Genetica Biotecnologia Bioinformatica Applicata

GeBBAlab

GeBBA Lab

Genomic and Bioinformatic Applied to Biotech

Sergio D’Ascia, NSI – Bologna – Italy - s.dascia@nsi-mail.it
Giuseppe Frangiamone, NSI – Bologna – Italy – g.frangiamone@nsi-mail.it

NSI - Nier Soluzioni Informatiche S.r.l.
Via Clodoveo Bonazzi 2
40013 CASTEL MAGGIORE (BO)
web site: http://www.nsi-online.it

> n s i

CINECA Consorzio Interuniversitario

IOR Istituti Ortopedici Rizzoli

DAMA

Regione Emilia-Romagna

Overview

PART 1 - GebbaLab project

PART 2 - HL7: the standard for clinical and genomics data exchange

PART 3 - Integration of HL7 into the GebbaLab project: the glue of the architecture

PART 4 – Conclusion

Part 1 – GebbaLab project

GebbaLab project

Key areas and goals of the project

Two key areas:

1. **Genomic area:**
   - Microarray;
   - SNPs;
   - Sequencing;
   - Proteomics.
2. **Clinical area:**
   - clinical data from health care providers (such as labs, hospitals, clinics and health care providers in general) operating in different clinical domain

Two main goals:

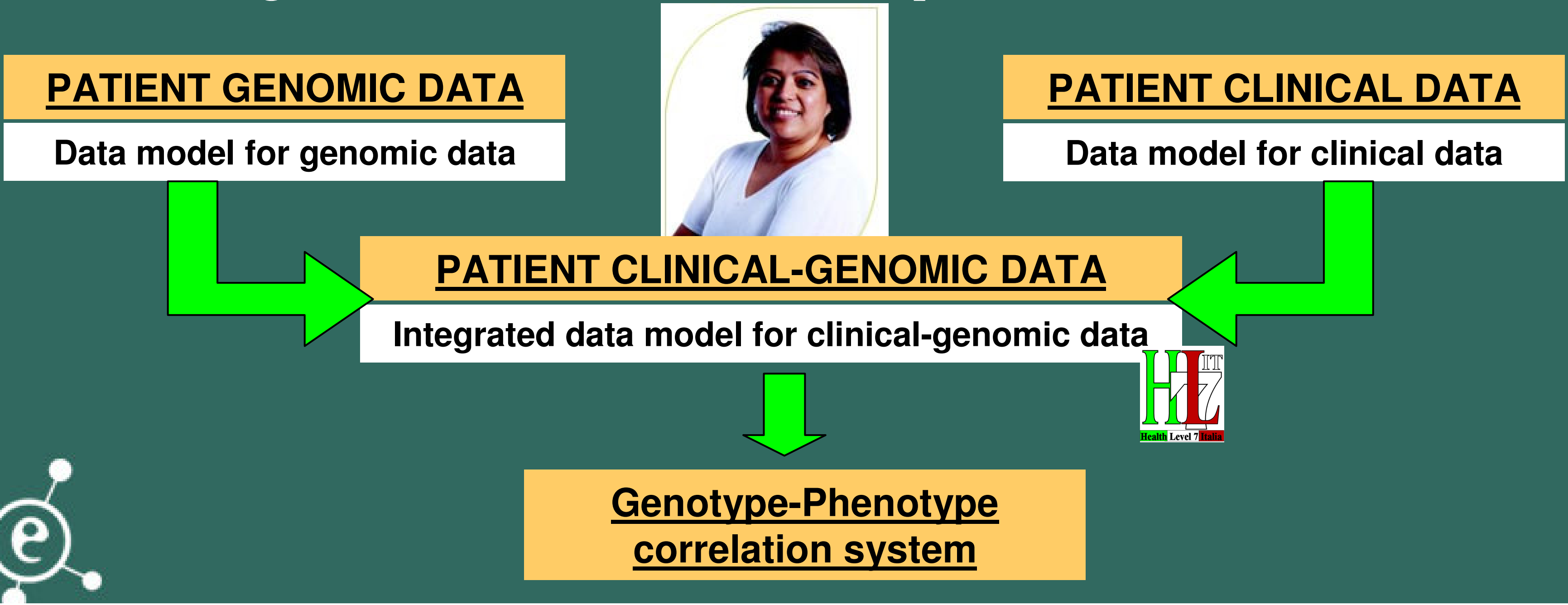
1. To provide **distributed services** to
   - Store
   - Retrieve
   - Manage
   - Analyze

   both **clinical** and **genomic data** in a **<u>patient centric approach</u>** (PCA);
2. To allow the **exchange** of both
   - Clinical and
   - genomic data

   among heterogeneous systems using standard exchange data representation such as HL7 v3

Need to integrate clinical and genomic data in a standard way to provide interoperability

1. Store genomic data;
2. Store clinical data;
3. Create a data model to integrate clinical and genomic data in a standard way to allow heterogeneous application interoperability;
4. Correlate genomic data to clinical data in a patient centric view



PATIENT GENOMIC DATA

Data model for genomic data

PATIENT CLINICAL DATA

Data model for clinical data

PATIENT CLINICAL-GENOMIC DATA

Integrated data model for clinical-genomic data

Genotype-Phenotype correlation system

Institutions working at the project and external collaborations

Participating institutions: technical and biological partners

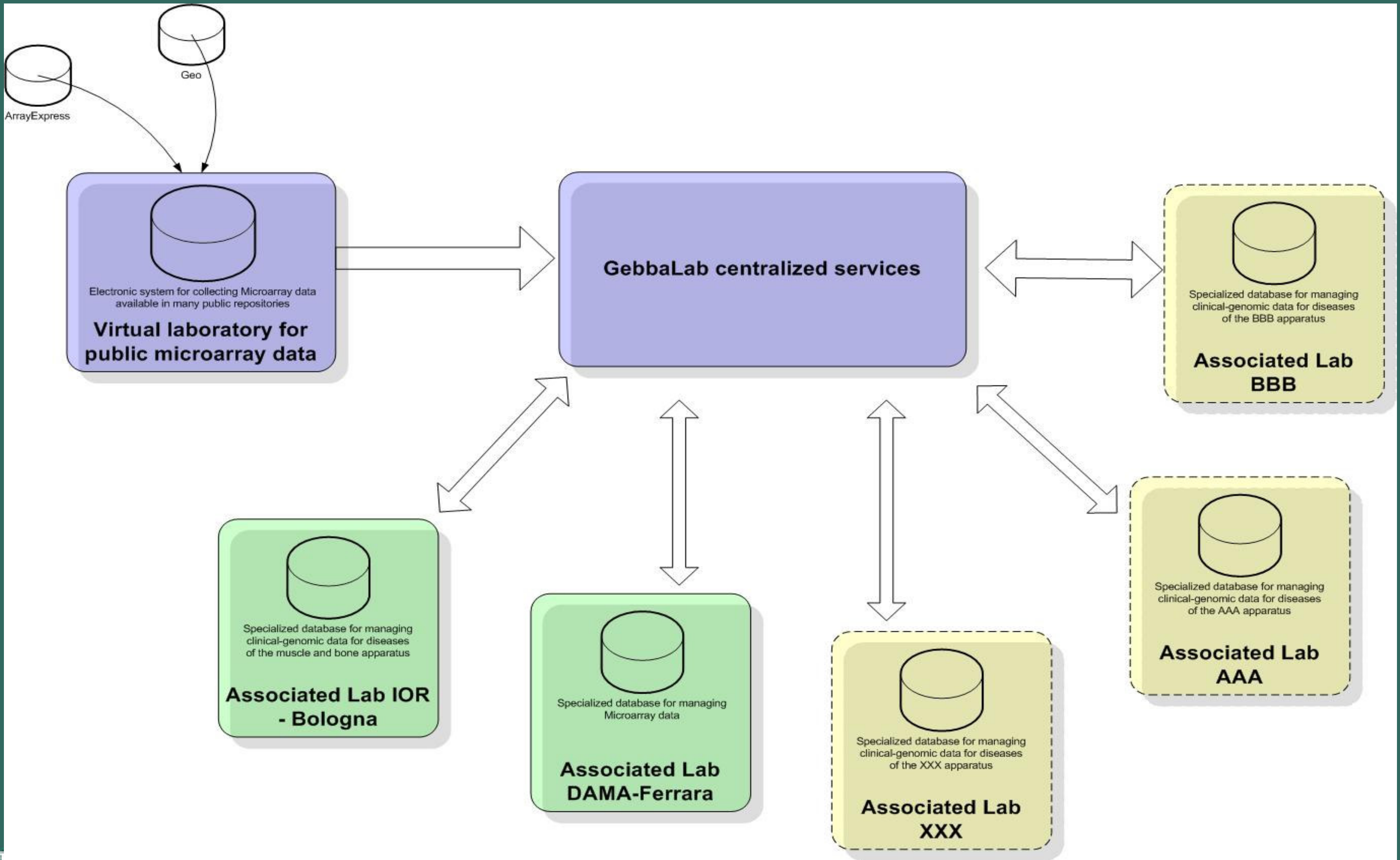
1. Genetics Unit, Rizzoli Orthopedic Institute (IOR) - Bologna;
2. Hbio-Lab, CINECA – Bologna;
3. Telethon DAMA - Ferrara University;
4. NSI – Nier ITC Solutions srl - Bologna

External collaborations

1. HL7 Italy – Gregorio Mercurio and Stefano Dalmiani (Clinical-Genomics domain), Paolo Marcheschi (Clinical Document Architecture v2)
2. HL7 Haifa – Amnon Shabo facilitator of the Clinical-Genomics Special Interest Group of HL7 International;
3. Sant’Orsola - University Hospital (Bologna)

