Hasso Plattner Institut

IT Systems Engineering | Universität Potsdam

Data Sources and Formats

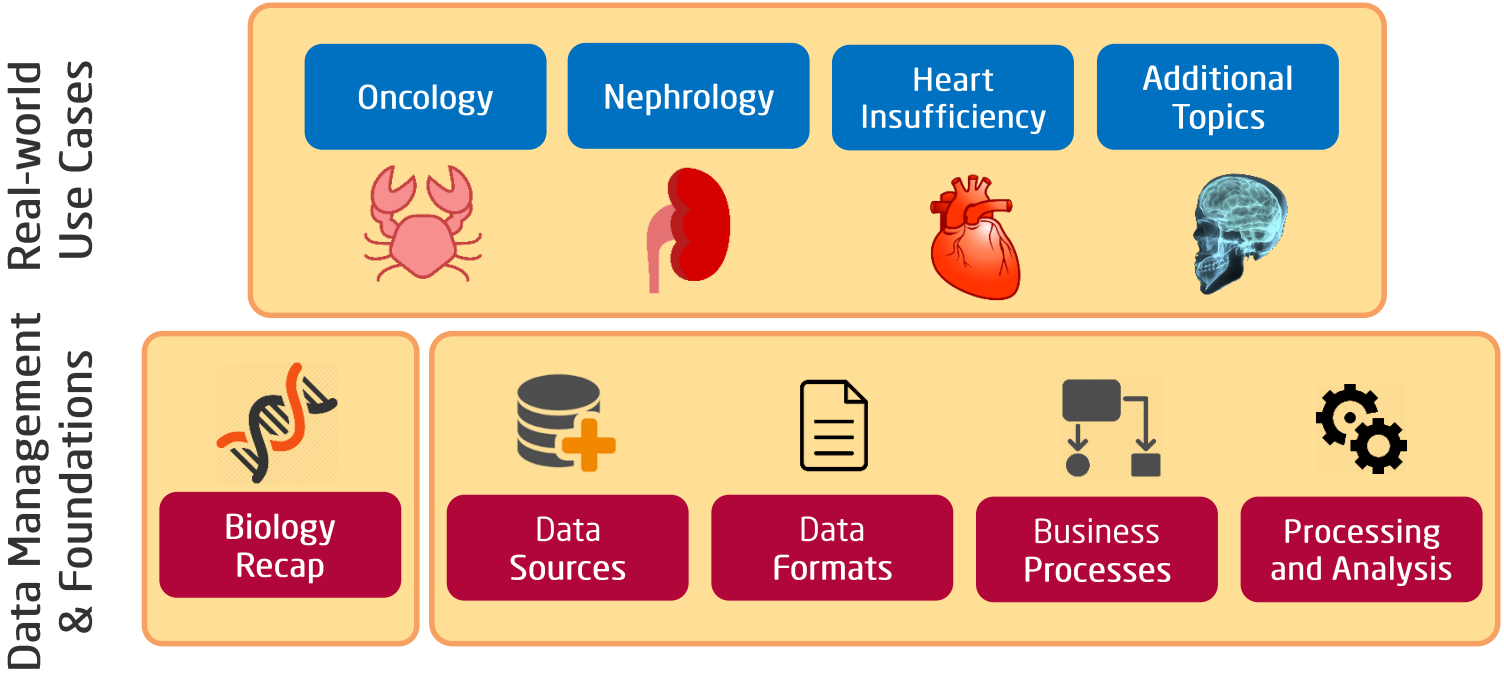
Dr. Matthieu-P. Schapranow

Data Management for Digital Health

Summer 2017

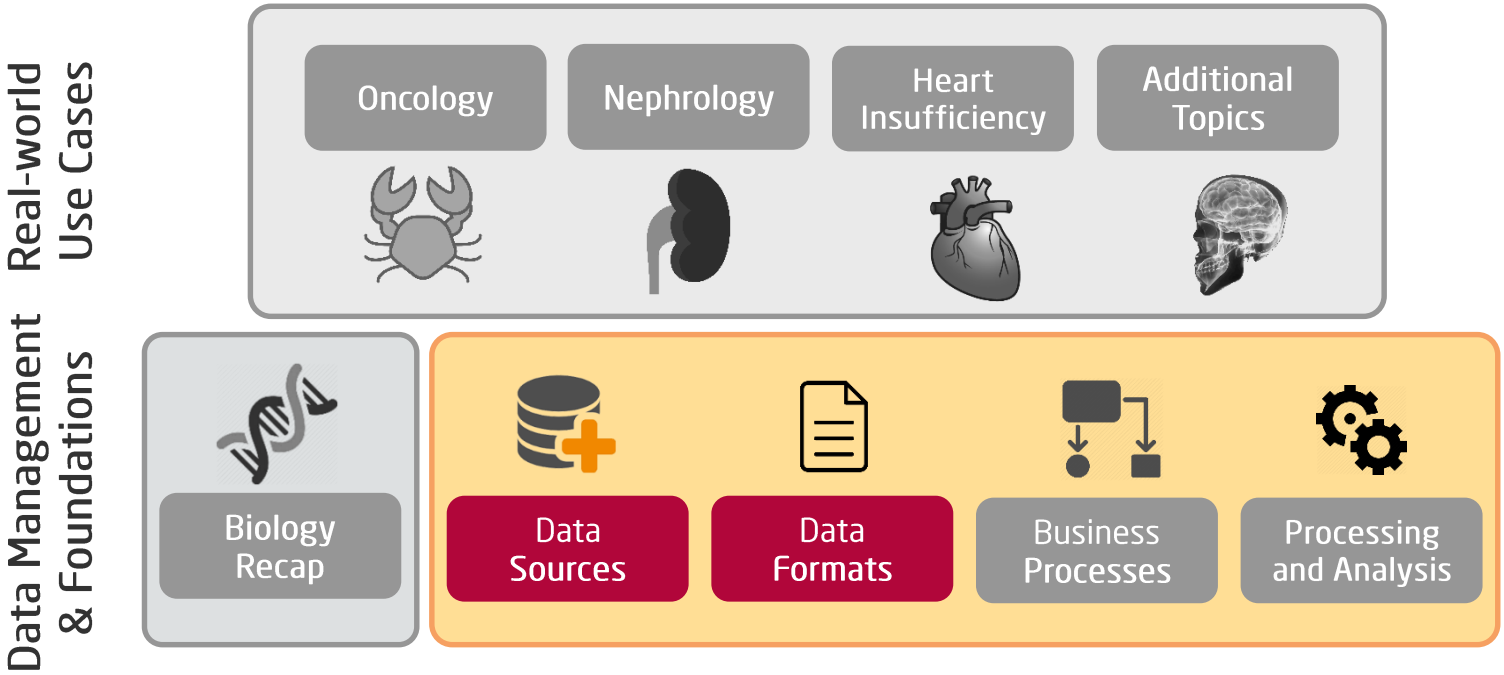
Where are we?

