricacies (=elaborately complex details);

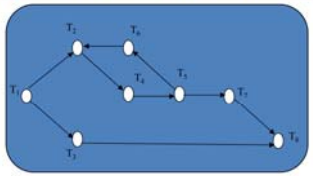
Malhotra, S., Jordan, D., Shortliffe, E. & Patel, V. L. (2007) Workflow modeling in critical care: Piecing together your own puzzle. *Journal of Biomedical Informatics, 40, 2, 81-92.*

Slide 10-6 Example: Formal workflow modeling 1/2

- A workflow is defined as a process that contains tasks T, and the respective rules on how those tasks are executed:
- Workflow W:= (T, P, C, A, S_0) where
  - $T = \{T_1, T_2, \ldots T_m\}$ A set of tasks, $m \geq 1$
  - $P = (p_{ij})_{m \times m}$ **Precedence matrix of the task set**
  - $C = (c_{ij})_{m \times m}$ **Conflict matrix of the task set**
  - $A = (A(T_1), A(T_2), \ldots, A(T_m))$ Pre-Condition set for each task
  - $S_0 \in \{0, 1, 2, 3\}_m$ is the initial state

J. Wang, D. Rosca, W. Tepfenhart & A. Milewski (2006) Dynamic Workflow Modeling and Analysis, Monmouth University

Slide 10-7 Example: Formal workflow modeling 2/2



$T = \{T_1, T_2, \ldots, T_8\}$,

$A(T_1) = \emptyset, A(T_2) = \{\{T_1\}, \{T_6\}\}, A(T_3) = \{\{T_1\}\}, A(T_4) = \{\{T_2\}\}, A(T_5) = \{\{T_4\}\}, A(T_6) = A(T_7) = \{\{T_5\}\}, A(T_8) = \{\{T_3, T_7\}\}.$

$S_0 = (1, 0, 0, 0, 0, 0, 0, 0).$

$$P = \begin{bmatrix} 0&1&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&1 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&1&1&0 \\ 0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&0&0 \end{bmatrix} \quad C = \begin{bmatrix} 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0 \end{bmatrix}$$

J. Wang, D. Rosca, W. Tepfenhart & A. Milewski (2006) Dynamic Workflow Modeling and Analysis, Monmouth University

Slide 10-8 Example: Modeling in UML – Use Case Diagram

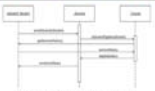
Name	Modify a wellness data entry
Purpose	The user edits the value of an existing wellness data entry
Summary	Clicking on an entry's row in the list of all wellness entries (except clicking on the info-icon) opens an input form allowing the user to modify and store the selected entry. After storing the entry on the server, an info message tells about success or possible failure. Instead of modifying the data, the user can also delete the entry using a button in this form.
Pre-conditions	The user is authenticated. The user has chosen an existing wellness data entry.
Post-conditions	The entry is updated or deleted on the server.

Holzinger, A., Dorner, S., Födinger, M., Valdez, A. C. & Ziefle, M. (2010) Chances of Increasing Youth Health Awareness through Mobile Wellness Applications. In: *Lecture Notes in Computer Science (LNCS 6389). Berlin, Heidelberg, Springer, 71-81.*

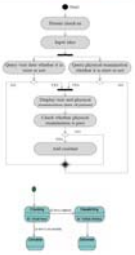
Slide 10-9 Modeling: Different views for different people

For Stakeholders **use case diagrams** describe sequences of actions, providing measurable values to actors

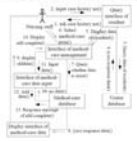
Class diagrams show interrelationships, inheritance, aggregation, association, operations and attributes



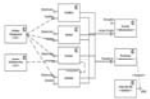
Sequence diagrams are interaction diagrams which details how operations are carried out, which messages are sent and at which time



Activity Diagrams for modelling the logic and **State Machine diagrams** show to analysers the possible states of objects and transitions that causes changes



Communication/Collaboration diagrams show the message flow between objects and imply the relationships between classes

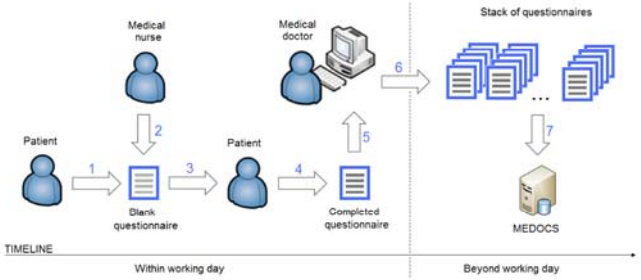


Component diagrams are used as an architecture-level artifact

Deployment diagrams model the static view of the run-time configuration of processing nodes and the components that run on those nodes; this shows also hardware & software requirements

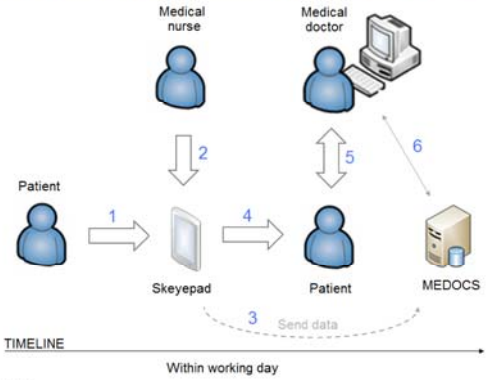
Juan, Y. C., Ma, C. M. & Chen, H. M. (2005) Applying UML to the development of medical care process management system for nursing home residents. *International Journal of Electronic Business Management, 3, 4, 322-330.*

Slide 10-10 Example Clinical Workflow



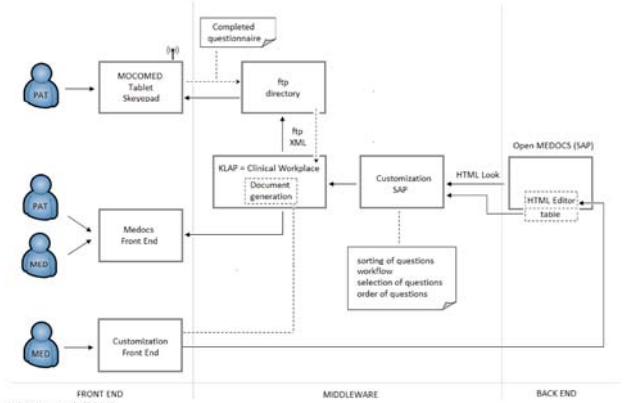
Holzinger, A., Kosec, P., Schwantzer, G., Debevc, M., Hofmann-Wellenhof, R. & Frühauf, J. (2011) Design and Development of a Mobile Computer Application to Reengineer Workflows in the Hospital and the Methodology to evaluate its Effectiveness. *Journal of Biomedical Informatics, 44, 6, 968-977.*