The GebbaLab functional architecture

The functional architecture



Part 2 – HL7

HL7: the standard of clinical and genomic data exchange

The HL7 standard: an overview

Health Level Seven (HL7): overview

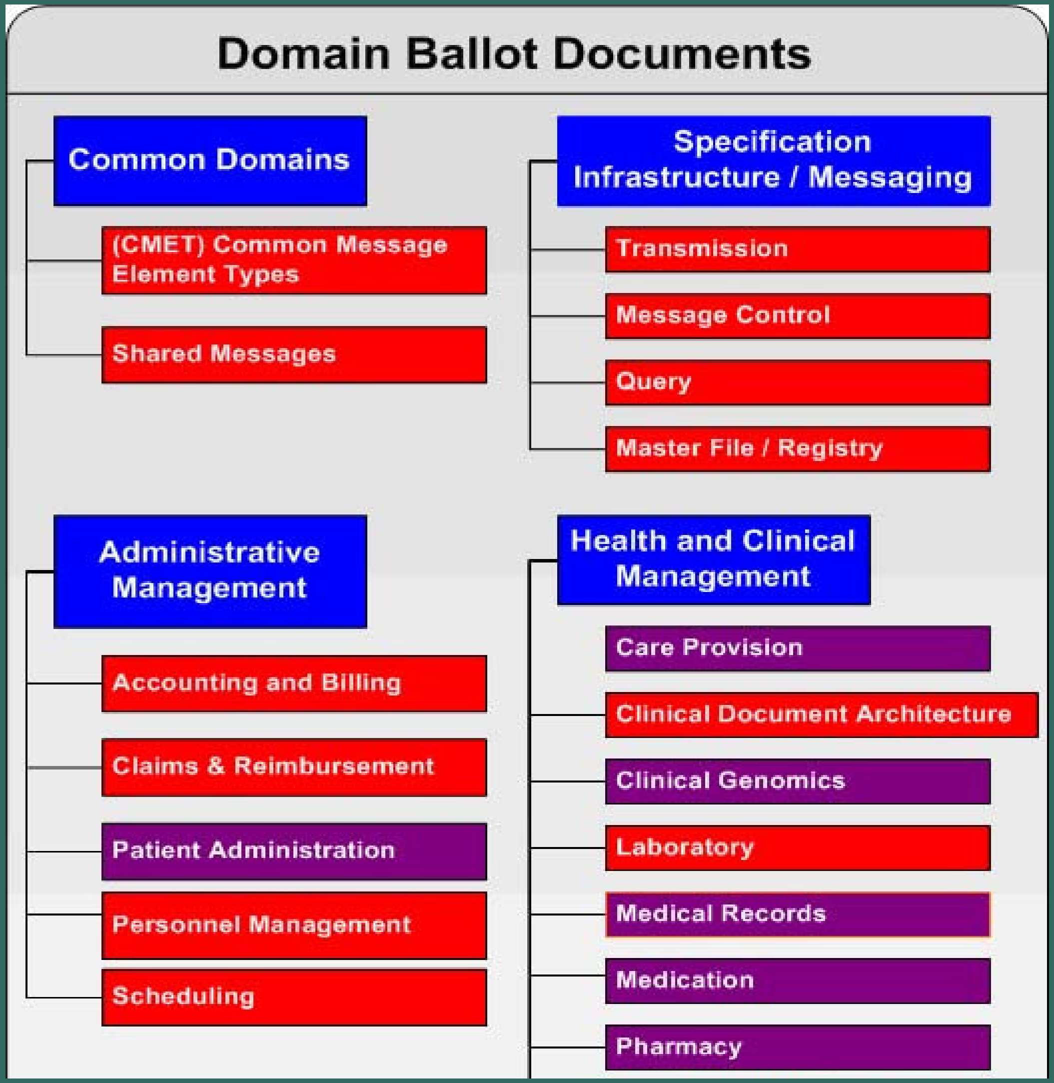
- Currently the most advanced standard for integrating, representing and exchanging both clinical (**phenotype**) and genomic (**genotype**) data among heterogeneous health care providers' systems;
- an American National Standards Institute (ANSI) accredited Standards Developing Organization (SDOs) founded in 1987;
- Built around different clinical domains (CDA, ClinicalGenomics, PatientAdministration, Account and Billings, etc.);
- Important version evolution over the time: version 2.x to version 3;
- Encapsulating and bubble-up paradigm for clinical-genomics domain data;
- Supporting international standards such as MAGE-ML and BSML for genomic data.

Clinical domains in HL7

HL7 clinical domain classification:

The following are HL7 Clinical domains:

1. Common;
2. Specific infrastructure-messaging;
3. Administrative;
4. Health and clinical management:
   - CDA;
   - ClinicalGenomics;
   - ………



Domain Ballot Documents

- Common Domains
  - (CMET) Common Message Element Types
  - Shared Messages
- Specification Infrastructure / Messaging
  - Transmission
  - Message Control
  - Query
  - Master File / Registry
- Administrative Management
  - Accounting and Billing
  - Claims & Reimbursement
  - Patient Administration
  - Personnel Management
  - Scheduling
- Health and Clinical Management
  - Care Provision
  - Clinical Document Architecture
  - Clinical Genomics
  - Laboratory
  - Medical Records
  - Medication
  - Pharmacy

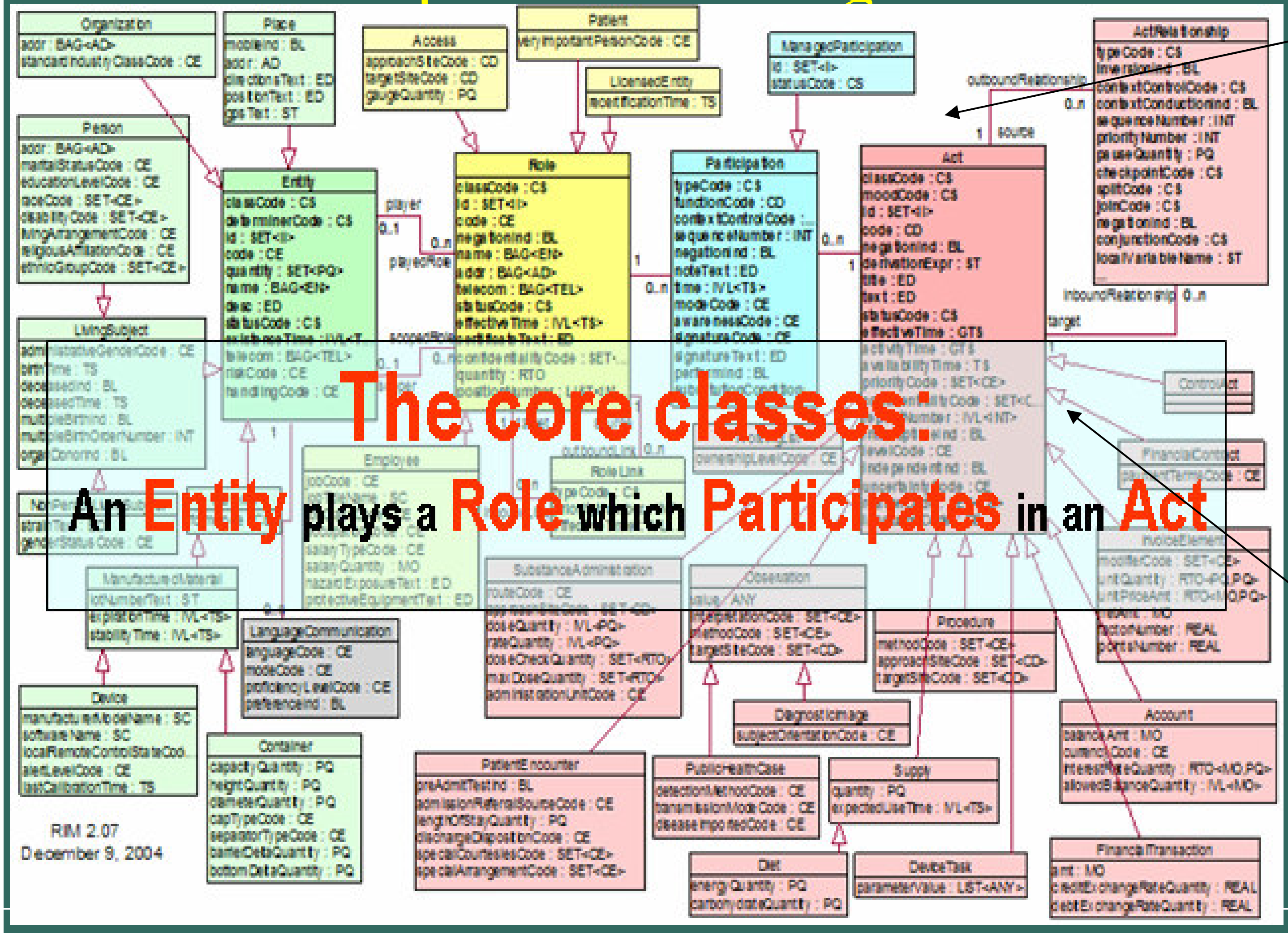
What is an HL7 message?
How is it built?

HL7 message in version 3

- HL7 evolution: from version 2.x (messages as ASCII flat files composed by segments of different types in specific position inside the document) to version 3 (messages as XML files validated against specific XML schema built for specific domains of interest);
- Development of the Reference Information Model (RIM), an object-oriented data model, which enable to build messages by including specific entities (classes in an object model) such as Patient, PatientEncounter, Observation, Person, GeneticLocus, Sequence, etc.