Hasso Plattner Institut



Agenda

Hasso Plattner Institut



IT Challenges
Distributed Heterogeneous Data Sources

Hasso Plattner Institut

Human genome/biological data
600GB per full genome
15PB+ in databases of leading institutes

Human proteome
160M data points (2.4GB) per sample
>3TB raw proteome data in ProteomicsDB

Hospital information systems
Often more than 50GB

PubMed database
>23M articles

Medical sensor data
Scan of a single organ in 1s
creates 10GB of raw data

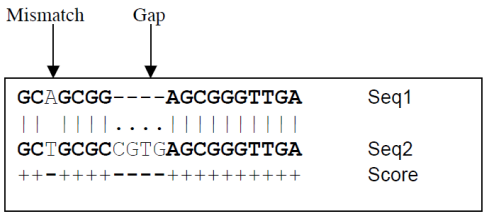
Cancer patient records
>160k records at NCT

Prescription data
1.5B records from 10,000 doctors and
10M Patients (100 GB)

Clinical trials
Currently more than 30k
recruiting on ClinicalTrials.gov

Recap Data Acquisition

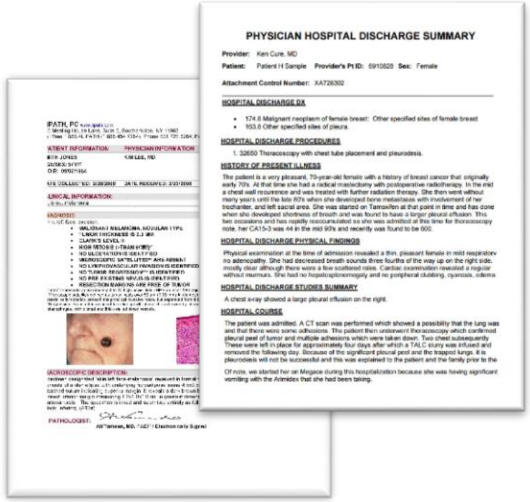
- Sequencing technologies
- Longitudinal data
- Sensor data
- Text documents
- Processing examples
- Alignment algorithms



http://www.hypothesisjournal.com

A. Rhoads and K. F. Au: PacBio Sequencing and Its Applications (2015)

SureSigns VM6 Vital Signs Patient Monitor w/ECG



www.sampletemplates.com

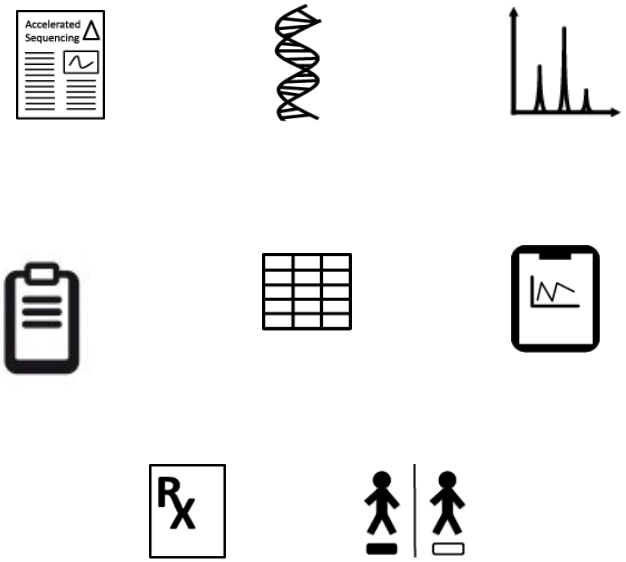
Data Sources and Formats

Agenda

- Data formats and sources
- Genomic (-omics data)
- Medical or hospital setting

Health Care Data Characteristics

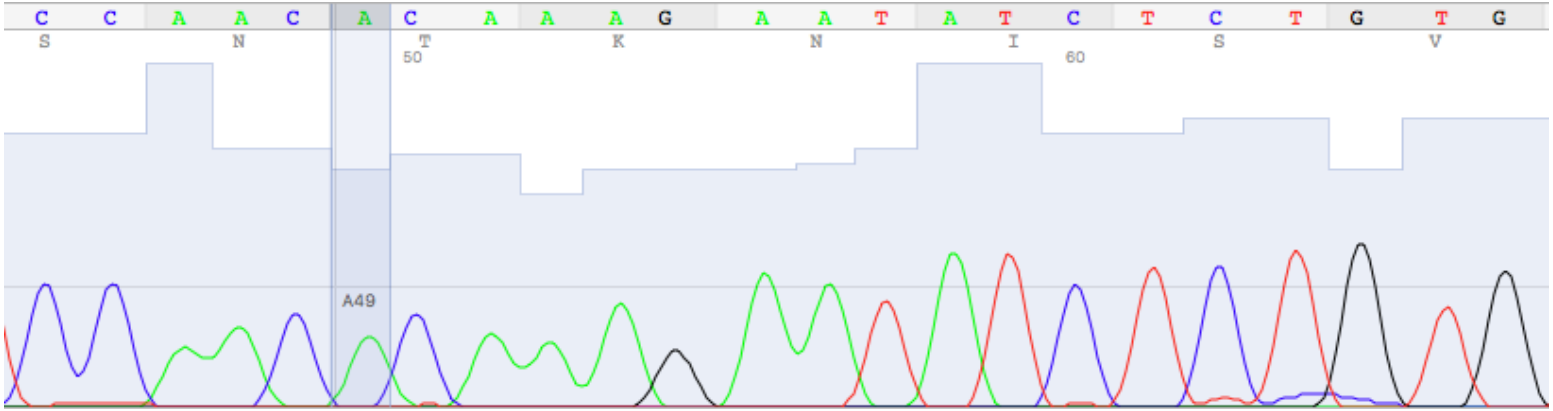
- Data silos
  - Variability and inconsistency
  - Sensitive / privacy concerns
- Structured
  - Electronic medical records
  - Sensor data
- Non-structured
  - Prescriptions
  - Discharge letters
  - Pathology reports
  - More…



Genomic Data Formats
From Raw Genome Data to Analysis

- **DNA Sequencing:** Transformation of analogues DNA into digital format
- **Alignment:** Reconstruction of complete genome with snippets
- **Variant Calling:** Identification of genetic variants
- **Data Annotation:** Linking genetic variants with research findings

Genomic Data Formats
Sequencing Signal



Data Sources and Formats

Data Management for Digital Health, Summer 2017
Chart 9

Genomic Data Formats
Sequencing ⇨ FASTQ files

Hasso Plattner Institut

@SRR831012.1 HWI-ST155_0742:7:1101:1284:1981/1

NGAGATGAAGCACTGTAGCTTGGAATTCTCGGGTGCCAAGGAACTCCAGT

+

%1=DDDFFHHHGFIHHIIIIIIIIIIIIIIIIIIEHIIIIIIIFIIIIIII

@SRR831012.2 HWI-ST155_0742:7:1101:2777:1998/1

NGAGATGAAGCACTGTAGCTCTTTGGAATTCTCGGGTGCCAAGGAACTCC

+

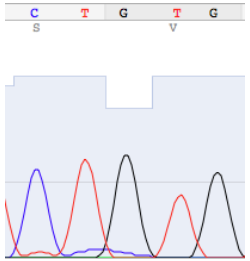
%1=DFFFFHHHHHIIIIIIIIIIIIIIIIIIIIGIIIIIIIIIIIIIIIIG

Quality score (increasing from worst to best):

!"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~

@SampleID.ReadNr

Experimental Setup



Data Sources and Formats

Data Management for Digital Health, Summer 2017

Chart **10**

Genomic Data Formats
Alignment ⇨ Reference genome (FASTA file)

- FASTA -> FAST-All (both proteins and nucleotides)

>Sequenz 1

;comment A

ACAAGATGCCATTGTCCCCCGGCCTCCTGCTGCTGCTGCTCTCCGGGGCCACGGCCACCGCTGCCCTG
CCCCTGGAGGGTGGCCCCACCGGCCGAGACAGCGAGCATATGCAGGAAGCGGCAGGAATAAGGAAA
AGCAGCCTCCTGACTTTCCTCGCTTGGTGGTTTGAGTGGACCTCCCAGGCCAGTGCCGGGCCCCTCA
TAGGAGAGGAAGCTCGGGAGGTGGCCAGGCGGCAGGAAGGCGCACCCCCCCAGCAATCCGCGCGCC
GGGACAGAATGCCCTGCAGGAACTTCTTCTGGAAGACCTTCTCCTCCTGCAAATAAAACCTCACCC
ATGAATGCTCACGCAAGTTTAATTACAGACCTGAA...