Slide 10-11 Example: WF-Optimization with mobile comp



Holzinger et al. (2011)

Slide 10-12 Mobile Computing Project MoCoMed-Graz



Holzinger et al. (2011)

Slide 10-13 Important: Macrolevel – the view of the Manager

Rechenmodell zum Ausfüllen eines Fragebogens in einer Ambulanz

ohne MoCoMed (Szenario 1)			Entwicklungs- und Betriebskosten MoCoMed	
Eingebendes Personal	Arzt (1 A)	Schreibkraft bzw Schalter (1 B)		
Personalkosten/h	€ 44,0	€ 19,0	Kosten pro Personenmonat Entwicklung (EU-Satz)	€ 4.000
Personalkosten/min	€ 0,73	€ 0,32	Anzahl Personenmonate Entwicklung einmalig	7
Anzahl Patienten/Tag	30	30	**Kosten Personal Entwicklung**	**€ 28.000**
Zeitbedarf pro Fragebogen in min.	10	7	investkosten pro Gerät	€ 1.600
Ambulantage pro Jahr	250	250	Anzahl Geräte	2
Jahreskosten Personal	**€ 55.000**	**€ 16.625**	**Kosten Geräte**	**€ 3.200**
			Summe Entwicklungskosten	**€ 31.200**
mit MoCoMed (Szenario 2)			Kosten pro Personenmonat Wartung und Betrieb	€ 4.000
		Schreibkraft bzw	Anzahl Personenmonate	
Eingebendes Personal	Arzt	Schalter	Wartung und Betrieb pro Gerät pro Jahr	0,25
Personalkosten/h	€ 44,0	€ 19,0	Anzahl Jahre (Lebenszyklus)	4
Personalkosten/min	€ 0,73	€ 0,32	**Kosten Wartung und Betrieb auf Lebenszyklus**	**€ 8.000**
Anzahl Patienten pro Tag	30	30	**Gesamtkosten auf Lebenszyklus**	**€ 39.200**
Zeitersparnis pro Fragebogen in %	90%	90%	**Gesamtkosten anteilig/Jahr**	**€ 9.800**
Zeitbedarf pro Fragebogen in min.	1,00	0,70		
Gerätemanipulationszeit pro Fragebogen in min.		1,00		
Ambulantage	250	250		
Jahreskosten Personal	**€ 7.875**	**€ 4.038**		
+ Gesamtkosten MoCoMed/Jahr	€ 9.800	€ 9.800		
= Jahreskosten TOTAL	**€ 17.675**	**€ 13.838**		
Ergebnis				
Eingebendes Personal	Arzt	Schreibkraft bzw Schalter		
Einsparungspotential/Jahr in EUR unter den oa. Annahmen:	**€ 37.325**	**€ 2.788**		
Einsparungspotential/Jahr in Std. unter den oa. Annahmen:	**848**	**147**		

Holzinger, A., Kosec, P., Schwantzer, G., Debevc, M., Hofmann-Wellenhof, R. & Frühauf, J. (2011) Design and Development of a Mobile Computer Application to Reengineer Workflows in the Hospital and the Methodology to evaluate its Effectiveness. *Journal of Biomedical Informatics, 44, **6, 968-977.***

Slide 10-14 MoCoMED-Graz in operation

Holzinger et al. (2011)

Slide 10-15 MoCoMedGraz Project – Benefit Summary

- This project was evaluated from a three level perspective: **patients – doctors – managers**
- Theoretical Background: Bronfenbrenner model
- Project demonstrates how both workflows and information quality can be optimized;
- thereby raising both medical quality and patient empowerment;
- Success factors for applications within the hospital include: **simplicity, usability, reliability**

Holzinger, A., Kosec, P., Schwantzer, G., Debevc, M., Hofmann-Wellenhof, R. & Frühauf, J. (2011) Design and Development of a Mobile Computer Application to Reengineer Workflows in the Hospital and the Methodology to evaluate its Effectiveness. *Journal of Biomedical Informatics, 44, **6, 968-977.***

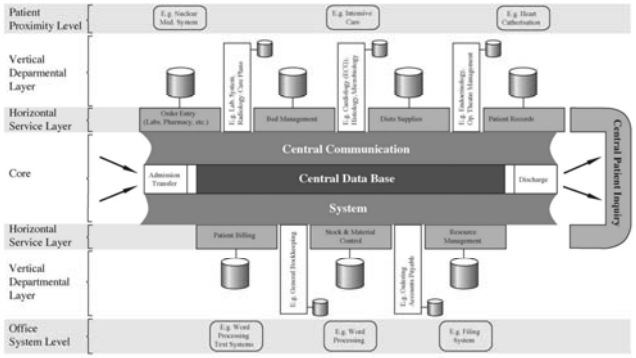
- A data-driven procedure consisting of one or more transformation processes -> nodes;
- Can be represented as a **directed graph;**
  - **Direction is time** – i.e. the order of transformations;
  - A set of transformation rules;
- The data flow origins from a source to a destination (or result) via a series of data manipulations;
- The specification is designed in a **Workflow Design System** (modeling component) and then run by a **Workflow Management System** (execution component).

Hasan, S., Daugelat, S., Rao, P. S. S. & Schreiber, M. (2006) Prioritizing genomic drug targets in pathogens: application to Mycobacterium tuberculosis. *PLoS Computational Biology, 2, 6, e61.*

- Def.: WMS = a system that defines, creates and manages the execution of workflows. Its main components include:
- 1) a graphical interface for composing workflows, entering data, watching execution, displaying results;
- 2) an data archive to store workflow descriptions, results of executions and related traces;
- 3) a registry of available services, either local or remote,
- 4) a scheduler able to invoke services included in the workflow at the appropriate time,
- 5) a set of programming interfaces able to dialogue with remote services,
- 6) a monitor tool for controlling the execution of the
- workflow,
- 7)a set of visualization capabilities for displaying different types of results.