The backbone classes and RIM

HL7 RIM: the process of deriving an HL7 v3 message



The core classes:

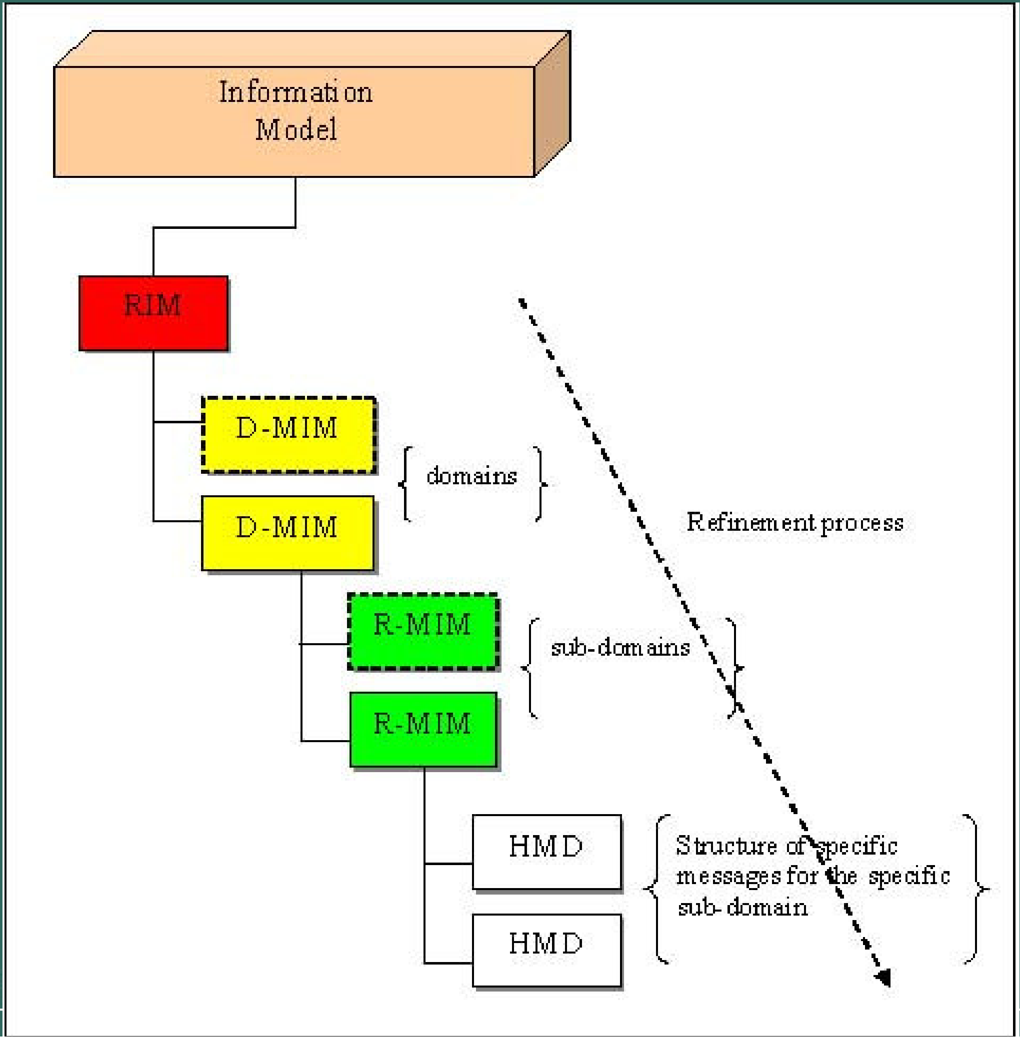
An **Entity** plays a **Role** which **Participates** in an **Act**

The RIM: about 60 classes derived from the backbone classes due to specialization. For example, from the Role class (backbone) has been derived the Patient, the Employee class and others. Each class has its own specific attributes

The backbone (the core classes): Entity, Role, Partecipation, Act, ActRelationShip and RoleLink. Only 6 classes to represent any specific entity in clinical domains

RIM, D-MIMs, R-MIMs and HMDs

The Information Model refinement and specialization process



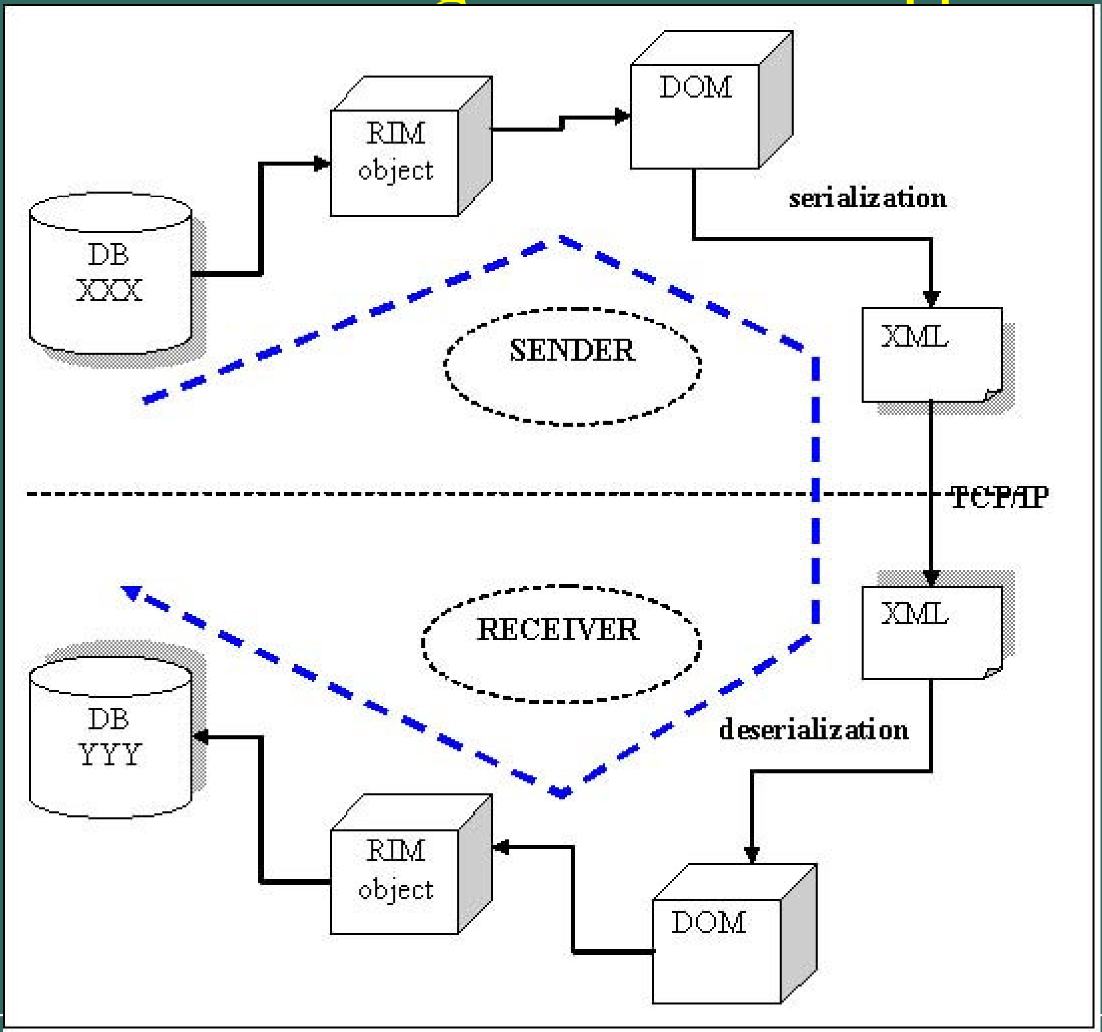
The refinement process involves the following models:

1. RIM: derived by the six backbone core classes;
2. D-MIM: derived by RIM, one D-MIM for each domain;
3. R-MIM: derived by D-MIM, one for each sub-domain;
4. HMD: the final hierarchical model for a specific message. HMD contains the final structure of the specific message

Using HL7 v3 inside applications

Example of exchanging an HL7 v3 message between two apps

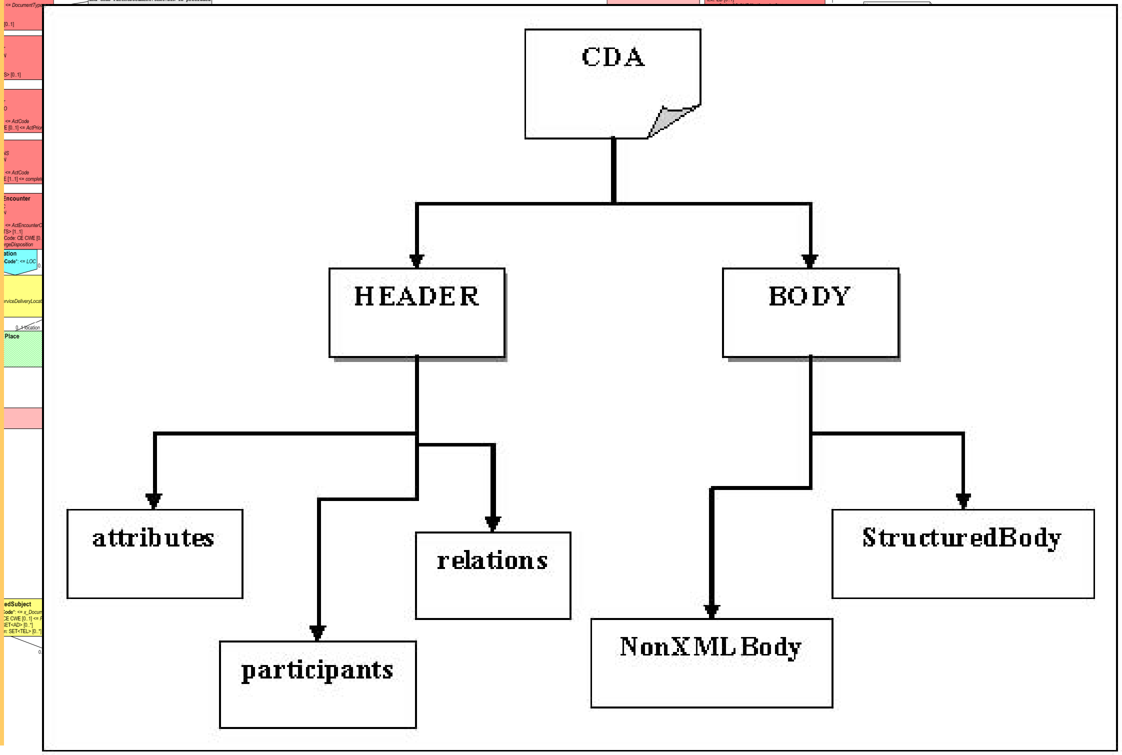
1. DB XXX (reading);
2. RIM object (Java API);
3. DOM object from Java classes;
4. Serialization (XML file from DOM);
5. sending over the net;
6. Deserialization (XML file to DOM);
7. Java classes from DOM object;
8. Rim object (Java API);
9. DB YYY (writing)



