Hospital experience in the use health data for improvement and research

Pablo Serrano Balazote
Planning Director
Hospital 12 de Octubre. Madrid Regional Health Service

Granted by Carlos III Health Institute projects PI15/00321 and PI15/00825
AGENDA

• About the hospital, primary care and research institute
  ▪ Regional health care information systems
  ▪ Data warehouse, business intelligence and management
  ▪ Healthcare, research and analytic systems

• Data reuse process

• Clinical information standardization
  ▪ regional shared repository
  ▪ methodology to define the normalized model
  ▪ controlled clinical terminologies

• Repositories experiences
  ▪ research repository ISO13606 standard base
  ▪ InSite platform

• Data quality assessment

• Challenges
University Hospital “12 de Octubre”

Primary Care

Madrid Regional Health Service SERMAS

public fudging and provision

population-base 420,000 people, south Madrid city

- 20 Primary Care centers,
- community hospital provide specialized care,
- 1,300 beds, 4 outpatient and 6 mental health centers
- tertiary referral regional and national, all specialties

Teaching: health professionals, grade and postgraduate

Research Institute “Instituto de Investigación i+12”

135 research professionals, 106 EUR/y

support units: genomics, biobank, clinical epidemiology…

9 research areas, 240 funding projects

220 clinical trials, ECRIN centre

year cumulate impact factor 2500
National repository *HCD SNS*

Regional view layer, professional *HORUS* and personal "*carpeta de salud*"

Mandate. Delay. Appointment

e prescription

**EHR Hospital**
*HCIS, LIS, PACs…*
H. 12 Octubre EHR 2011, HIMSS EMRAM stage 6

**EHR Primary Care**
*AP Madrid*
EHR 1996, Regional normalized 2007

**Population information** *Cibeles*
The Hospital has Business Intelligence applications

Data origins from the secondary standby databases of EHR and departmental systems

- Created by ODBCs which gets the data.
- Data are transformed by dynamic tables
- Distribution to professionals
- Created by automatic processes that allows data extraction, transformation and load in data marts (management, quality and research)
- Scorecard access to professionals
HEALTHCARE, RESEARCH AND ANALYTIC SYSTEMS

Analytic information systems

Quality Bench
Secondary use repositories
Research ISO 13606 repositories
InSite
Pharma

Research information systems

Registries
Genomic
Biobank
Research data bases
eCRF

Healthcare information systems

EHR HOS
EHR PC
Phar HOS
LIS
GEN
Path
ONC
EMR
CAR
RIS
RAD
NEP
DOL
NEU
HEM
...
CLINICAL DATA REUSE PROCESS

- Governance
- EHR normalization
- Clinical models definition
- Archetypes and terminology mappings
- ETL
- Use case
- Data quality and Validation studies
- Repositories performance tests
- EHR improvement

Iterations new data sources
Iterations new purposes
• Progress in share clinical information to facilitate continuity of care
• Use existing EHR systems in primary care, hospitals and electronic prescriptions
• The new shared system must evolve current regional and national repositories based on document
• Towards the organizational and semantic interoperability
• Enable the Secondary use of health data
Apply the lessons learned with ISO 13606 in advanced platforms (UPV LinkERH and ISCIII) about:

- normalized heterogeneous EHR
- communicate to other systems
- persist in sharable data repository based on the dual model

Need a common and corporate information model, transversal to care levels and providers, and cumulative longitudinal in time

They has to cover high organizational concepts, record context, clinical data structures and terminology
Concurrent used of health informatics ISO advanced standards

- ISO 13940 for the concepts of continuity of care
- ISO 13606 reference model for clinical information

Sequence

1. organization information model with the intersection of the standards: organizational concepts, processes, and the reference of the clinical record context
2. clinical detailed information models: editing archetypes and binding terminology

Heterogeneous concepts of current systems have been matched through a correspondence schema based in the model of ISO 13940 and ISO 13606 reference model.
## CONTROLLED CLINICAL TERMINOLOGIES

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Scope</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD 10 MC</td>
<td>Hospital diagnostics and procedures</td>
<td>EHR HCIS</td>
</tr>
<tr>
<td>ICD 9 MC (-2016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICPC2</td>
<td>GP health problems</td>
<td>Primary Care</td>
</tr>
<tr>
<td>Mapping to SNOMED CT</td>
<td>Alerts, Allergies, Diets, Anatomical location, Observations, Orders</td>
<td>EHR HCIS, LIS</td>
</tr>
<tr>
<td>SNOMED CT, ICD O</td>
<td>Morphology and topography</td>
<td>Pathology</td>
</tr>
<tr>
<td>NANDA-NIC-NOC</td>
<td>Nurse diagnostics, interventions and objectives</td>
<td>EHR HCIS</td>
</tr>
<tr>
<td>LOINC</td>
<td>Laboratory</td>
<td>LIS, EHR HCIS</td>
</tr>
<tr>
<td>SNOMED CT, ATC</td>
<td>Medications</td>
<td>EHR HCIS, Pharmacy, Primary Care</td>
</tr>
<tr>
<td>ICPC2</td>
<td>Emergency diagnostics</td>
<td>Emergency</td>
</tr>
<tr>
<td>SERAM</td>
<td>Spanish radiology procedures</td>
<td>RIS</td>
</tr>
</tbody>
</table>
ISCIII