Newest version:

- Indexed hg38 (Human Genome version 38), 2013
- HG consists of approx. 3.2B base pairs

Hasso Plattner Institut

wellcome trust sanger institute

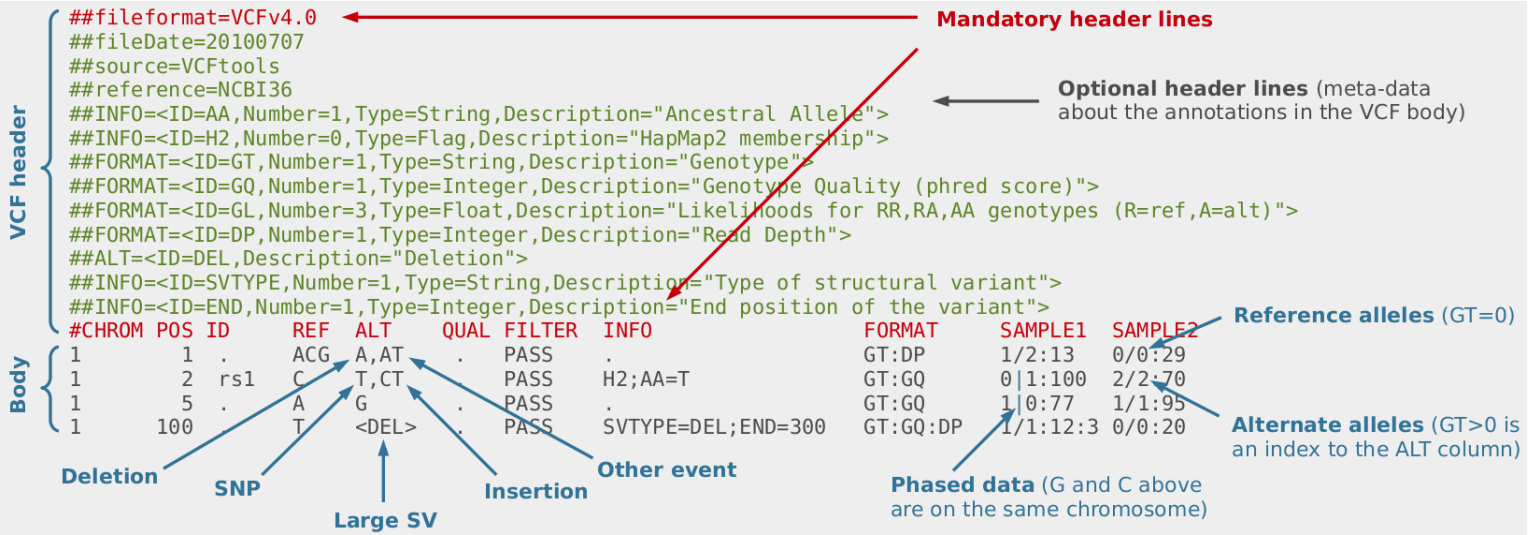
McDonnell Genome Institute

EMBL-EBI

NCBI

Genomic Data Formats
Variant Calling ⇨ Find variation (VCF file)

- Variant Call Format
- Standard format for SNP, Indels and structural variation
- Used by GATK (Gene Analysis Toolkit)



https://bioinf.comav.upv.es/courses/sequence_analysis/snp_calling.html

```
##fileformat=VCFv4.1
##fileDate=20140930
##source=23andme2vcf.pl https://github.com/arrogantrobot/23
##reference=file://23andme_v3_hg19_ref.txt.gz
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype"
#CHROM  POS  ID  REF ALT QUAL  FILTER  INFO  FORMAT  GEN
chr1  82154  rs4477212  a  .  .  .  .  GT  0/0
chr1  752566  rs3094315  g  A  .  .  .  GT  1/1
chr1  752721  rs3131972  A  G  .  .  .  GT  1/1
chr1  798959  rs11240777  g  .  .  .  .  GT  0/0
chr1  800007  rs6681049  T  C  .  .  .  GT  1/1
chr1  838555  rs4970383  c  .  .  .  .  GT  0/0
chr1  846808  rs4475691  C  .  .  .  .  GT  0/0
chr1  854250  rs7537756  A  .  .  .  .  GT  0/0
chr1  861808  rs13302982  A  G  .  .  .  GT  1/1
chr1  873558  rs1110052  G  T  .  .  .  GT  1/1
chr1  882033  rs2272756  G  A  .  .  .  GT  0/1
chr1  888659  rs3748597  T  C  .  .  .  GT  1/1
chr1  891945  rs13303106  A  G  .  .  .  GT  0/1
```

Header

Records

http://genomeintelligence.org/

Genomic Data Formats
Annotations

- Purpose:
  - Assess impact of genetic changes
  - Understand gene function and possible medical therapy options

- Input: List of genetic variants

CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE
chr7	140753336	**rs113488022**	T	A	61	PASS	NS=1	GT	0/1

- Output: Details about certain genetic locus

Genomic Data Formats
Annotations for Gene structure (GTF file)

- Gene transfer format (GTF)
- 20k-25k protein coding genes representing small part of the genome
- Using the annotation to speed up processing
- If the discovery of new genes in a sample is expected, a custom annotation can be calculated

Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 8	Col 9
chr21	HAVANA	transcript	10862622	10863067	.	+	.	gene_id "ENSG00000169..
chr21	HAVANA	exon	10862622	10862667	.	+	.	gene_id "ENSG00000169..
chr21	HAVANA	CDS	10862622	10862667	.	+	0	gene_id "ENSG00000169..
chr21	HAVANA	start_codon	10862622	10862624	.	+	0	gene_id "ENSG00000169..
chr21	HAVANA	exon	10862751	10863067	.	+	.	gene_id "ENSG00000169..
chr21	HAVANA	CDS	10862751	10863064	.	+	2	gene_id "ENSG00000169..
chr21	HAVANA	stop_codon	10863065	10863067	.	+	0	gene_id "ENSG00000169..
chr21	HAVANA	UTR	10863065	10863067	.	+	.	gene_id "ENSG00000169..

- Example name: geneshg19.gtf

Genomic Data Formats
State of the Art for Processing

Hasso Plattner Institut

bwa aln hg19.fa sample.fastq | bwa samse hg19.fa – sample.fastq | samtools view -Su - | samtools sort …

- Concatenation of command line tools reading/writing files
- Lack of standardization and exchangeability

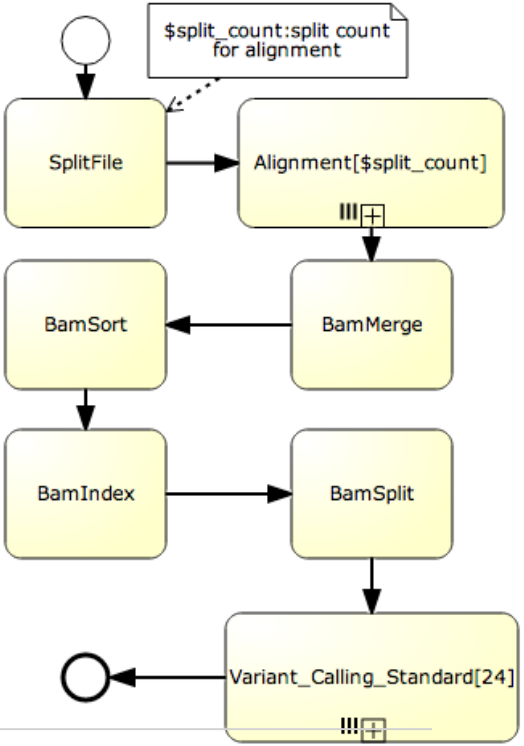
- Requires dedicated expertise for
  - Setup and configuration,
  - Error handling, and
  - Scalable processing

- Requirement: Enable modeling, configuration, and execution of reproducible pipelines without involving IT experts

Genomic Data Formats
Standardized Graphical Modeling

- Graphical modeling notation compliant with Business Process Modeling and Notation 2.0 extended by
  - Modular structure
  - Parallelization annotations
  - Parameters and variables

- Model descriptions are stored within database
- Model instances are transformed into graph structure for execution by dedicated runtime environment

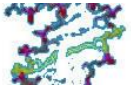


Genomic Data Formats
Challenges for Annotation Data

- Manual, time-consuming Internet search, e.g. publications, annotations, guidelines
- International consortiums provide fragmented information
- Missing linkage across individual data sources
- Annotations vary in completeness and correctness

Genomic Data Formats
Interpreting Annotations: BRAF Gene

dbSNP
Short Genetic Variations



Hasso Plattner Institut

- https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=**rs113488022**

Reference SNP (refSNP) Cluster Report: rs113488022 **** With Pathogenic allele ****