CDA: Clinical Document Architecture

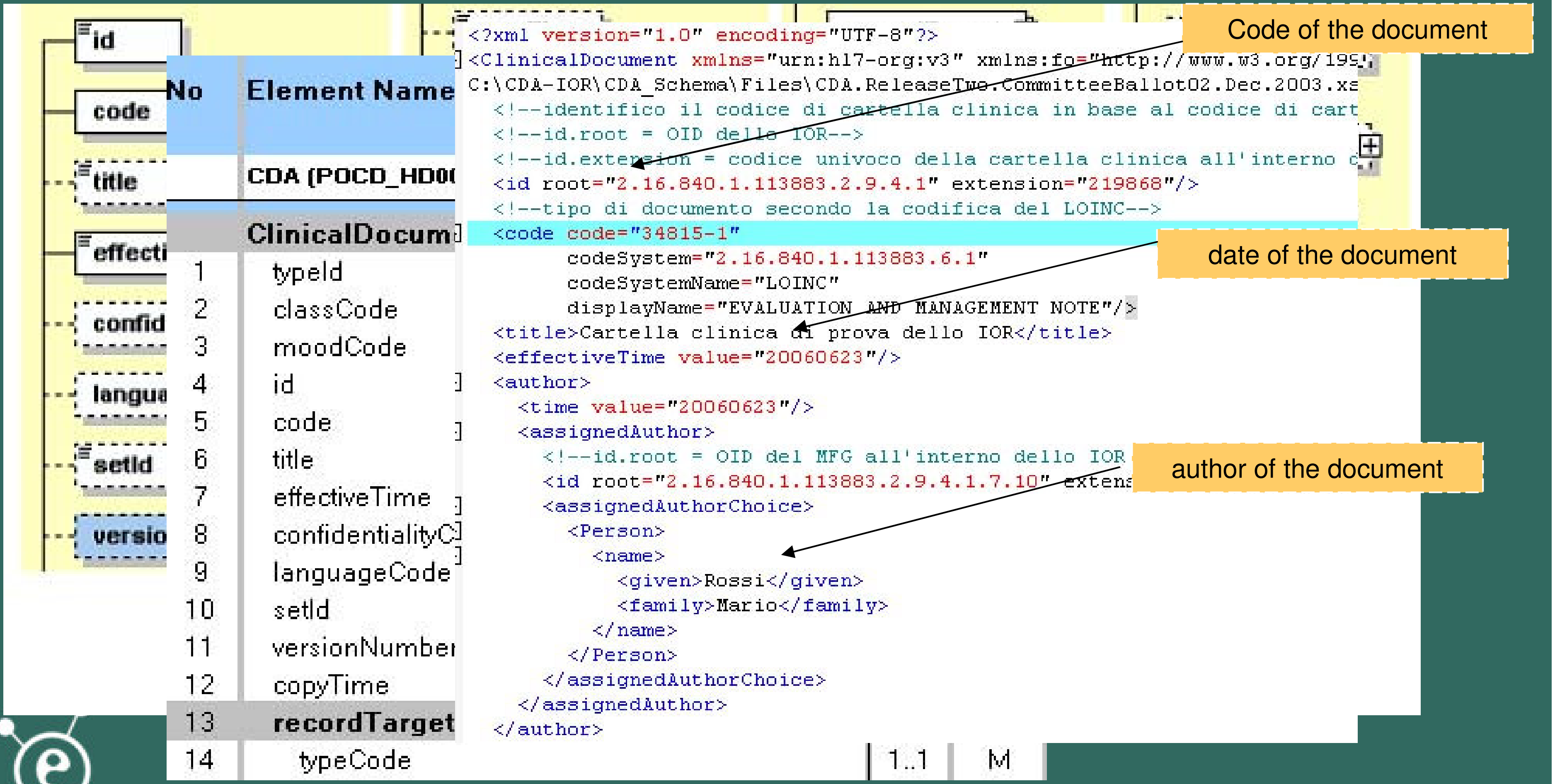
The CDA for health care records exchange

1. HEADER: contains general information about document id and type, relations with other CDAs, list of participants in different roles, personal data for the patient;
2. BODY: contains specific clinical data for the patient in a structured way.



The detailed CDA structure

The structure of a CDA message in details: an example of CDA



No	Element Name
	CDA (POCD_HD0
	ClinicalDocum
1	typeId
2	classCode
3	moodCode
4	id
5	code
6	title
7	effectiveTime
8	confidentialityC
9	languageCode
10	setId
11	versionNumber
12	copyTime
13	recordTarget
14	typeCode

1..1 M

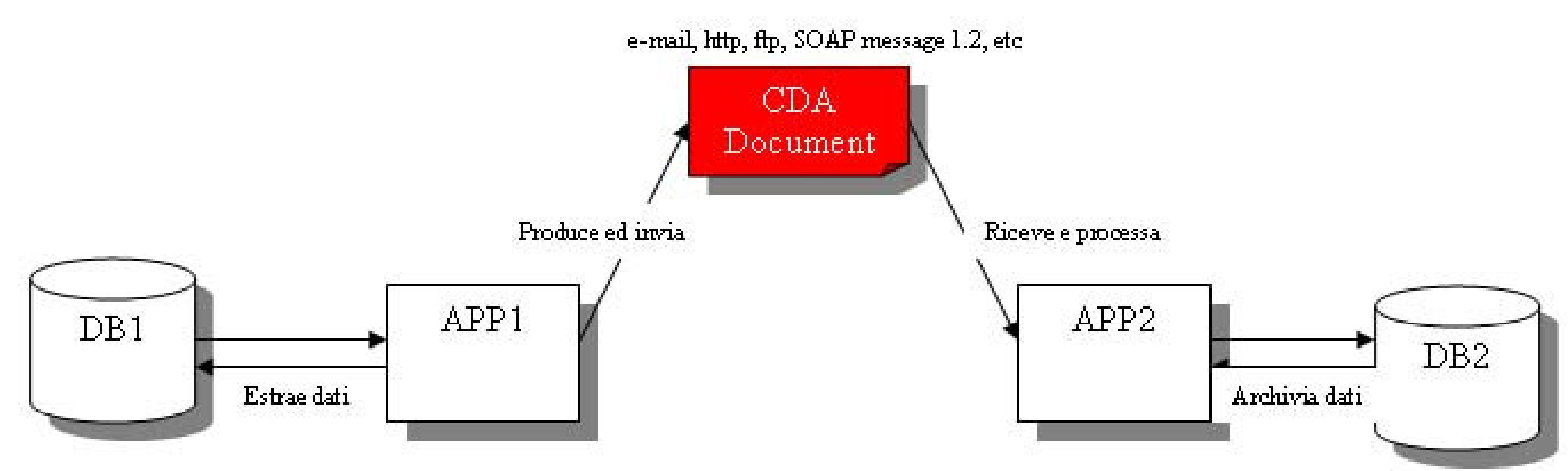
```
<?xml version="1.0" encoding="UTF-8"?>
<ClinicalDocument xmlns="urn:hl7-org:v3" xmlns:fo="http://www.w3.org/19
C:\CDA-IOR\CDA_Schema\Files\CDA.ReleaseTwo.CommitteeBallot02.Dec.2003.x
  <!--identifico il codice di cartella clinica in base al codice di cart
  <!--id.root = OID dello IOR-->
  <!--id.extension = codice univoco della cartella clinica all'interno d
  <id root="2.16.840.1.113883.2.9.4.1" extension="219868"/>
  <!--tipo di documento secondo la codifica del LOINC-->
  <code code="34815-1"
        codeSystem="2.16.840.1.113883.6.1"
        codeSystemName="LOINC"
        displayName="EVALUATION AND MANAGEMENT NOTE"/>
  <title>Cartella clinica di prova dello IOR</title>
  <effectiveTime value="20060623"/>
  <author>
    <time value="20060623"/>
    <assignedAuthor>
      <!--id.root = OID del MFG all'interno dello IOR
      <id root="2.16.840.1.113883.2.9.4.1.7.10" extens
      <assignedAuthorChoice>
        <Person>
          <name>
            <given>Rossi</given>
            <family>Mario</family>
          </name>
        </Person>
      </assignedAuthorChoice>
    </assignedAuthor>
  </author>
```