COMBINE

• PRIMARY CARE
• HOSPITALS
IMPLEMENTATION STRATEGY OF STANDARD

Data Format

is represented by

Information sources

AP Madrid
Pharmacy
HCIS
...

LinkEHR

Rules

Integration engine

generate

Uses

Extract 13606

Repository

viewer

Design | Execution

Repository

Repository

LinkEHR Information sources

Data Format

is represented by

Archetypes
INSITE
SOURCE SYSTEMS, TERMINOLOGIES AND PHASES

Demographics
- 1: IS demographics

Contacts
- 1: EHR hospital
- 3: EHR primary care

Diagnostic Procedures
- 1: EHR hospital
- 2: EHR hospital
- 3: EHR primary care

Medication
- 1: EHR hospital
- 2: IS pharmacy
- 3: e-prescription

Laboratory
- 1: IS laboratory

Pathology
- 1: Cancer Registry
- 2: IS pathology

Clinical findings
- 1, 2, 3: EHR hospital
- 3: EHR primary care

Inpatient
Outpatient
Primary care

terminologies
- natives
- official mappings
- local mappings

phases 1, 2, 3

Use case
### Demograph (ID, Age, Sex...)  
### Patient Mapping (EHR ID, regional...)  
### Provider (Physicians, nurses...)  
### Contacts (A clinical transaction generating a set of data)  
### Diagnoses  
### Procedures  
### Lab Test  
### Medications

<table>
<thead>
<tr>
<th>Period / Context</th>
<th>Inpatients</th>
<th>Inpatients</th>
<th>Inpatients</th>
<th>Inpatients</th>
<th>Inpatients</th>
<th>All</th>
<th>Inpatients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inpatients</strong></td>
<td>2.672.821</td>
<td>8.018.462</td>
<td>43.053</td>
<td>22.462.945</td>
<td>4.337.910</td>
<td>3.205.190</td>
<td>148.761.307</td>
</tr>
<tr>
<td><strong>#</strong></td>
<td>100%</td>
<td>&gt;99%</td>
<td>100%</td>
<td>&gt;99%</td>
<td>100%</td>
<td>100%</td>
<td>95.9%</td>
</tr>
<tr>
<td><strong>% Load</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1191 without birth_date</strong></td>
<td>Every patient has 3 linked PatientIds (main code, EHR, IUP)</td>
<td>Rejected visits are due to the time difference between the extraction of the demographic and visit data. 12O data is extracted from a live system</td>
<td>Diagnosis ICD code load depends on 12O coding department</td>
<td>Procedure ICD code load depends on 12Ocoding department</td>
<td>Missing patients, providers &amp; visits are due to the time difference between their extraction lab data. 12O data is extracted from a live system</td>
<td>Missing patients &amp; visits are due to the time difference between their extraction and medicine data. 12O data is extracted from a live system</td>
<td></td>
</tr>
<tr>
<td>DQ dimension</td>
<td>Hospital Dataset</td>
<td>Hospital 12 de Octubre</td>
<td>Hospital Virgen del Castillo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------</td>
<td>------------------------</td>
<td>------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PRE-stand. n=1949</td>
<td>POST-stand. n=1948</td>
<td>PRE-stand. n=3781</td>
<td>POST-stand. n=3776</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Uniqueness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-replicated identifiers</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Completeness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-missing data, weighting</td>
<td>76.71%</td>
<td>8.44%</td>
<td>56.60%</td>
<td>18.03%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>obligatory and optative elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Consistency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformance to basic schema rules</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Temporal stability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data concordance over time</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multi-source stability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data concordance among different sources (1-GPD metric)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.08 (among hospitals)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Correctness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possibly anomalous records</td>
<td>3.39%</td>
<td>0.62%</td>
<td>1.38%</td>
<td>0.45%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Predictive value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usefulness of data to predict breastfeeding at one month</td>
<td></td>
<td></td>
<td></td>
<td>Not applicable</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TECHNOLOGICAL CHALLENGES

ETL normalization using artefacts standard ISO 13606

Extend repositories standards based for primary and secondary use

Automated population of secondary use repositories with define specific models: i2b2 and OMOP
Acknowledgments

Teams projects

Standard ISO 13606 repository

Grant

Advisor

In Site

Contact pserranob@salud.madrid.org