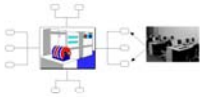
Romano, P. (2008) Automation of in-silico data analysis processes through workflow management systems. *Briefings in Bioinformatics, 9, 1, 57-68.*

03
Hospital Information Systems (HIS)



Reichertz, P. L. (2006) Hospital information systems - Past, present, future. *International Journal of Medical Informatics, 75, 3-4, 282-299.*

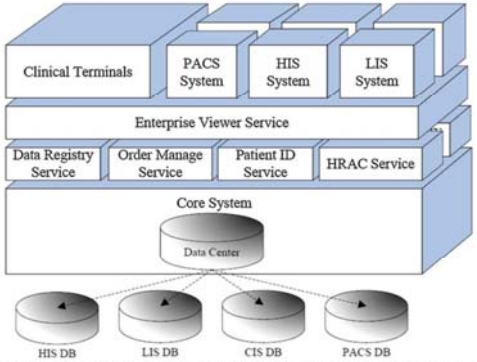
1970+ **"Vertical Approach"** – monolithic mainframes
Central computer systems mainly for accounting, typical "data processing" ("EDV")



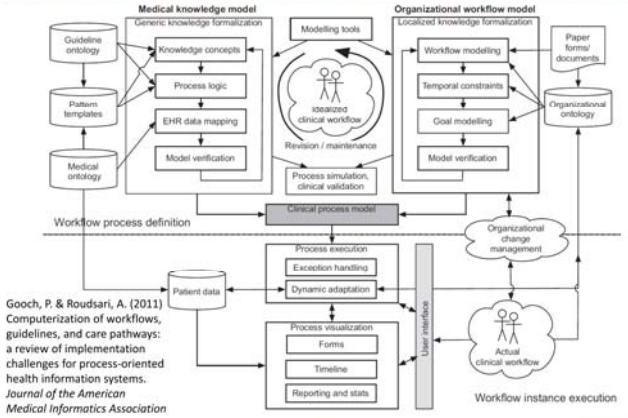
1985+ **"Horizontal Approach"** – evolutionary systems
Departmental clinical information systems, local area networks, distributed systems

2000+ **"Integrated Approach"** – open, distributed systems
Hospital Intranets, electronic patient/health record, mobile computing, "information quality focus"

Holzinger, A. (2002) *Basiswissen IT/Informatik Band 1: Informationstechnik. Das Basiswissen für die Informationsgesellschaft des 21. Jahrhunderts. Wuerzburg, Vogel Buchverlag.*

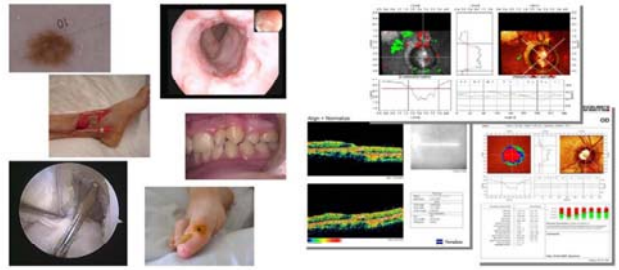


El Azami, I., Cherkaoui Malki, M. O. & Tahon, C. (2011) Integrating Hospital Information Systems in Healthcare Institutions: A Mediation Architecture. *Journal of Medical Systems, (in print), 1-12.*



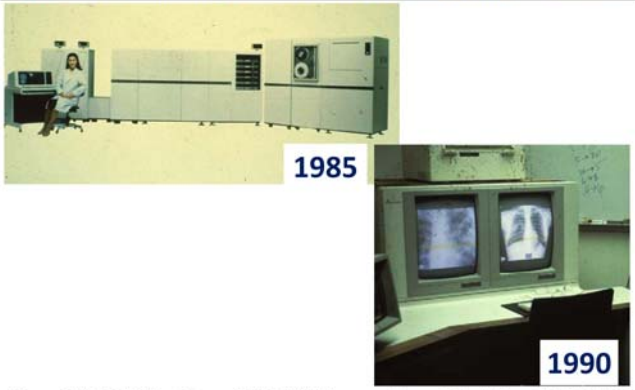
Gooch, P. & Roudsari, A. (2011) Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems. *Journal of the American Medical Informatics Association*

04
Multimedia Data in the Hospital: PACS



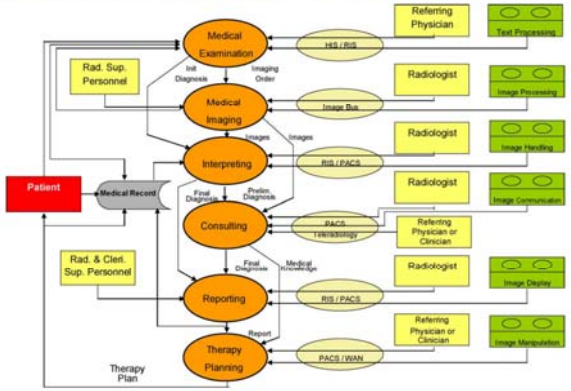
Bellon, E., Feron, M., Deprez, T., Reynders, R. & Van den Bosch, B. (2011) Trends in PACS architecture. *European Journal of Radiology, 78, 2, 199-204.*

Slide 10-23 1st commercial Picture Archiving & Comm. System



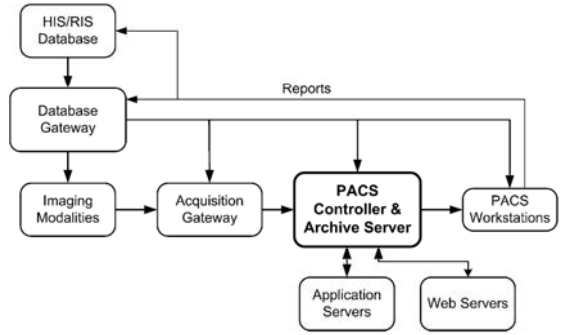
Huang, H. K. (2011) Short history of PACS (USA). *European Journal of Radiology, 78, 2, 163-176.*

Slide 10-24 Typical Workflow in a European PACS



Lemke, H. (2011) Short history of PACS (Europe). *European Journal of Radiology, 78, 2, 177-183.*

Slide 10-25 Generic PACS components and data flow



Huang, H. K. (2011) Short history of PACS (USA). *European Journal of Radiology, 78, 2, 163-176.*

05

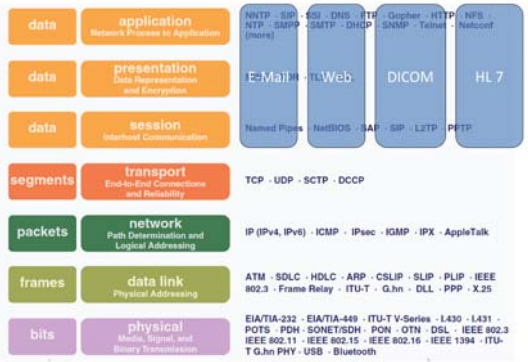
Data Standards for Communication and Representation