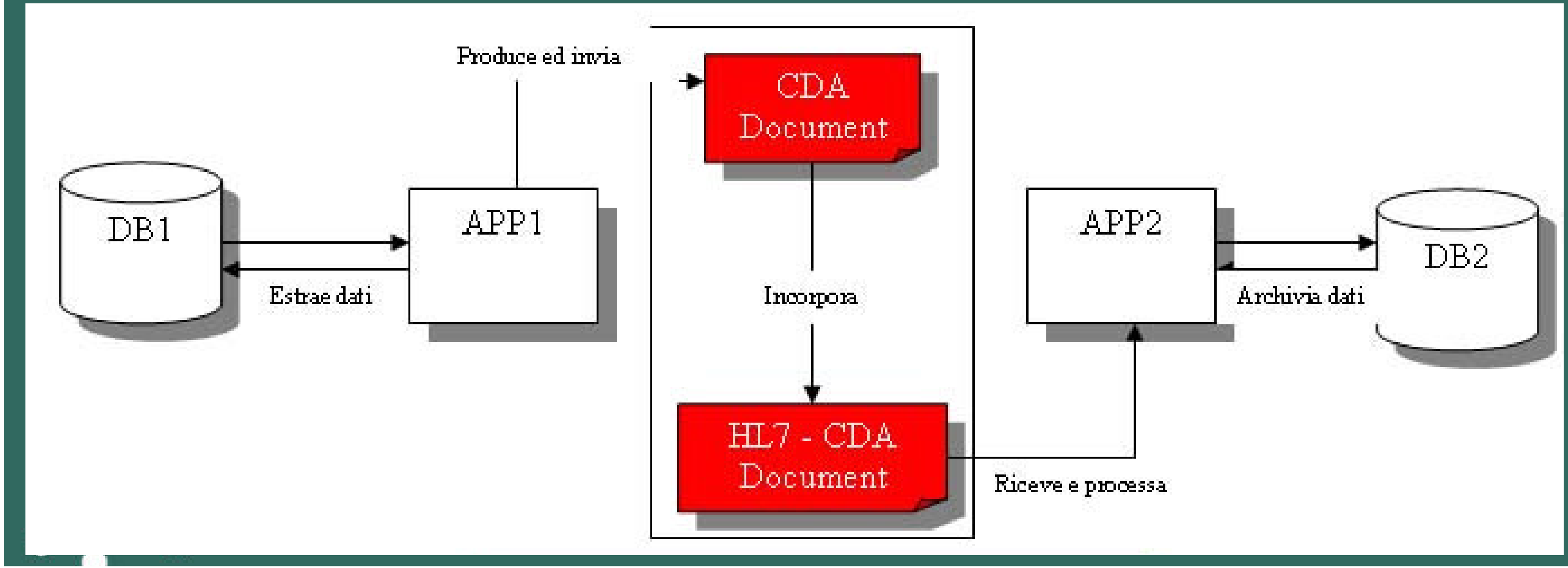
Code of the document

date of the document

author of the document

CDA message exchange example

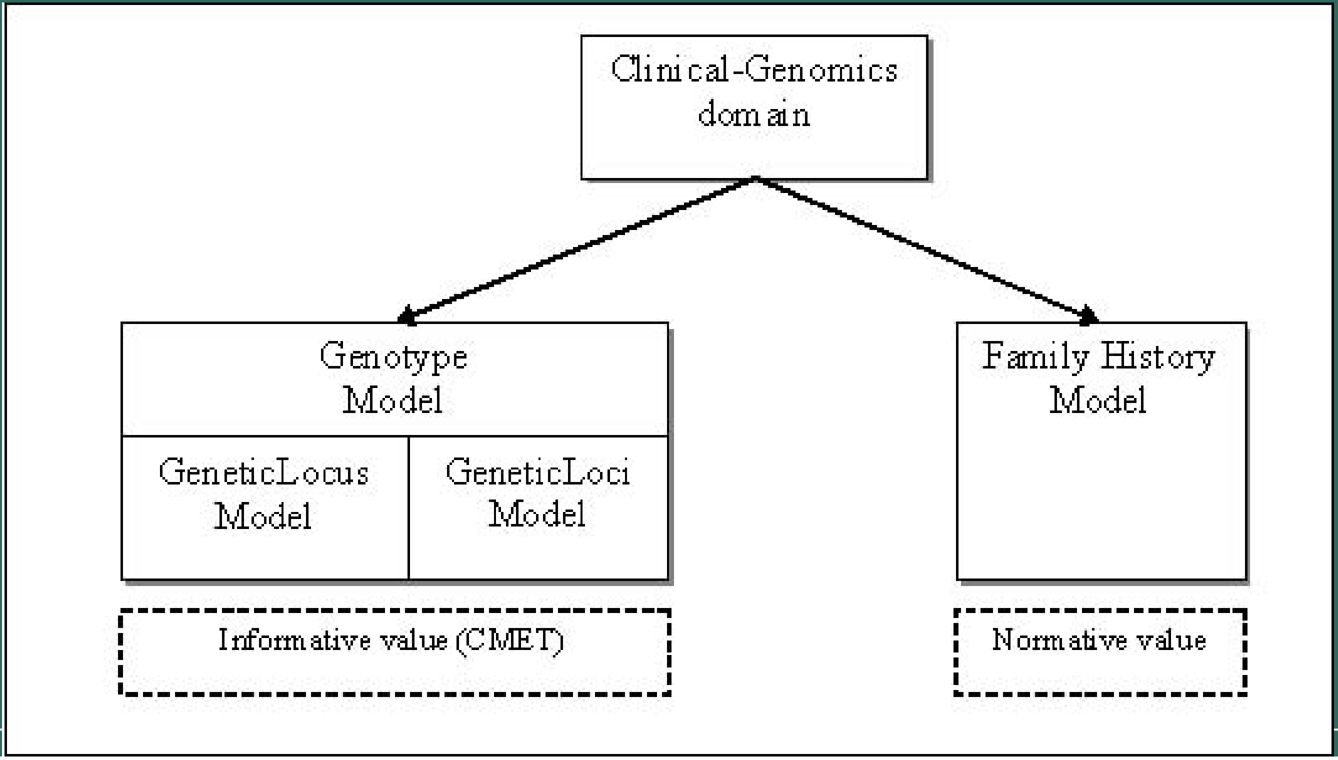




ClinicalGenomics model

CG HL7 SIG

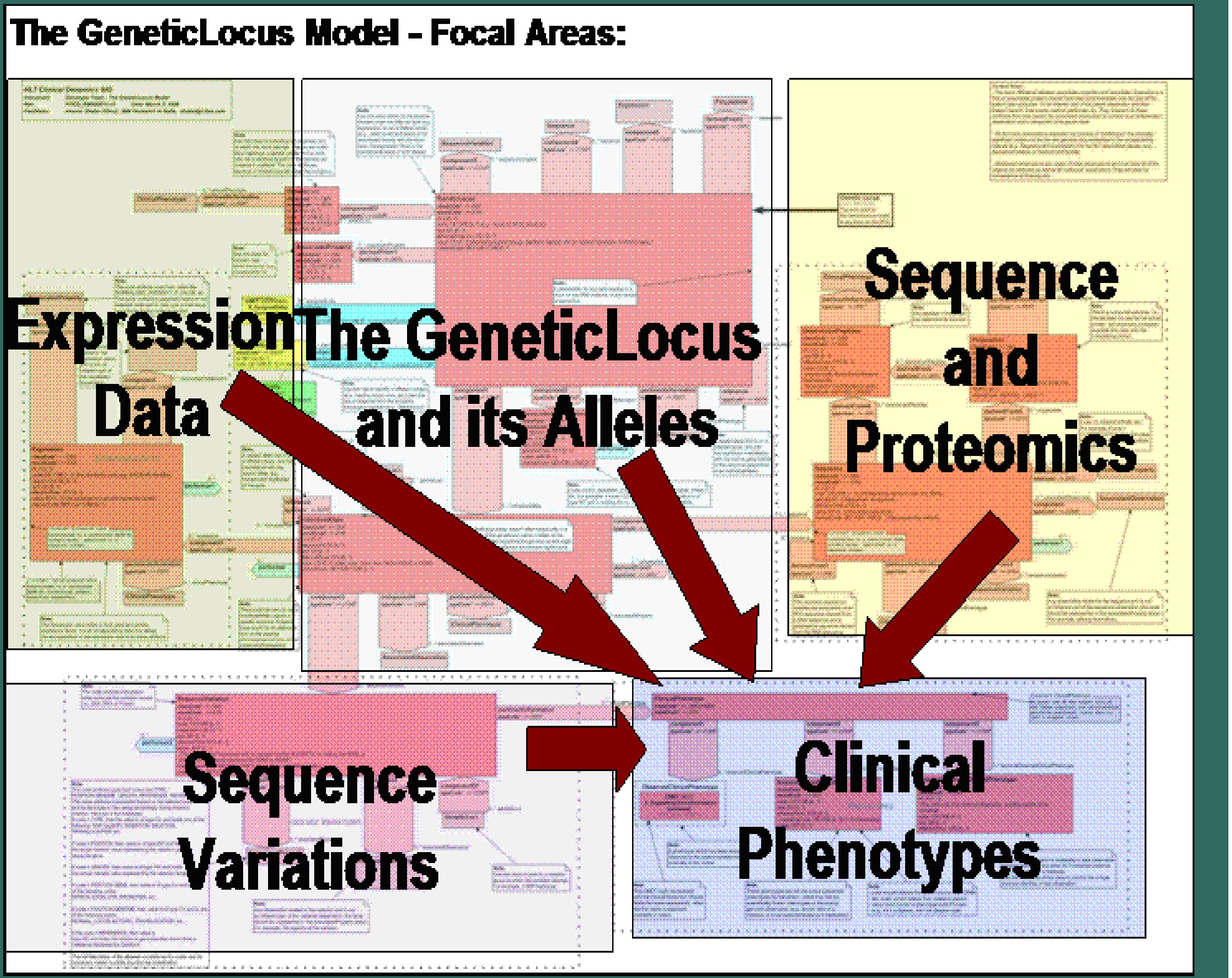
Inside HL7 a Special Interest Group for developing Clinical-Genomics specs was created to enable the communication between interested parties of the clinical and personalized genomic data. The main goal was to correlate clinical (phenotype) to genomic data (genotype) in order to facilitate the development of a personalized medicine.