Clinical Significance: With Pathogenic allele

Primary Assembly Mapping

Assembly	SNP to Chr	Chr	Chr position	Contig	Contig position	Allele
GRCh38.p2	Rev	7	140753336	NT_007933.16	78246557	A

RefSeqGene Mapping

RefSeqGene	Gene (ID)	SNP to RefSeqGene	Position	Allele
NG_007873.3	BRAF (673)	Fwd	176429	T

Gene Model(s)

Function	mRNA				Protein		
	SNP to mRNA	Accession	Position	Allele change	Accession	Position	Residue change
missense	Fwd	NM_004333.4	1860	GTG ⇒ GAG	NP_004324.2	600	V [Val] ⇒ E [Glu]
missense	Fwd	XM_005250045.1	1807	GTG ⇒ GAG	XP_005250102.1	600	V [Val] ⇒ E [Glu]
missense	Fwd	XM_005250046.1	1807	GTG ⇒ GAG	XP_005250103.1	600	V [Val] ⇒ E [Glu]
missense	Fwd	XM_011516529.1	1807	GTG ⇒ GAG	XP_011514831.1	600	V [Val] ⇒ E [Glu]
noncoding	Fwd	XR_242100.1	1807	NA ⇒ NA	NA	NA	NA

	Sample Ascertainment				Genotypes	Alleles	
ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	HWP	A	T
ss1688979043	ExAc_Aggregated_Populations		121410	AF		0.00001647	0.99998355

Genomic Data Formats
Interpreting Annotations: BRAF Gene

GeneCards HUMAN GENE DATABASE

Hasso Plattner Institut

Aliases for BRAF Gene

Aliases for BRAF Gene

elonging to the raf/mil family of serine/threonine protein kinases. This protein plays a role in regulating the MAP kinase/ERKs signaling pathway, which affects d secretion. Mutations in this gene are associated with cardiofaciocutaneous syndrome, a disease characterized by heart defects, mental retardation and a Mutations in this gene have also been associated with various cancers, including non-Hodgkin lymphoma, colorectal cancer, malignant melanoma, thyroid carcinoma, and adenocarcinoma of lung. A pseudogene, which is located on chromosome X, has been identified for this gene. [provided by RefSeq, Jul 2008]

, Serine/Threonine Kinase) is a Protein Coding gene. Diseases associated with BRAF include Cardiofaciocutaneous Syndrome and Lung Cancer. Among its ment Slit-Robo signaling and MAPK Signaling: Oxidative Stress Pathway. GO annotations related to this gene include *calcium ion binding* and *transferase* ıs-containing groups. An important paralog of this gene is RAF1.

External Ids for BRAF Gene

HGNC: 1097 Entrez Gene: 673 Ensembl: ENSG00000157764 OMIM: 164757 UniProtKB: P15056

Previous GeneCards Identifiers for BRAF Gene

GC07M138705, GC07M139764, GC07M139841, GC07M139754, GC07M139887, GC07M140080, GC07M140424, GC07M134729

Export aliases for BRAF gene to outside databases

Summaries for BRAF Gene

Genomic Data Formats
Interpreting Annotations: BRAF Gene

Neuropathology. 2016 Oct 28. doi: 10.1111/neup.12347. [Epub ahead of print]

A comprehensive analysis identifies BRAF hotspot mutations associated with gliomas with peculiar epithelial morphology.

Hatae R[1], Hata N[2], Suzuki SO[3], Yoshimoto K[2], Kuga D[2], Murata H[2], Akagi Y[2], Sangatsuda Y[2], Iwaki T[3], Mizoguchi M[4], Iihara K[2].

Author information

Abstract

Brain tumors harbor various BRAF alterations, the vast majority of which are the BRAF kinase-activating V600E mutation. BRAF mutations are most frequently detected in certain subtypes of low-grade glioma, such as pilocytic astrocytoma (PA), pleomorphic xanthoastrocytoma (PXA), ganglioglioma (GG) and dysembryoplastic neuroepithelial tumor (DNT). However, it is unclear whether gliomas harboring BRAF mutations can be invariably regarded as these glioma subtypes or their derivatives. To address this question, we analyzed 274 gliomas in our institutional case series.

Genomic Data Formats
Olelo: Intuitive exploration of biomedical literature

Olelo

Hasso Plattner Institut

A

What are the symptoms of a zika virus infection?

List of Symptoms | Dysgeusia

HEADACHE
B
DYSGEUSIA
FEVER
PURPURA
HYDROPS FETALIS
NEUROBEHAVIORAL MANIFESTATIONS
MYALGIA
VOMITING
SKIN MANIFESTATIONS
PAIN
NAUSEA

Dysgeusia C

A condition characterized by alterations of the sense of taste which may range from mild to severe, including gross distortions of taste quality.

DISEASES 4
GEOGRAPHICALS 3
SPECIES 1
SYMPTOMS 1

SHOW 1 DOCUMENT MATCHING YOUR SEARCH

D

Abstracts of 1/1 document

Your search: Dysgeusia

The Clinical Spectrum of Zika Virus in Returning Travelers. F

G Gottesman, E Leshem, Y Lustig, E Meltzer, E Schwartz
Am. J. Med., 2016, 129(10)
PMID: 27260832

The clinical spectrum of Zika virus had, to date, been described in small series from endemic/epidemic countries and is not well established.We describe the clinical manifestations of laboratory-proven Zika virus infection in Israeli travelers during December 2015-February 2016, and review all published cases of travel-related Zika virus.During the study period, 8 returning Israeli travelers were diagnosed with Zika virus infection. In addition, 41 published cases were included, mostly from Latin America to Europe and North America. Overall, 65.3% were diagnosed by polymerase chain reaction. rash was the most frequent symptom, present in 95.7% of cases, followed by fever and arthralgia. Conj...

CREATE SUMMARY | EXPORT AS BIOC

Summary of 1 documents E

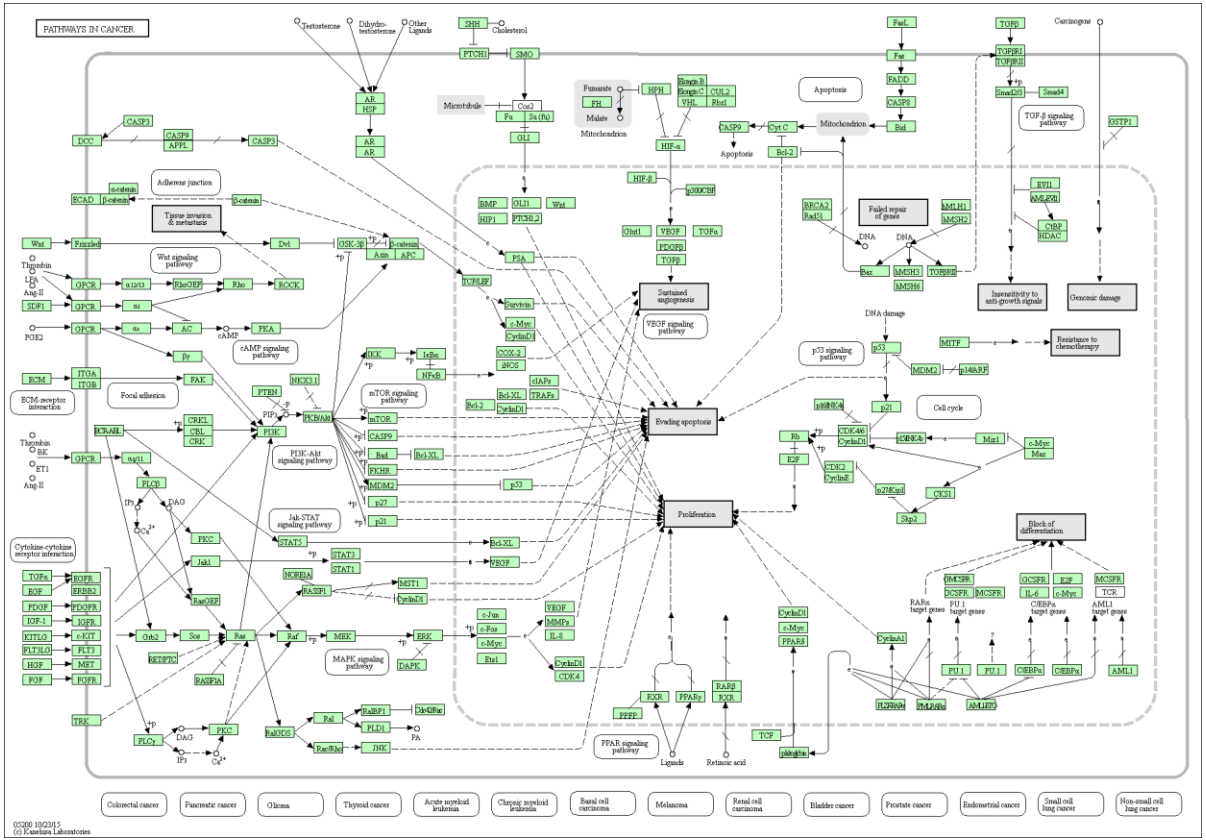
rash was the most frequent symptom, present in 95.7% of cases, followed by fever and arthralgia. conjunctivitis was present in 53.1%; however, only 40.3% presented with a triad of conjunctivitis, fever, and rash. Less frequent symptoms included dysgeusia and nightmares, which, together with arthralgia, persisted for several weeks in some travelers.Zika virus clinical picture in travelers is diverse.