Slide 10-26 Data Standards for Comm. & Representation

- **DICOM 3.0 – Digital Imaging and Communication in Medicine** (1993)
  - 1) a set of protocols for network communication;
  - 2) a syntax and semantics for commands and info;
  - 3) a set of media storage services (standard compliant);
- **HL 7 - Health Level 7**
  - 1) HL 7 v2.x messaging protocol, to provide exchange of textual healthcare data between hospital information systems;
  - 2) Reference Implementation Model (RIM) contains data types, classes, state diagrams, use case models, ant terminology to derive domain-specific information models;
  - 3) Clinical Document Architecture (CDA) is a document markup standard to specify structure and semantics of clinical documents in XML;
- **LOINC - Logical Observation Identifier Names and Codes**
  - 1) Laboratory data (e.g. molecular pathology observations used for identification of genetic mutations, tumor genes, gene deletions, etc.
  - 2) Clinical Observations (e.g. non-laboratory diagnostic studies, critical care, nursing measures, patient history, instrument surveys, etc.
  - 3) Claims attachments (e.g. handles the definition of new LOINC terms and codes to manage claims-related data

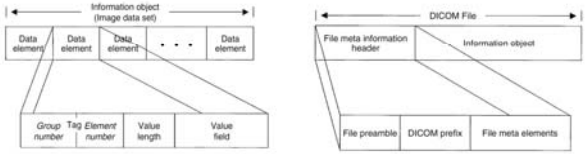
Bui, A. A. T. & Taira, R. K. (2010) *Medical Imaging Informatics. New York, Heidelberg, London, Springer.*

Slide 10-27 DICOM and HL 7



Williams, R. (2011) Network Concepts.
In: *An Introduction to Trading in the Financial Markets. San Diego, Academic Press, 201-225.*

Slide 10-28 DICOM Object containing multiple data elements

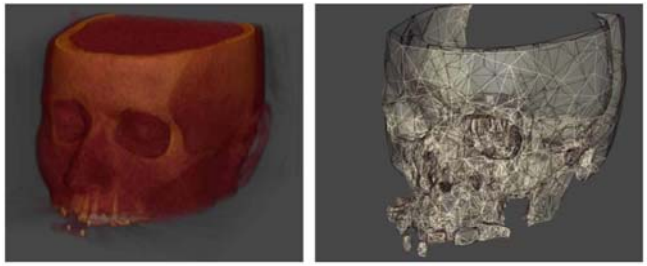


Each data element is uniquely identified by its corresponding tag composed of a group number and an element number. Pixel data of the image is stored in element 0010 within group 7FE0.

A DICOM file consists of a file meta-information header and an information object (image data set). The file meta-information header is made of a file preamble, a DICOM prefix, and multiple file meta-elements.

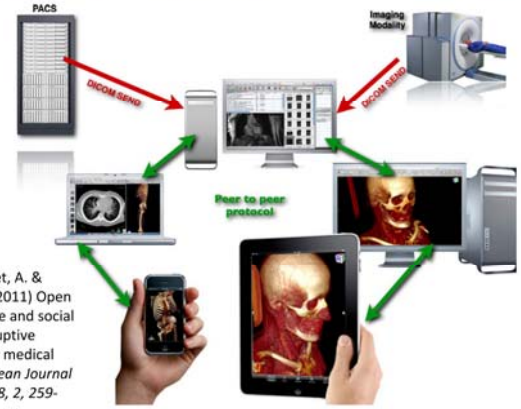
Wong, A. & Lou, S. L. (2009) Medical Image Archive, Retrieval, and Communication. In: Isaac, N. B. (Ed.) *Handbook of Medical Image Processing and Analysis (Second Edition). Burlington, Academic Press, 861-873.*

Slide 10-29 DICOM instances allowing the storage of meshes



Lemke, H. U. & Berliner, L. (2011) PACS for surgery and interventional radiology: Features of a Therapy Imaging and Model Management System (TIMMS). *European Journal of Radiology, 78, 2, 239-242.*

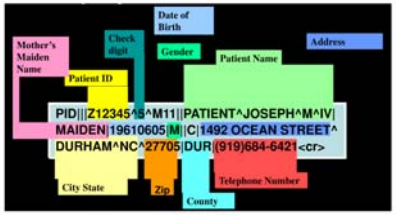
Slide 10-30 Image distribution workflow



Ratib, O., Rosset, A. & Heuberger, J. (2011) Open Source software and social networks: Disruptive alternatives for medical imaging. *European Journal of Radiology, 78, 2, 259-265.*

Result delivered by Dr. Howard H. Hippocrates. The lab test was ordered by Dr. Patricia Primary for patient Eve E. Everywoman. The use case takes place in the US Realm.

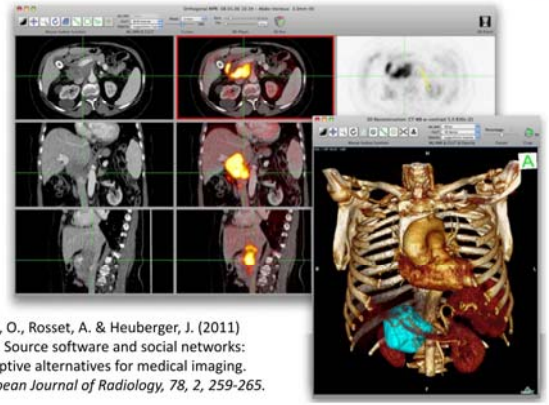
```
MSH|^~\&|GHH LAB|ELAB-3|GHH OE|BLDG4|200202150930||ORU^R01|CNTRL-3456|P|2.4<cr>
PID|||555-44-4444||EVERYWOMAN^EVE^E^^^^L|JONES|19620320|F|||153 FERNWOOD DR.^
^STATESVILLE^OH^35292||(206)3345232|(206)752-121|||AC555444444||67-A4335^OH^20030520<cr>
OBR|1|845439^GHH OE|1045813^GHH LAB|15545^GLUCOSE|||200202150730||||||||
555-55-5555^PRIMARY^PATRICIA P^^^^MD^^|||||||F|||||444-44-4444^HIPPOCRATES^HOWARD H^^^^MD<cr>
OBX|1|SN|1554-5^GLUCOSE^POST 12H CFST:MCNC:PT:SER/PLAS:QN||^182|mg/dl|70_105|H|||F<cr>
```



| - FIELD DELIMITER
^ - SUB FIELD DELIMITER
& - SUB-SUB FIELD DELIMITER
~ - SEPARATOR - REPEATING FIELDS
\- ESCAPE CHARACTER