The ClinicalGenomics model: focal areas

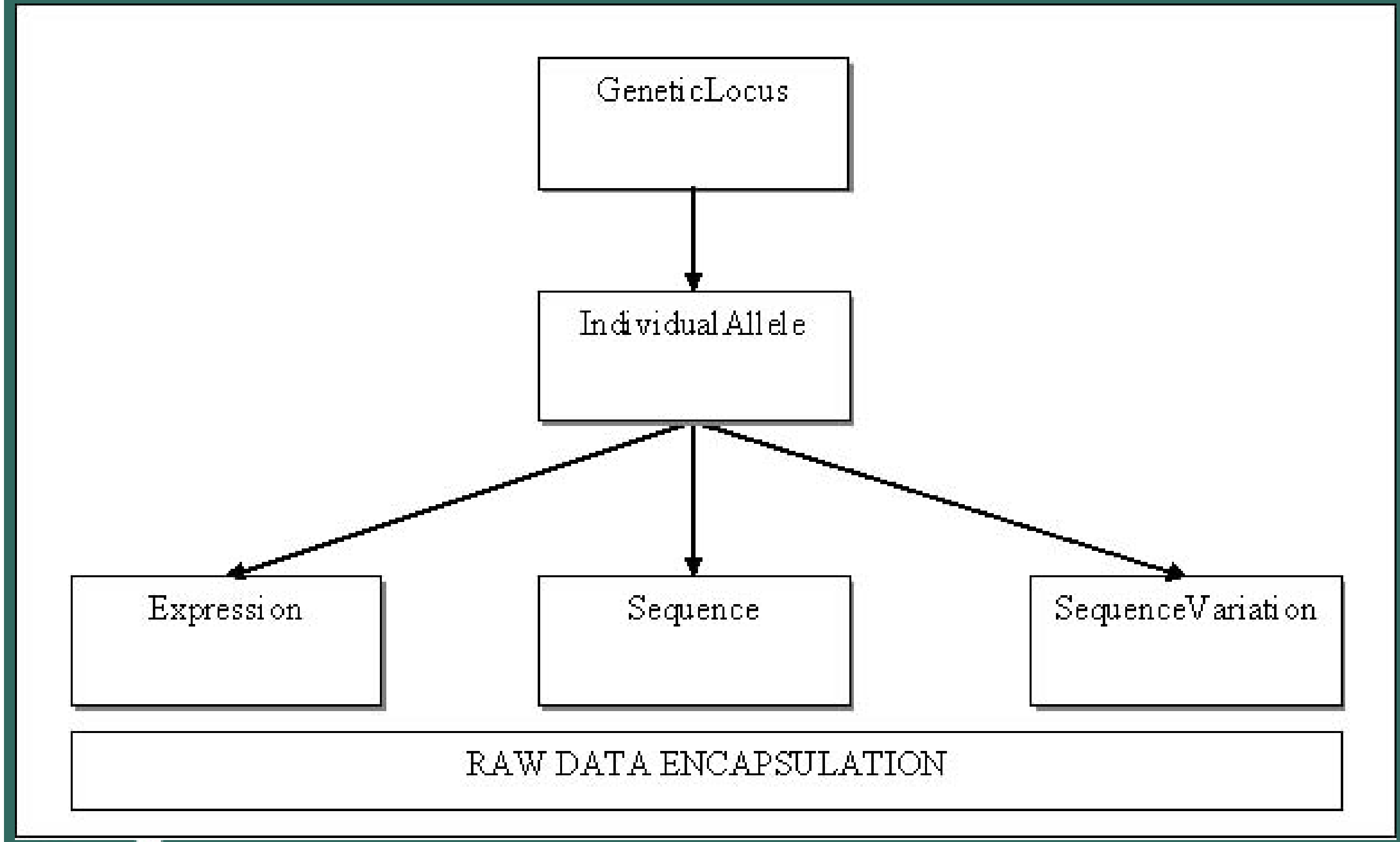
Focal areas

- Expression data (microarray)
- Genetic Locus and Alleles;
- Sequence and Proteomics;
- Sequence Variations (BSML);
- Clinical Phenotypes.



GeneticLocus Model

Classes for encapsulating and bubbling-up



GeneticLocus

IndividualAllele

Expression

Sequence

SequenceVariation

RAW DATA ENCAPSULATION

The GeneticLocus Model is built around a single Gene or Allele and provides classes for encapsulating genomic data expressed in existing markup language such as MAGE-ML (microarray data), BSML (sequence data), and others.

GeneticLoci and FamilyHistory Model

The GeneticLoci Model

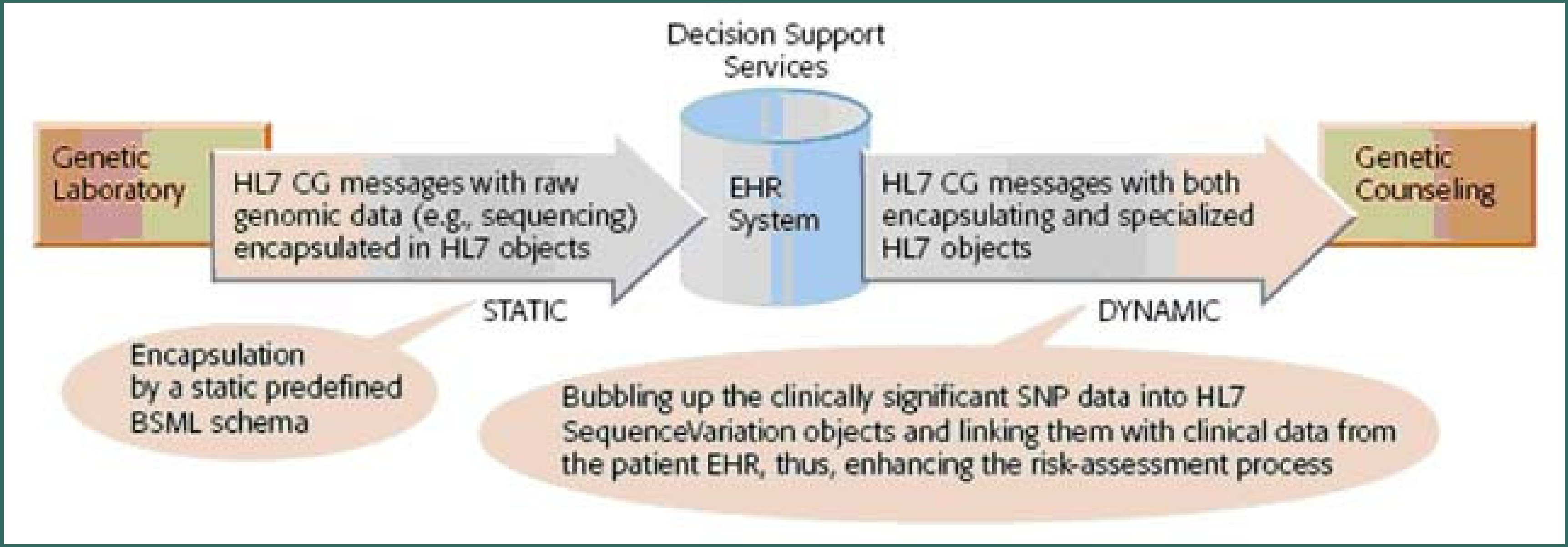
The GeneticLoci Model is a generalization of the GeneticLocus model: it represents a set of GeneticLocus instances related each other by different causes (haplotype, pathway, etc).

The FamilyHistory Model

The FamilyHistory model has been created to represent a patient's pedigree information as associated with clinical and genomic data.

ClinicalGenomics Model: exchange data

An example for exchanging data with the encapsulate and bubble-up paradigm



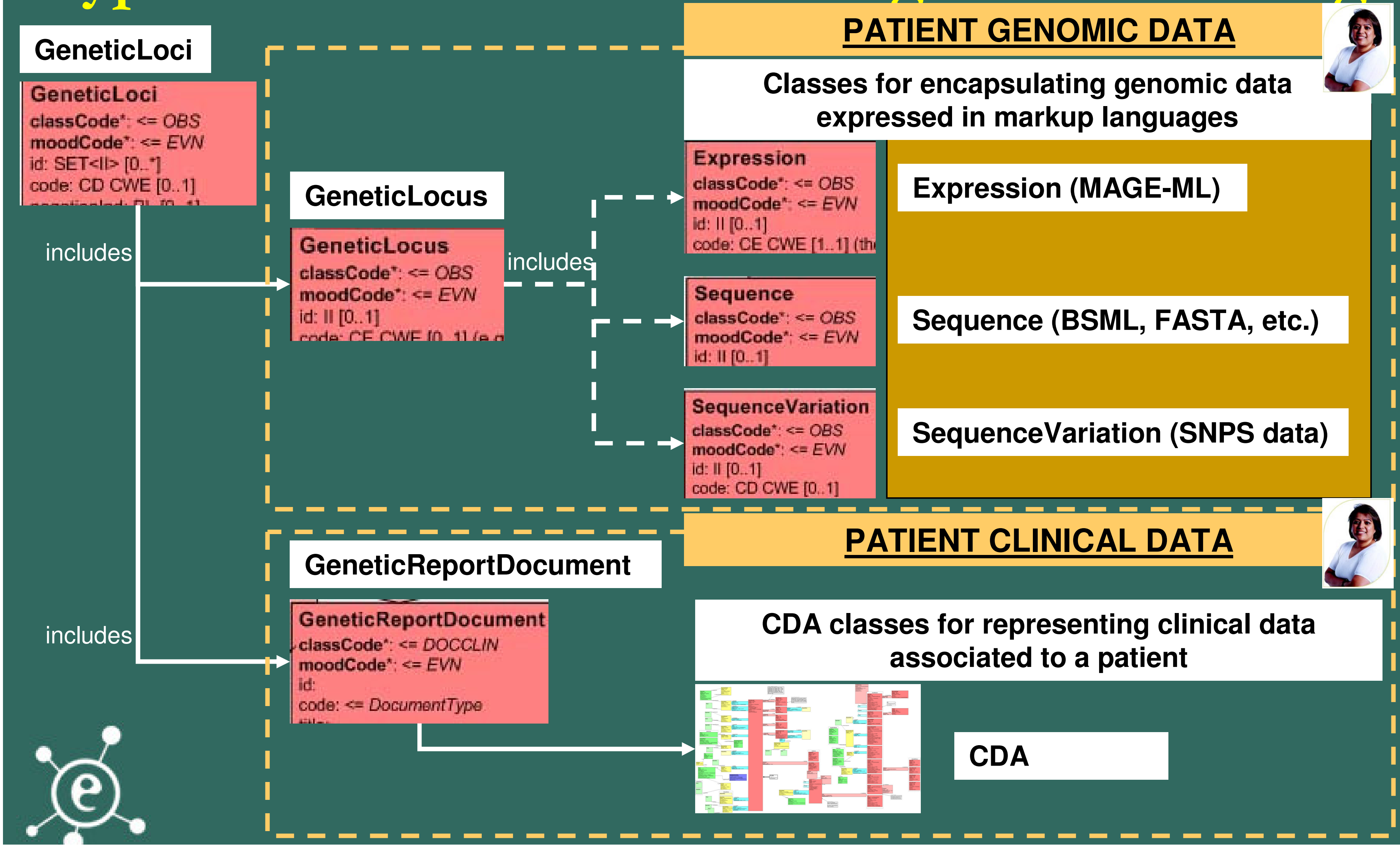
HL7 CG messages with raw genomic data (e.g., sequencing) encapsulated in HL7 objects

HL7 CG messages with both encapsulating and specialized HL7 objects

Encapsulation by a static predefined BSML schema

Bubbling up the clinically significant SNP data into HL7 SequenceVariation objects and linking them with clinical data from the patient EHR, thus, enhancing the risk-assessment process

Typical structure of a clinical-genomic message



GeneticLoci

GeneticLoci
classCode*: <= *OBS*
moodCode*: <= *EVN*
id: SET<II> [0..*]
code: CD CWE [0..1]

includes

GeneticLocus

GeneticLocus
classCode*: <= *OBS*
moodCode*: <= *EVN*
id: II [0..1]
code: CE CWE [0..1] (e.g

includes

PATIENT GENOMIC DATA

Classes for encapsulating genomic data expressed in markup languages

Expression
classCode*: <= *OBS*
moodCode*: <= *EVN*
id: II [0..1]
code: CE CWE [1..1] (th

Expression (MAGE-ML)

Sequence
classCode*: <= *OBS*
moodCode*: <= *EVN*
id: II [0..1]

Sequence (BSML, FASTA, etc.)