TRANSLATE | CORRESPONDING DOCUMENTS

(A) Question asked
(B) Selected term 'Dysgeusia'
(C) MeSH definition
(D) Collection card
(E) Summary
(F) Link to primary source

Kraus et al. Olelo: a web application for intuitive exploration of biomedical literature. Nucleic Acids Research (2017)

Genomic Data Formats
Graph Data: Pathways (Kegg example)



Data Sources and Formats

Data Management for Digital Health, Summer 2017

Medical Data Formats

Source: Armed Forces Institute of Cardiology & National Institute of Heart Diseases (Pakistan)

Medical Data Formats
Electronic Medical Records (EMR) in Germany

- Only 60% of German clinics use EMR[1]
- Germany occupies 10th place among 20 EU countries[2]
- Leaders: Denmark, Sweden and Estonia
- In Denmark: 99% of discharge letters are digital[3]
- Estonia has a national EMR since 2008[4]

Stiftung Münch, Studie zur elektronischen Patientenakte im Ausland (2015)

[1] http://www.pwc.de/de/pressemitteilungen/2014/papierlose-klinik-bleibt-vision-deutschland-liegt-bei-ehealth-zurueck.html
[2] http://www.stiftung-muench.org/studie-zur-elektronischen-patientenakte-im-ausland-klare-vorgaben-des-gesetzgebers-sind-voraussetzung-fuer-erfolgreiche-implementierung/
[3] http://www.sum.dk/~/media/Filer%20-%20Publikationer_i_pdf/2012/Sundheds-IT/Sundheds_IT_juni_web.ashx
[4] https://e-estonia.com/component/electronic-health-record/

Medical Data Formats
EMR Adoption Model (EMRAM)

HIMSS, EMR Adoption Model, Cumulative Capabilities

7	Complete EMR, Data Analytics to improve care
6	Physician Documentation (templates), Full CDSS, Closed Loop Medication Administration
5	Full R-PACS
4	CPOE; Clinical Decision Support (clinical protocols)
3	Clinical Documentation, CDSS (error checking)
2	CDR, Controlled Medical Vocabulary, CDS, HIE Capable
1	All Three Ancillaries Installed — Lab, Rad, Pharmacy
0	All Three Ancillaries Not Installed

http://www.himss.eu/healthcare-providers/emram

Medical Data Formats
EMR Adoption Model (EMRAM)

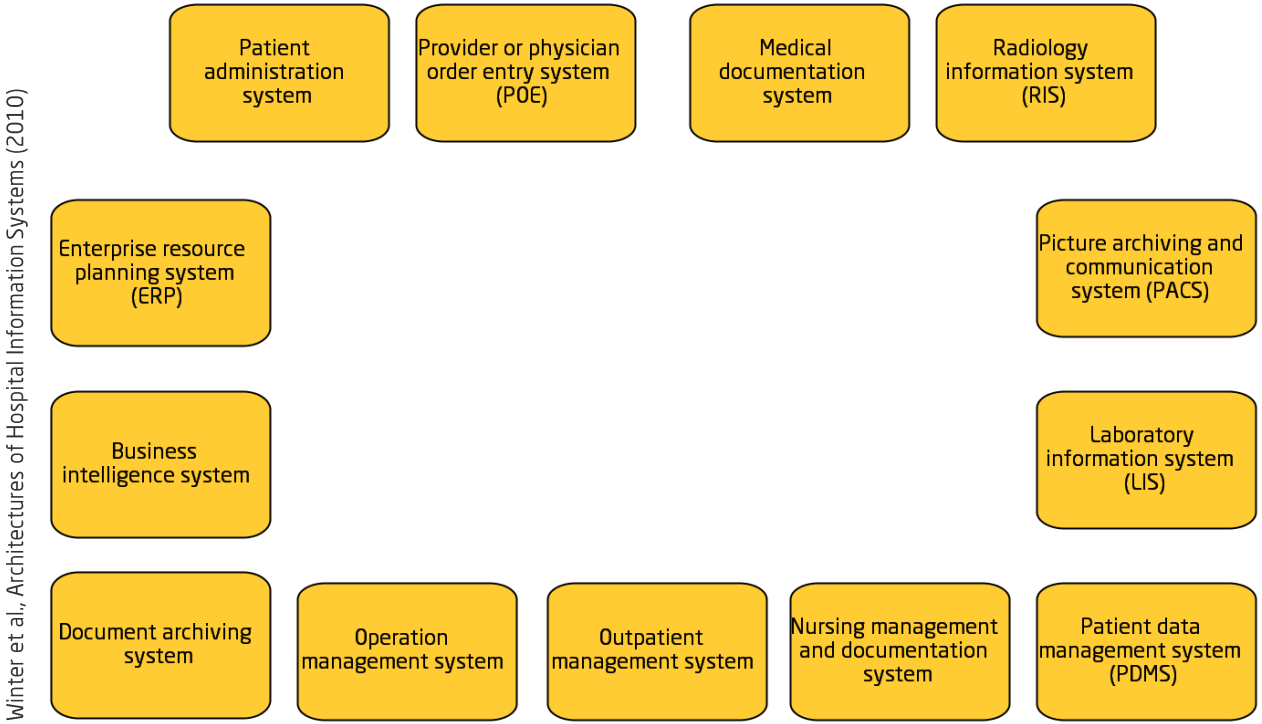
Belgium	2
France	4
Germany	1
Ireland	1
Italy	4
Netherlands	8
Portugal	1
Spain	15
Switzerland	2
Turkey	16
United Kingdom	3
TOTAL:	57

(data from 1/1/2017)