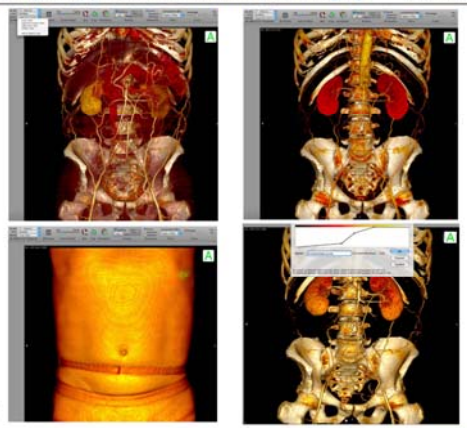
Examples: http://www.ringholm.com/docs/04300_en.htm

Ratib, O., Rosset, A. & Heuberger, J. (2011) Open Source software and social networks: Disruptive alternatives for medical imaging. *European Journal of Radiology, 78, 2, 259-265.*

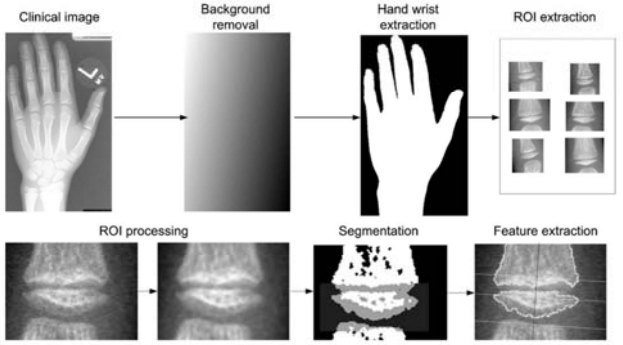


Ratib, O., Rosset, A. & Heuberger, J. (2011) Open Source software and social networks: Disruptive alternatives for medical imaging. *European Journal of Radiology, 78, 2, 259-265.*

Faggioni, L., Neri, E., Castellana, C., Caramella, D. & Bartolozzi, C. (2011) The future of PACS in healthcare enterprises. *European Journal of Radiology, 78, 2, 253-258.*



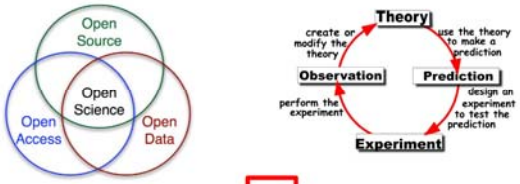
Faggioni et al. (2011)



Piętka, E., Kawa, J., Spinczyk, D., Badura, P., Więcławek, W., Czajkowska, J. & Rudzki, M. (2011) Role of radiologists in CAD life-cycle. *European Journal of Radiology, 78, 2, 225-233.*

06 Towards Open …

- Billions of biological data sets are openly available, here only some examples:
- General Repositories:
  - GenBank, EMBL, HMCA, …
- Specialized by data types:
  - UniProt/SwissProt, MMMP, KEGG, PDB, …
- Specialized by organism:
  - WormBase, FlyBase, NeuroMorpho, …
- Details: http://hci-kdd.org/open-data-sets



License	Apache	BSD/MIT	GPL	LGPL	MPL/CDDL	CPL/EPL
Reciprocity	No	No	Yes	Maybe	No	No
Modification release	No	No	Yes	Yes	Yes	Yes
Patent	Yes	No	No	No	Yes	Yes
Jurisdiction	Silent	Silent	Silent	Silent	California	New York
Freedom	PR	Free	PR	PR	Free	PR

Braun, M. L. & Ong, C. S. 2014. Open Science in Machine Learning. Implementing Reproducible Research, 343.

Holzinger, A. 2010. Process Guide for Students for Interdisciplinary Work in Computer Science/Informatics. Second Edition, Norderstedt, BoD.

- Grand Challenges in this area:
- – Production of Open Data Sets
- – Synthetic data sets for learning algorithm testing
- – Privacy preserving machine learning
- – Data leak detection
- – Data citation
- – Differential privacy
- – Anonymization and pseudonymization
- – Evaluation and benchmarking

Please visit:
http://hci-kdd.org/privacy-aware-machine-learning-for-data-science/

87 % of the population in the USA can be uniquely re-identified by Zip-Code, Gender and date of birth

Hospital Patient Data

Birthdate	Sex	Zipcode	Disease
1/21/76	Male	53715	Flu
4/13/86	Female	53715	Hepatitis
2/28/76	Male	53703	Brochitis
1/21/76	Male	53703	Broken Arm
4/13/86	Female	53706	Sprained Ankle
2/28/76	Female	53706	Hang Nail

Voter Registration Data

Name	Birthdate	Sex	Zipcode
Andre	1/21/76	Male	53715
Beth	1/10/81	Female	55410
Carol	10/1/44	Female	90210
Dan	2/21/84	Male	02174
Ellen	4/19/72	Female	02237

Samarati, P. 2001. Protecting respondents identities in microdata release. IEEE Transactions on Knowledge and Data Engineering, 13, (6), 1010-1027, doi:10.1109/69.971193.

Sweeney, L. 2002. Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10, (05), 571-588.

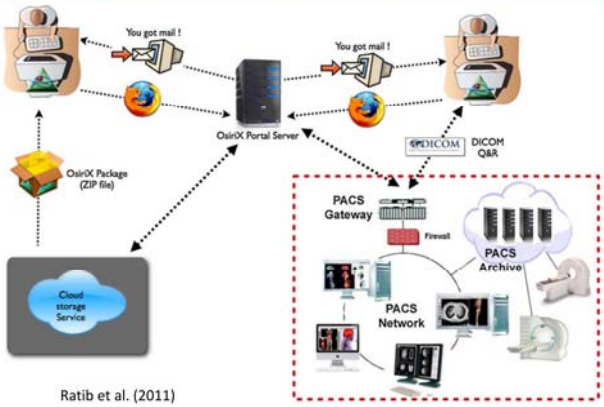
open source ≠ FDA CE

- Attention: Medical certifications such as FDA in the US and CE marking in Europe **do not apply** to Open Source software;
- These certifications require a legal commercial entity to be identified as the owner of the product and warrant the legal liability of its distribution and commercial support.
- Open Source software being often developed outside commercial enterprises; such as academic groups or university research labs, do not have the proper legal structure to apply for such certifications.
- Also, most Open Source products being distributed free of charge lack the legal binding between the provider and the user that is required for software distribution under FDA and CE certification.