SequenceVariation
classCode*: <= *OBS*
moodCode*: <= *EVN*
id: II [0..1]
code: CD CWE [0..1]

SequenceVariation (SNPS data)

includes

GeneticReportDocument

GeneticReportDocument
classCode*: <= *DOCCLIN*
moodCode*: <= *EVN*
id:
code: <= *DocumentType*

PATIENT CLINICAL DATA

CDA classes for representing clinical data associated to a patient

CDA

Part 3 – Integrating HL7 models into the project

HL7: the standard of clinical and genomic data exchange

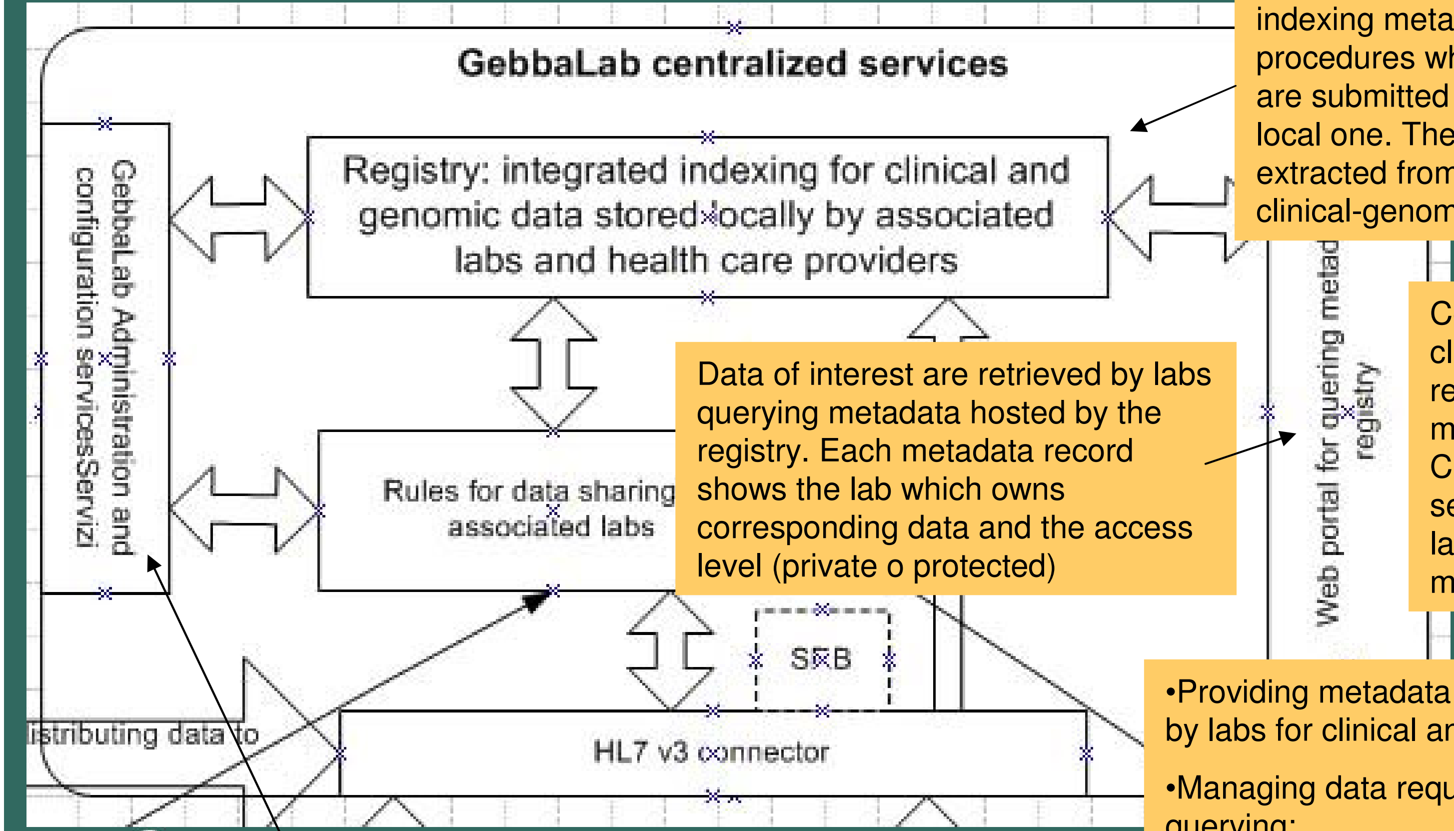
CDA and CG inside the GebbaLab project

Many advantages coming from adopting the CDA and CG models

1. HL7 models are patient centric: genotype and phenotype are wrapped around a specific patient;
2. Encapsulation of genetic data using native markup languages such as MAGE-ML, BSML: no need for defining further genomic standards;
3. Possibility of exchange clinical and genomic data among heterogeneous health care organizations using HL7 messages;
4. HL7 models useful for building database objects strictly mapped to CDA and ClinicalGenomics D-MIM entities;
5. Possibility of extending already available Java API (MAGE-ML and HL7 API) saving time and efforts;
6. High level of scalability and modularity that allows to build new modules around new labs which want to join the GebbaLab platform.

GebbaLab centralized services

Centralized services: components and sub-components



Registry: component for storing and indexing metadata extracted by specific procedures when clinical and genomic data are submitted to the central repository or a local one. The registry contains all metadata extracted from both public and private clinical-genomic data

Data of interest are retrieved by labs querying metadata hosted by the registry. Each metadata record shows the lab which owns corresponding data and the access level (private o protected)

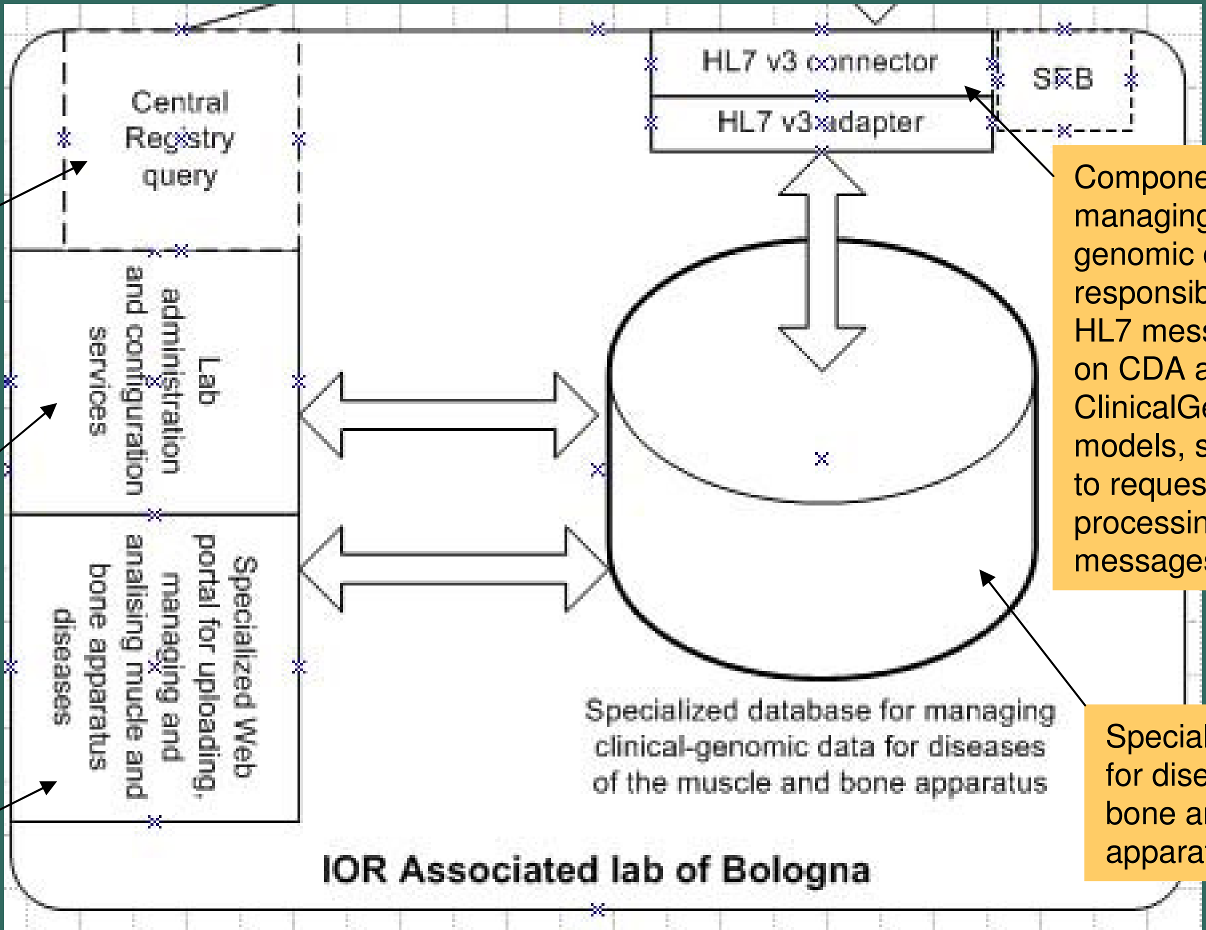
Component for managing clinical-genomic data responsible for building HL7 messages based on CDA and ClinicalGenomics models, sending them to requesting labs and processing received messages

- Providing metadata with the access levels defined by labs for clinical and genomic data;
- Managing data requests through metadata querying;
- Managing the messaging system for distributing requested data among labs.