EMRAM stage 6 / 7 awardees

http://www.himss.eu/healthcare-providers/emram

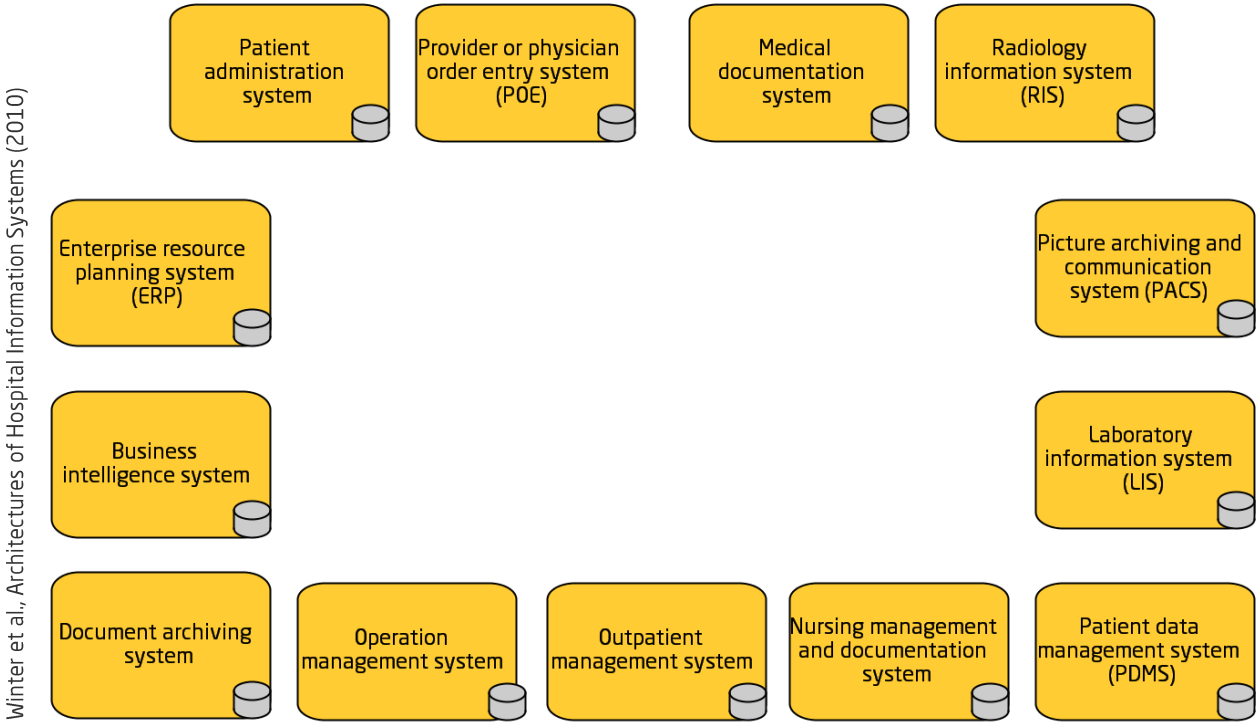
Medical Data Formats
Hospital Information Systems (Reference Architecture)



Winter et al., Architectures of Hospital Information Systems (2010)

A. Winter, R. Haux, E. Ammenwerth, B. Brigl, N. Hellrung, F. Jahn. Health Information Systems: Architectures and Strategies (2010)

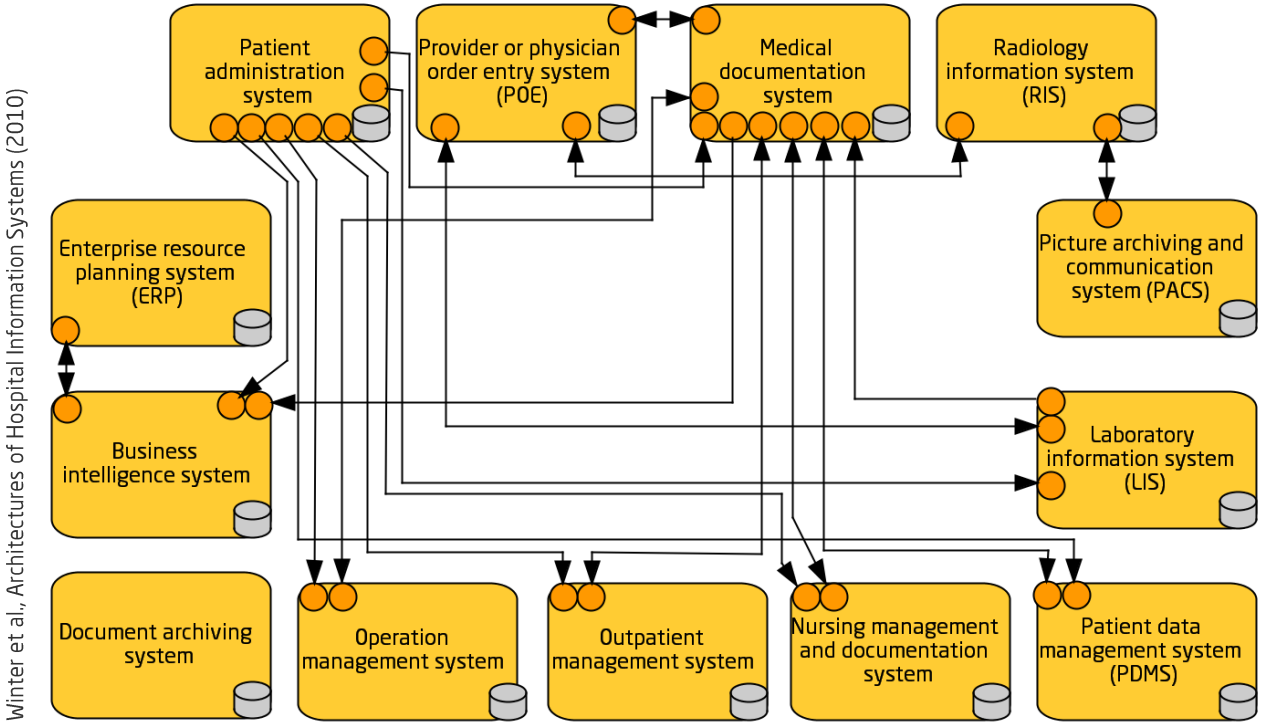
Medical Data Formats
Hospital Information Systems (Reference Architecture)



Winter et al., Architectures of Hospital Information Systems (2010)

A. Winter, R. Haux, E. Ammenwerth, B. Brigl, N. Hellrung, F. Jahn. Health Information Systems: Architectures and Strategies (2010)

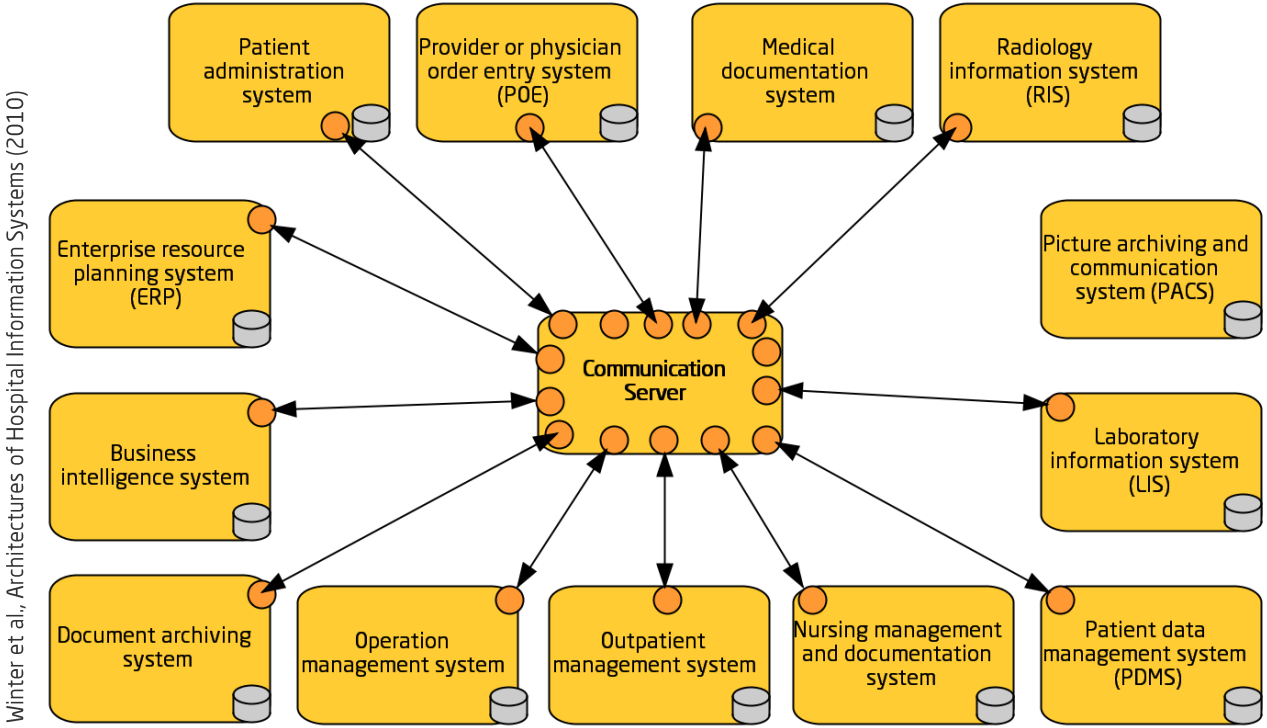
Medical Data Formats
Hospital Information Systems (Reference Architecture)