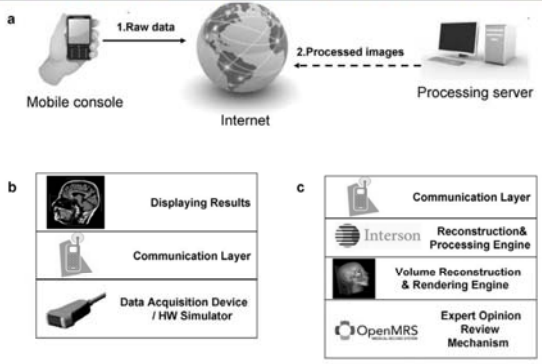
Ratib et al. (2011)

Conclusion and Future Outlook

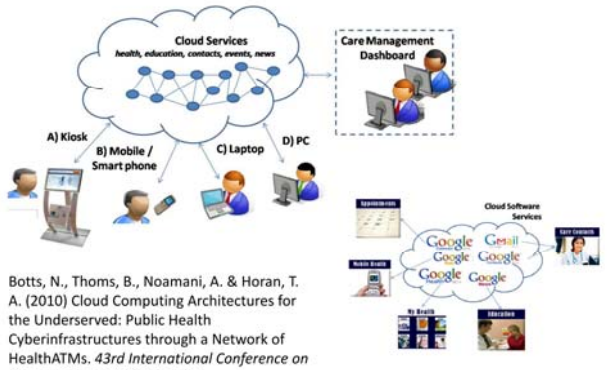
- 1) the world (even the hospital world ;-) of tomorrow will be **mobile …**
- 2) the data will be stored in the cloud as well as software-as-a-service will be used (Key problem: **Privacy, Security, Safety and Data Protection).**
- 3) Search tools will be integrated within the clinical workplace to provide a quick **overview first,** and then zoom and filter on demand …
- 4) Integrated ML will enable **knowledge discovery with the doctor-in-the-loop …**



Ratib et al. (2011)



Meir, A. & Rubinsky, B. (2009) Distributed Network, Wireless and Cloud Computing Enabled 3-D Ultrasound; a New Medical Technology Paradigm. *Plos One, 4, 11, e7974.*



Botts, N., Thoms, B., Noamani, A. & Horan, T. A. (2010) Cloud Computing Architectures for the Underserved: Public Health Cyberinfrastructures through a Network of HealthATMs. *43rd International Conference on System Sciences (HICSS). 1-10.*

Instead of pure databases learning and knowledge extraction

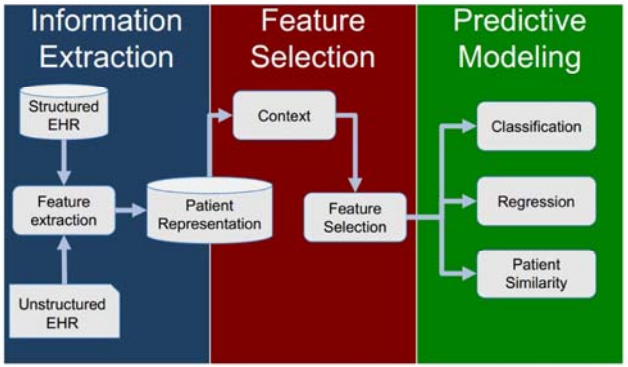


Image credit to Jimeng Sun and Chandan K. Reddy

Thank you!

Questions

Sample Questions (1)

- How is a workflow defined?
- How can a workflow be described formally?
- Why is workflow modeling important in health care?
- Please describe the different tools of the Unified Modeling Language (UML) on some medical examples!
- Which benefits can be gained by optimization of workflows?
- Which three stakeholders have which interests within an Hospital?
- Please describe the basic idea of a typical bioinformatics workflow management system!
- What is the difference of system quality versus information quality?
- What are the advantages/disadvantages of the three basic system architecture approaches of hospital information systems?
- Which functional parts does the classic conceptual model of a HIS include?
- In which aspects does the process-oriented health information systems model differ from the classic conceptual model?
- Please describe the typical workflows within a PACS System?
- What are the typical modalities of a PACS System?

Sample Questions (2)

- What are the generic PACS components and the respective data flows?
- What are the typical advantages/disadvantages of the six PACS implementation models?
- Why are communication standards important for biomedical informatics?
- Please describe the purpose and advantages of DICOM?
- What is the basic idea of HL7?
- Why is open source software problematic in the medical domain?
- What are the advantages/disadvantages of cloud computing for health care?
- What are the advantages/disadvantages of the paradigm "Software as a Service"?
- What is an electronic Personal Health Record?
- Which is still the biggest problem of such PHRs?

Appendix

Appendix: Bioinformatics Workflow Management Systems

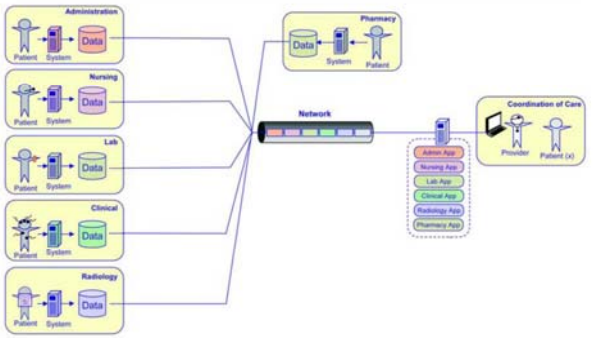
- **BioExtract** = for creating and customizing workflows; you can query online sequence data, analyze it using an array of informatics tools, create and share custom workflows for repeated analysis, and save the resulting data and workflows in standardized reports; http://www.bioextract.org
- **CellProfiler** = open source modular image analysis software developed at the Broad Institute; algorithms for image analysis; http://www.cellprofiler.org
- **Discovery Net** = was one of the earliest examples of scientific workflow systems (e-Science project by the Imperial College London), having many features, e.g. chemical compounds represented in the widely used SMILES (Simplified molecular input line entry specification) format can be imported and rendered using three-dimensional representation or the structural formula, of historic interest, see: http://www.computer.org/portal/web/csdl/doi/10.1109/HPDC.2002.1029946
- **Ergatis** = to create, run, and monitor reusable computational analysis pipelines, contains pre-built components for common bioinformatics analysis tasks. These components can be arranged graphically to form highly-configurable pipelines. Each analysis component supports multiple output formats, including the Bioinformatic Sequence Markup Language (BSML), http://ergatis.sourceforge.net
- **GenePattern** = genomic analysis platform that provides access to 150+ tools for gene expression analysis, proteomics, SNP analysis, flow cytometry, RNA-seq analysis as well as standard data processing tasks. A web-based interface provides easy access to these tools and allows the creation of multi-step analysis pipelines that enable reproducible in silico research; http://www.broadinstitute.org/cancer/software/genepattern
- **Triana** = open source problem solving environment developed at Cardiff University that combines an intuitive visual interface with powerful data analysis tools. Already used by scientists for a range of tasks, such as signal, text and image processing, Triana includes a large library of pre-written analysis tools and the ability for users to easily integrate their own tools. is an open source problem solving environment developed at Cardiff University that combines an intuitive visual interface with powerful data analysis tools; http://www.trianacode.org