Accessible throughout the web portal by system administrators, the component allows to manage users, groups, access levels and criteria (CRUD)

IOR associated lab of Bologna

Specialized services managed by IOR associated lab of Bologna



Local registry that enables lab's users to query directly the local repository and remotely the entire repository throughout the central registry. Local and central registries are always synchronized

Accessible throughout the specialized web portal by system administrators, the component allows to manage users, groups, access levels and criteria (CRUD) inside the lab

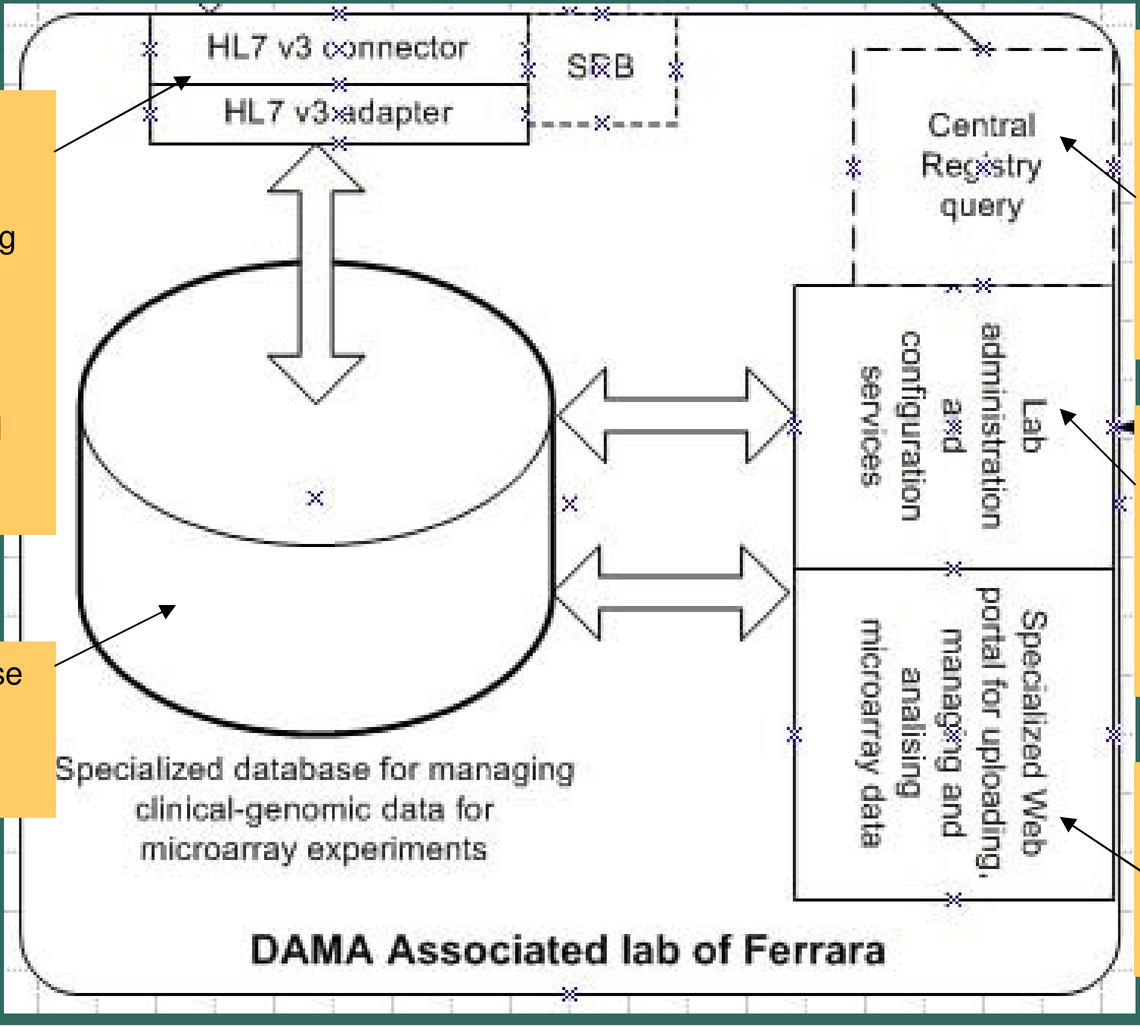
Specialized web portal with specific applications built around clinical-genomic needs of the IOR lab

Components for managing clinical-genomic data responsible for building HL7 messages based on CDA and ClinicalGenomics models, sending them to requesting labs and processing received messages

Specialized database for diseases of the bone and muscle apparatus

DAMA associated lab of Ferrara

Specialized services managed by DAMA associated lab of Ferrara



Components for managing clinical-genomic data responsible for building HL7 messages based on CDA and ClinicalGenomics models, sending them to requesting labs and processing received messages

Specialized database for expression data from microarray experiments

Local registry that enables lab's users to query directly the local repository and remotely the entire repository throughout the central registry. Local and central registries are always synchronized

Accessible throughout the specialized web portal by system administrators, the component allows to manage users, groups, access levels and criteria (CRUD) inside the lab

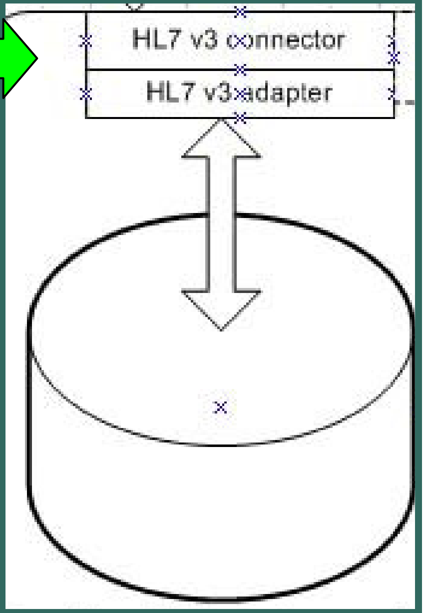
Specialized web portal with specific applications built around genomic (microarray experiments) needs of the DAMA lab

HL7 connectors: the glue for platform labs

The data exchange system among heterogeneous labs

Labs that need to exchange clinical-genomic data to each other are allowed to use components responsible for sending and receiving HL7 v3 message enveloped, for example, inside SOAP messages.

The glue of the architecture are **HL7 v3 connectors**



HL7 v3 connectors are responsible for:

- Reading and writing raw data from and to specific DBs;
- Mapping raw data to CDA and ClinicalGenomics HL7 v3 entities;
- Building HL7 v3 messages;
- Enveloping HL7 v3 messages, i.e. SOAP messages;
- Sending and receiving HL7 v3 messages to and from specialized web services responsible for processing data.

HL7 connectors: creating a message

An example of building an HL7 message to send data to associated labs that made a request

Typical operations executed by a reading-sending HL7 connector (in a Java thread):

Reading from the proprietary repository or filesystem of the lab that is requested to share its clinical and/or genomic data

Marshalling data to build in memory Java objects

Serializing in memory Java object to build HL7 v3 messages using the XML format

Sending HL7 v3 messages to another specific component responsible for managing message queuing

HL7 connectors: reading a message

An example of reading an HL7 message received by a sending lab

Typical operations executed by a receiving-writing HL7 connector (in a Java thread):

Receiving HL7 v3 messages from another specific component responsible for managing message queuing

↓

Deserialising HL7 v3 messages to build in memory Java object

↓

Unmarshalling in memory Java objects to build data

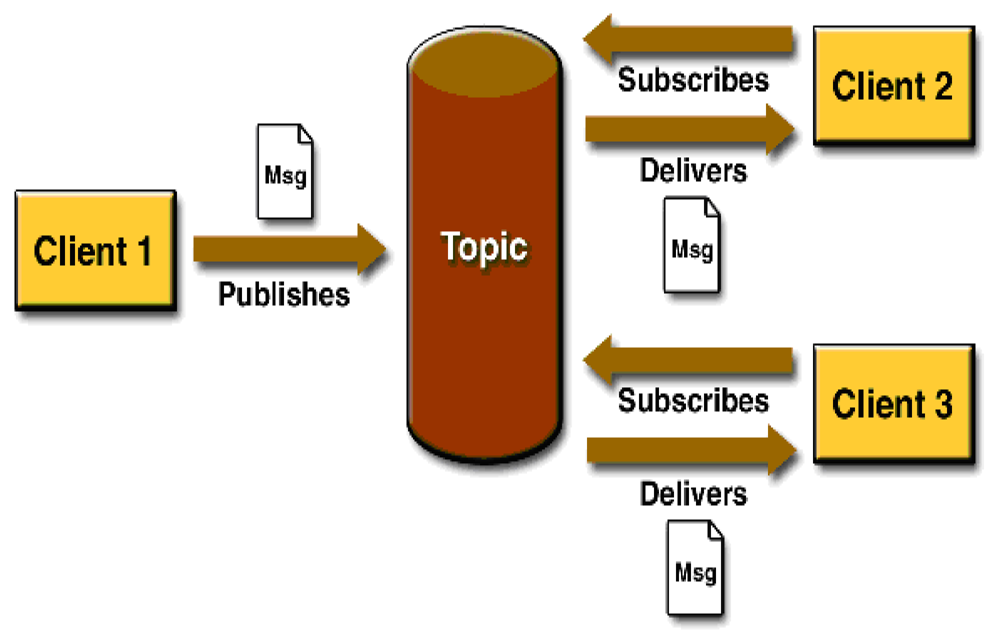
↓

Writing data to the proprietary repository or filesystem of the lab that has requested to share clinical and/or genomic data

The HL7 message queuing system

HL7 messages need a message queuing system

1. Once an HL7 message is built, it is sent to a "Message Queuing Manager" (MQM);
2. The MQM is a specialized component could be built using Java Message System (JMS) which listens to arriving messages;
3. The MQM collect messages and reads associated header to create dispatching rules (who, when);
4. The managing paradigm for messages is "publish and subscribe" in a n asynchronous way (see the figure)



Part 4 – Conclusion

- **Conclusion**

Privacy issues and sharing data

- Each lab is responsible for its data maintained locally, even if in case of virtualization of the lab (the lab decides not to store locally its data but to buy disk space on the central server). Labs manage access policies for groups and users;
- Clinical and genomic metadata are always public and allow users from different labs to search and retrieve data by means of the local or the central registry;
- Only authorized users are allowed to receive data from other labs (nodes). User who are interested in data from other labs, are allowed to ask them through a platform service. Once authorized, they can obtain data directly from the owner lab;
- Clinical and genomic data exchanged among labs are always transmitted in an anonymous way: no one possibility to link the patient or person data are relative to;
- Each transaction need to be based on private-public key cryptographic systems for the best level of security

Conclusion

- A modular platform opened to new health care providers;
- Every new lab, interested in associating to the "virtual laboratory", could need a customized application and an HL7 specialized connector;
- According to the previous need, we are interested in new projects connected to GebbaLab;

- We are evaluating different GRID technologies to individuate the best solution for the GebbaLab approach.

Thanks for your attention

??? QUESTIONS ???