Winter et al., Architectures of Hospital Information Systems (2010)

A. Winter, R. Haux, E. Ammenwerth, B. Brigl, N. Hellrung, F. Jahn. Health Information Systems: Architectures and Strategies (2010)

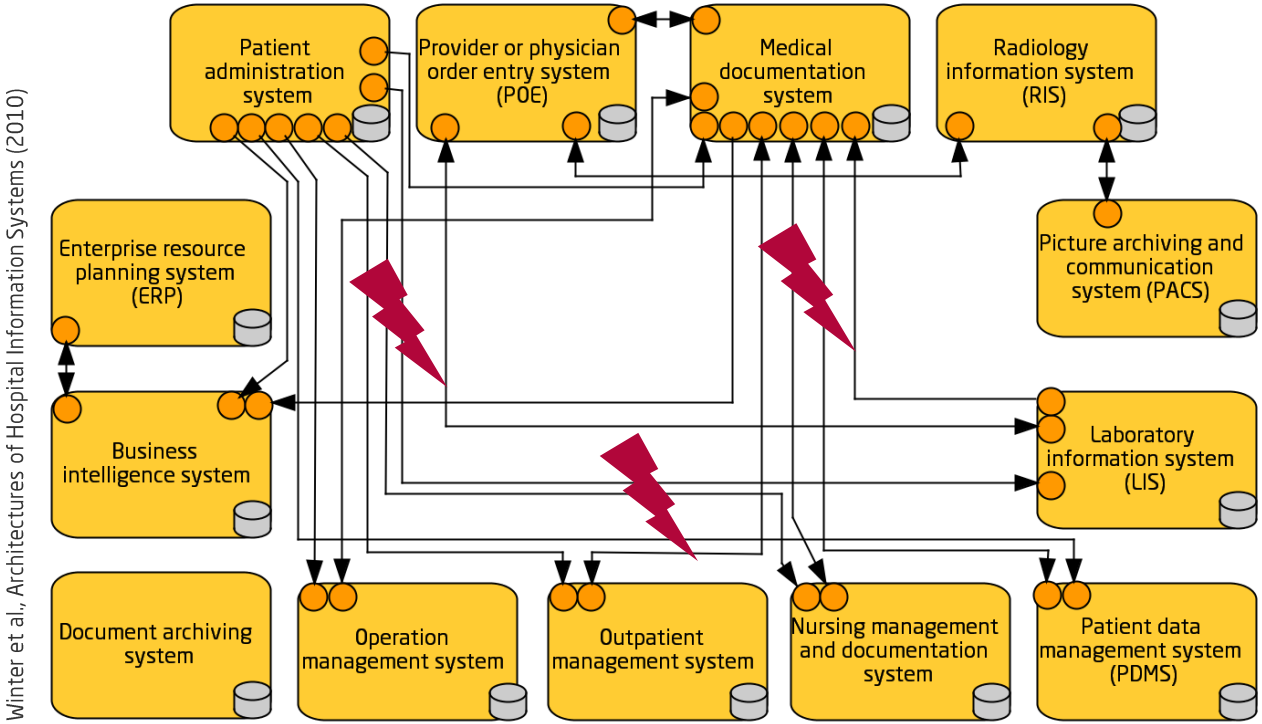
Medical Data Formats
Hospital Information Systems (Reference Architecture)



Winter et al., Architectures of Hospital Information Systems (2010)

A. Winter, R. Haux, E. Ammenwerth, B. Brigl, N. Hellrung, F. Jahn. Health Information Systems: Architectures and Strategies (2010)

Medical Data Formats
Hospital Information Systems (Reference Architecture)



Winter et al., Architectures of Hospital Information Systems (2010)

A. Winter, R. Haux, E. Ammenwerth, B. Brigl, N. Hellrung, F. Jahn. Health Information Systems: Architectures and Strategies (2010)

Medical Data Formats
Standards for Data Exchange

- HL7
- FIHR
- DICOM
- ISO/IEEE 11073
- CCOW
- IHE
- EDIFACT
- CDA
- ...

Medical Data Formats
Standards for Data Exchange: HL7

- Health Level 7
- Dedicated message types between application components
- Event-driven communication
- Example messages:
  - ADT (Admission Discharge Transfer)
  - ORM (Order entry message)

Adapted from Winter et al. (2010)

Medical Data Formats
Standards for Data Exchange: HL7

- ADT (Admission Discharge Transfer)
- 51 different types of ADT messages

```
MSH|^~\&|AcmeHIS|StJohn|ADT|StJohn|20050518073622||ADT^A01|MSGID
20050518073622|P|2.3
EVN|A01
```

PID|||12001||Jones^John^^^Mr.||19670822|M|||123 West St.^^Denver^CO^80020^USA||(850)555-0809|||||99345|460-99-2928	PID – Patient Info
PV1||I|Main^802^1||||^Quacker^John|||IP|||||||||1||||||||||||||||||||||||||20050518073622	PV1 – Visit Info
IN1|1|EPO|80|AETNA US HEALTHCARE|PO BOX 981114^"""^EL PASO^TX^79998^""|||1500004000001|AETNA SERVICES INC|19|AETNA US HEALTHCARE|""|""||2|SOUTAR^RENEE^D|3|19700722|13324 WHITE CEMETERY RD^"""^HANNIBAL^NY^130740000^""||||||||||||||||||124705454||||||1|F|225 GREENFIELD PARKWAY^^LIVERPOOL^NY^13088|185428 IN2|1||124705454||461-1200||||||	IN1 & IN2 Insurance Info

https://corepointhealth.com/resource-center/hl7-resources/hl7-adt

Medical Data Formats
Standards for Data Exchange: FHIR

- Fast Healthcare Interoperability Resources
- Builds on HL7 version 2.x and HL7 version 3.x
- Supports both XML and JSON (RESTful API)

Adapted from: http://wiki.galenhealthcare.com/HL7_FHIR

```
<Patient xmlns="http://hl7.org/fhir">
  <extension>
    <url value="http://www.goodhealth.org/consent/trials"/>
    <valueCode value="renal"/>
  </extension>
  <text>
    <status value="generated"/>
    <div xmlns="http://www.w3.org/1999/xhtml">
      <p>Henry LEVIN the 7th, DOB 24-Sept 1932</p>
      <p>MRN: 123456</p>
    </div>
  </text>
</Patient>
```

https://www.hl7.org/fhir/

Medical Data Formats
Standards for Data Exchange: FHIR

- Open source implementations available (e.g. Java HAPI FHIR)
- msgString := XML FIHR resource

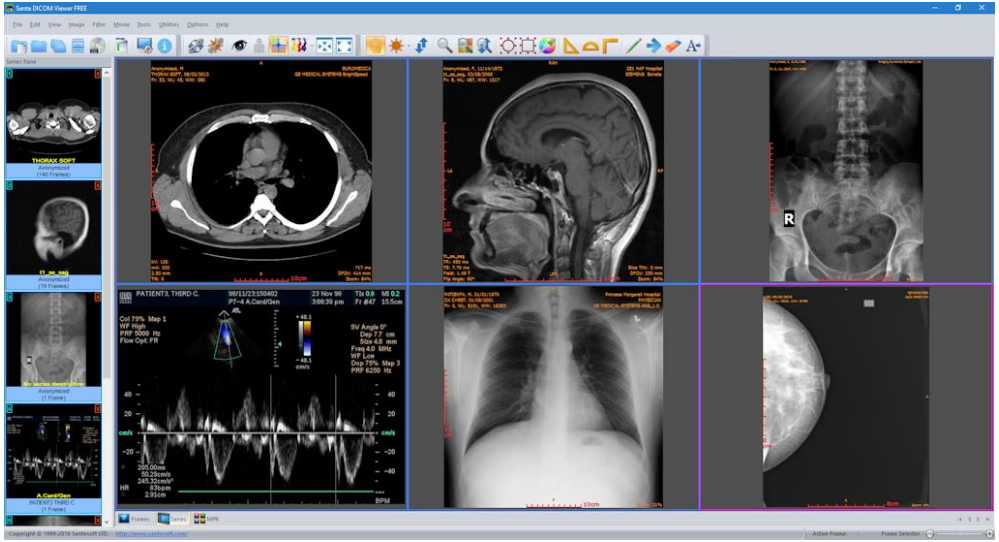
```
// The hapi context object is used to create a new XML parser
// instance. The parser can then be used to parse (or unmarshall) the
// string message into a Patient object
IParser parser = ctx.newXmlParser();
Patient patient = parser.parseResource(Patient.class, msgString);

// The patient object has accessor methods to retrieve all of the
// data which has been parsed into the instance.
String patientId = patient.getIdentifier().get(0).getValue();
String familyName = patient.getName().get(0).getFamily().get(0).getValue();
String gender = patient.getGender();

System.out.println(patientId); // PRP1660
System.out.println(familyName); // Cardinal
System.out.println(gender); // M
```