Some useful links

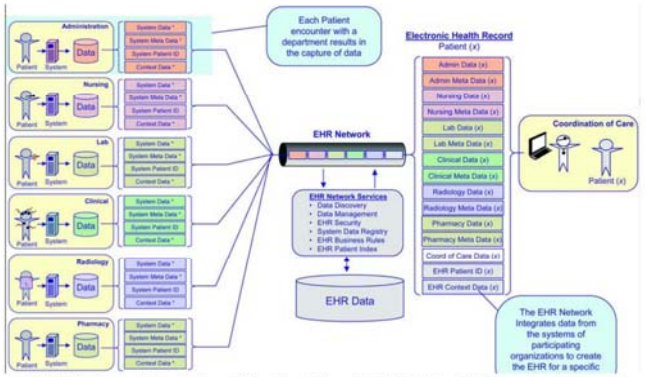
- http://www.gimias.org/download/sampledata (some useful sample data)
- https://www.biomedtown.org/biomed_town/MSV/reception/wikis/Data (on-line community open and free to anyone has a professional or educational interest in biomedical research & practice)
- http://rad.usuhs.edu/medpix/index.html (Medical Image Database)
- http://www.incits.org (International Committee for Information Technology Standards)
- http://medical.nema.org (about DICOM)
- http://www.aycan.de/main/lp/dicom-bilder-zum-download.html (DICOM examples for download)
- http://www.ringholm.com/docs/04300_en.htm (HL7 Message Examples)
- http://www.openehr.org (Open HER – good UML examples)
- http://www.sparxsystems.com/uml-tutorial.html (UML 2.0 Tutorial)
- http://www.agilemodeling.com/essays/umlDiagrams.htm (Excellent description of UML diagrams)
- http://www.wfmc.org (Workflow Management Coalition)

Appendix: Open Source in medical imaging

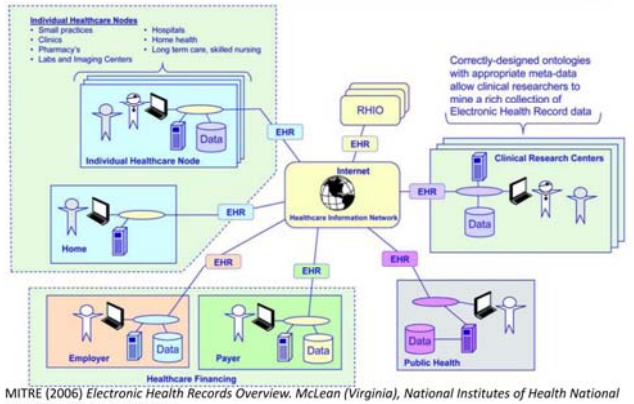
http://www.idoimaging.com



MITRE (2006) *Electronic Health Records Overview. McLean (Virginia), National Institutes of Health National Center for Research Resources.*

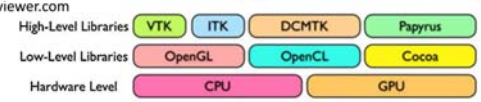


MITRE (2006) *Electronic Health Records Overview. McLean (Virginia), National Institutes of Health National Center for Research Resources.*

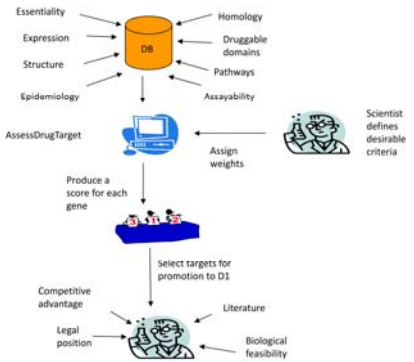


MITRE (2006) *Electronic Health Records Overview. McLean (Virginia), National Institutes of Health National Center for Research Resources.*

http://www.osirix-viewer.com



Rosset, A., Spadola, L. & Ratib, O. (2004) OsiriX: an open-source software for navigating in multidimensional DICOM images. *Journal of Digital Imaging, 17,* ***3, 205-216.***



Hasan, S., Daugelat, S., Rao, P. S. S. & Schreiber, M. (2006) Prioritizing genomic drug targets in pathogens: application to Mycobacterium tuberculosis. *PLoS Computational Biology, 2, 6, e61.*

http://screencast.g2.bx.psu.edu/pacBio-demo

Goecks, J., Nekrutenko, A., Taylor, J. & Team, T. G. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology, 11, 8, R86.*

Blankenberg, D., Kuster, G. V., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A. & Taylor, J. (2010) Galaxy: A Web-Based Genome Analysis Tool for Experimentalists. *Current Protocols in Molecular Biology. John Wiley & Sons, Inc.*

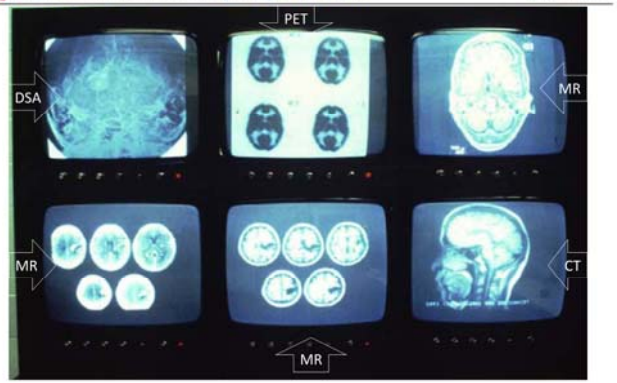
- Experts consider health IT key for improving efficiency and quality of health care processes.
- Please, carefully determine between:
- Quality of Systems
  and
- Quality of Data
- Quality of Information
- Quality of Knowledge
- Quality of Action

Chaudhry, B., Wang, J., Wu, S. Y., Maglione, M., Mojica, W., Roth, E., Morton, S. C. & Shekelle, P. G. (2006) Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine, 144,* ***10, 742-752.***

Decade	R&D progress	R&D topics
1980s Late 1980s	Medical imaging technology development Imaging systems integration	CR, MRI, CT, US, DR, WS, storage, networks PACS, ACR/NEMA, DICOM, high-speed networks
Early 1990s Late 1990s – present 2000s – present	Integration of HIS/RIS/PACS Workflow & application servers Imaging Informatics	DICOM, HL7, Intranet and Internet IHE, ePR, enterprise PACS, Web-based PACS Computer-aided diagnosis (CAD), image contents indexing, knowledge base, decision support tools, image-assisted diagnosis and treatment