HAPI FHIR: Parsing a resource from a String

http://hapifhir.io/doc_intro.html

Medical Data Formats
Standards for Data Exchange: DICOM

- Digital Imaging and Communications in Medicine
- Picture Archiving and Communication System (PACS)



http://www.santesoft.com/win/sante-dicom-viewer-free/img/modalities.png

Radiology information system (RIS)

Medical Data Formats
Standards for Data Exchange: IHE

- Specifications, tools and services for interoperability
- Sharing of health information across systems
- Coordinated use of established standards (DICOM, HL7)
- Addresses specific clinical needs (**domains**)
- Targets care provision workflows (**profiles**)

- Example Domain: Patient Care Device (PCD)
  - Alert Communication Management
  - Subscribe to Patient Data

Source: http://www.ihe.net/

IHE Connectathon

HE North American Connectathon 2017 Test Floor, Huntington Convention Center of Cleveland

Medical Data Formats
Data Formats for Clinical Studies: CDISC ODM

- Operational Data Model (ODM)-XML
- Enables exchange of clinical study data
- Models a Case Report Form (CRF)

ADVERSE EVENT		
Did any adverse event occur after the last visit?—*Indicator Question*		Yes ☐ 1 No ☐ 2
If " Yes", please provide the information in the adverse events (AE) page (page no. XX) and give details below:–Skip		
AE page number	AE serial number	Did any unscheduled visit happen after the last visit? *(Please provide details on page no. YY)–Skip*
☐☐	☐☐	Yes ☐ 1 No ☐ 2
☐☐	☐☐	Yes ☐ 1 No ☐ 2
☐☐	☐☐	Yes ☐ 1 No ☐ 2

LIMS Wiki: Basics of case report form designing in clinical research

Clinical Data Interchange Standards Consortium

Data Sources and Formats

Data Management for Digital Health, Summer 2017

https://www.cdisc.org/

Medical Data Formats
Data Formats for Clinical Studies: CDISC ODM

- Operational Data Model (ODM)-XML
- Enables exchange of clinical study data
- Models a Case Report Form (CRF)

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE ODM (View Source for full doctype...)>
- <ODM TransferType="Snapshot" Granularity="Study" TransferID="987-654-321"
  PriorTransferID="987-654-320" CreationDateTime="2001-10-16T13:27:45">
  - <Study ID="123-456-789">
    - <GlobalVariables>
        <StudyName>CDISC Test Study II</StudyName>
        <StudyDescription>This file contains test data for the CDISC
          Connect-A-Thon event scheduled for the DIA 11th Annual
          European Workshop on Clinical Data
          Management.</StudyDescription>
        <ProtocolName>CDISC-123-456-789</ProtocolName>
      </GlobalVariables>
    + <BasicDefinitions>
    + <MetaDataVersion ID="v1.01" Name="Version 1.01">
    </Study>
  - <AdminData StudyID="123-456-789">
    - <User ID="USR.cdisc001">
        <FullName>Shirley Williams</FullName>
        <FirstName>Shirley</FirstName>
        <LastName>Williams</LastName>
```

http://www.xml4pharma.com/CDISC/images/CDISC_ODM11_draft_example.gif

Clinical Data Interchange Standards Consortium

Data Sources and Formats

Data Management for Digital Health, Summer 2017

https://www.cdisc.org/

What to Take Home?

- Health care data speaks many languages
- Either structured or non-structured
- Formats in an -omics context
- Formats in the hospital setting

Save the Date
HIMSS Impact: Big Data in Medicine

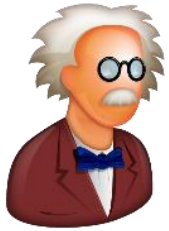
- When / where?
  - 20-21 November 2017
  - Potsdam, Germany, HPI ☺
- Topics
  - Impact of Big Data
  - How will health care delivery change?
  - State of the art and challenges

HIMSS Europe

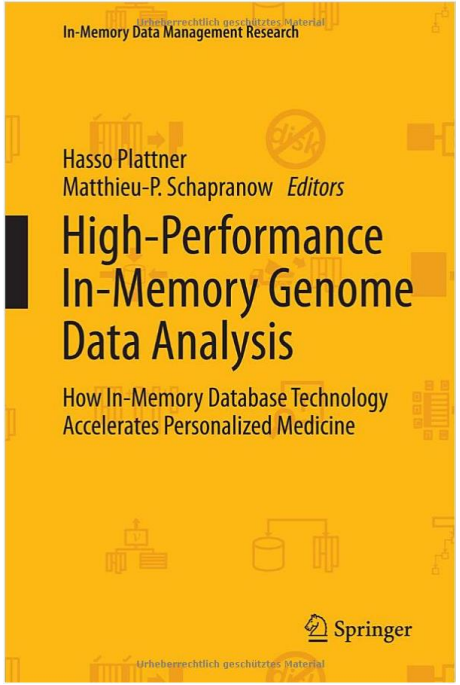
http://www.himss.eu/events/himss-impact-2017

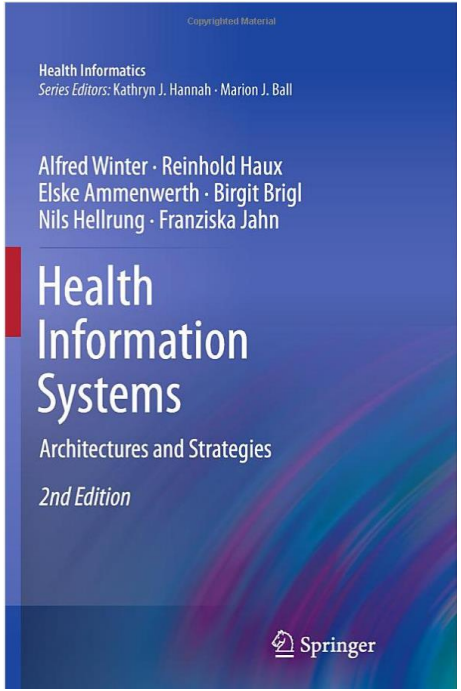
Save the Date (cont'd.)
Upcoming Expert Talks

- Dates confirmed
  - Christof von Kalle, Oncology, 01.06.2017 (National Center for Tumor Diseases)
  - Klemens Budde, Nephrology, 13.06.2017 (Charité)
  - Peter Kohlmann, 27.06.2017 (German Heart Center)



To Know More


Urheberrechtlich geschütztes Material
In-Memory Data Management Research
Hasso Plattner
Matthieu-P. Schapranow Editors
High-Performance In-Memory Genome Data Analysis
How In-Memory Database Technology Accelerates Personalized Medicine
Springer
Urheberrechtlich geschütztes Material


Copyrighted Material
Health Informatics
Series Editors: Kathryn J. Hannah · Marion J. Ball
Alfred Winter · Reinhold Haux
Elske Ammenwerth · Birgit Brigl
Nils Hellrung · Franziska Jahn
Health Information Systems
Architectures and Strategies
2nd Edition
Springer

What's Next?

- Use case Nephrology
- Health care analytics
- Clinical prediction models