- CR = Computed Radiography (vs. Direct Radiography (DR)); MRI = Magnetic resonance imaging; CT = Computed Tomography; US = Ultrasonography; WS = Web services;
  ACR= American College of Radiology; NEMA = National Electrical Manufacturers Association;

Huang, H. K. (2011) Short history of PACS (USA). *European Journal of Radiology, 78, 2, 163-176.*



Huang, H. K. (2011) Short history of PACS (USA). *European Journal of Radiology, 78, 2, 163-176.*

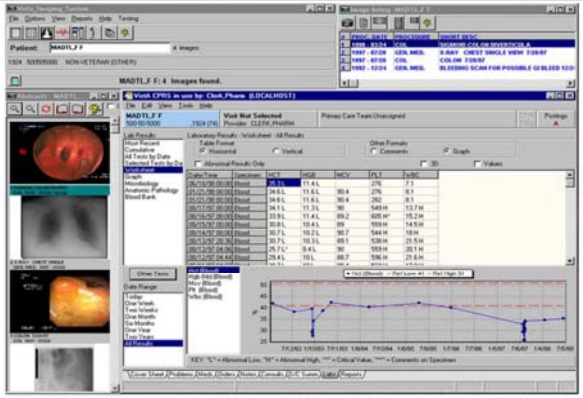
Pros/Cons of six PACS implementation models

Method	Advantages	Disadvantages
1. Home-grown system	Built to specifications, state-of-the-art technology, continuously upgrading, not dependent on a single manufacturer	Difficult to assemble a team, one-of-a-kind system, difficult to service and maintain
2. Two-team effort	Specifications written for a certain clinical environment, implementation delegated to the manufacturer	Specifications over ambitious, underestimate technical and operational difficulty, manufacturer lacks clinical experience, expensive
3. Turnkey	Lower cost, easier maintenance	Too general, not state-of-the-art technology
4. Partnership	System will keep up with technology advancement, health center does not have to worry of the system becoming obsolete, manufacturer has long-term contract to plan ahead	Expensive to the health center, manufacturer may not want to sign a partnership contract with a lesser prominent center, center has to consider the longevity and stability of the manufacturer
5. ASP	Minimize initial capital cost, may accelerate potential return on investment, no risk in technology obsolescence, provide flexible growth, no space requirement in data center	More expensive over 2–4-year time frame comparing to a capital purchase, customer has no ownership in equipment
6. Open source	Healthcare provider purchases its computer and communication equipment, use open source software, good for special PACS application server, lower cost	Open source software may not be robust for daily clinical use, maintenance and upgrade of the software may be a problem, may not be good for a full large-scale PACS

ASP = application service provider model

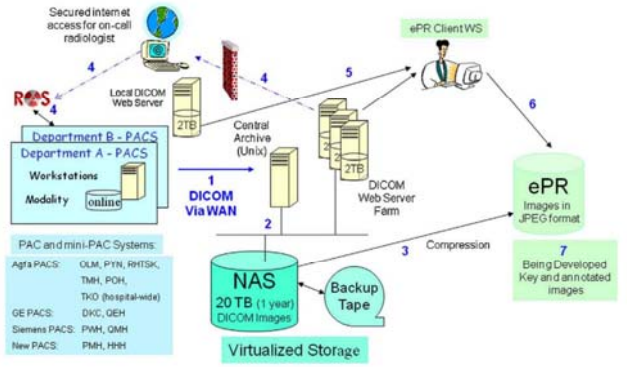
Huang, H. K. (2011) Short history of PACS (USA). *European Journal of Radiology, 78, 2, 163-176.*

Example: Images within the EPR



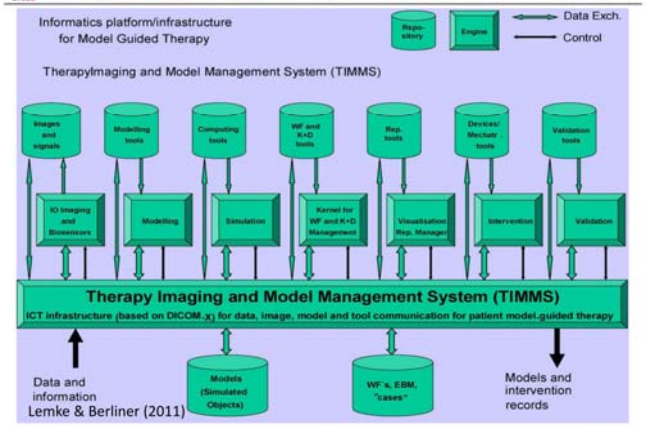
Huang, H. K. (2011) Short history of PACS (USA). *European Journal of Radiology, 78, 2, 163-176.*

Ex.: The Hong Kong Hospital Clinical Information System



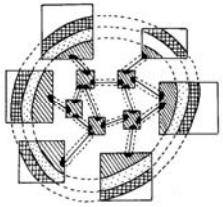
Huang, H. K. (2011) Short history of PACS (USA). *European Journal of Radiology, 78, 2, 163-176.*

Therapy Imaging and Model Management System: TIMMS

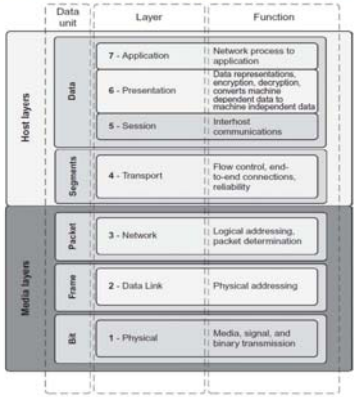


Lemke & Berliner (2011)

ISO/OSI Layer Model (ISO 7498) Communication Standards



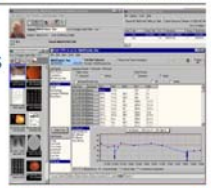
Zimmermann, H. (1980) OSI reference model-The ISO model of architecture for open systems interconnection. *IEEE Transactions on Communications, 28,* ***4, 425-432.***



Paulsen, K. (2011) Networked Storage. *Moving Media Storage Technologies. Boston, Focal Press, 429-467.*

Big Fact: Personal Health Records: Poor Usability

- 300+ different systems in the US
- Only 1% of the US population have access (2008)
- "Most EHR are 100 % bad." (Forrester Research, 2008)
- Biggest obstacles:
  - Cost factors
  - Poor usability
- Need for UCD
- Need for proactive wellness management



Alagoez, F., Valdez, A. C., Wilkowska, W., Ziefle, M., Dorner, S. & Holzinger, A. (2010) From cloud computing to mobile Internet, from user focus to culture and hedonism: The crucible of mobile health care and Wellness applications. *5th International Conference on Pervasive Computing and Applications (ICPCA). IEEE, 38-45.*

Degree of Integration