Semantic Interoperability:
HL7 Vs "Self Describing" XML

Stephen Chu

NZ HL7 User Group
HINZ
stephen.chu@auckland.ac.nz

Quarterly Seminar – 16 June 2005, Auckland

The Fallacy of "Self-Describing" XML

- **XML (eXtensible Markup Language) has been described as:**
  - **SEXY**
  - **EASY (to use and implement)**
  - **Highly FLEXIBLE (you can create your own tags)**
  - **HUMAN READABLE**
- **These views also raise the questions of:**
  - **Why do we need HL7 any more?**
  - **Shouldn't HL7 now go away?**

The Fallacy of "Self-Describing" XML

- XML is
  - A technology/language framework for structuring data for transmission from the source system to one or more target systems
  - Used in formatting data items using tags that are delimited by <> pairs, e.g.
    - <name>Stephen Chu</name>

The tag represents "attributes" that qualifies/describes the data

The data that are placed between the tags has data type constraints and values

The Fallacy of "Self-Describing" XML

- Tagged data elements may be ordered and grouped to express values of more complex concepts:
  - <Person_Name>
    - <First_Name>Stephen</First_Name>
    - <Middle_Name>Popular</Middle_Name>
    - <Last_Name>Chu</Last_Name>
  - </Person_Name>

The Fallacy of "Self-Describing" XML

- Relationships between data items can also be expressed:
  - <Physical_Exam>
    - <Finding>Erythematous rash, palmer surface, left index finger
      <renderMultiMedia referencedObject MMID="MM1">
    - </Finding>
    - <observationMedia MMID="MM1">
      <id root="10.23.4567.345"/>
      <value xsi.type="ED"
      mediaType="image/jpeg">
      <reference value="left-hand-image.jpeg"/>
      </value>
    - </observationMedia>
  - </Physical_Exam>

Textual description → <Finding>Erythematous rash, palmer surface, left index finger

Graphics → <reference value="left-hand-image.jpeg"/>

related to

The Fallacy of "Self-Describing" XML

- Express relationships between data items:

```
<Observation>
   <code code="F-0196A" codeSystem="2.16.840.1.113883.6.5"
   codeSystemName="SNOMED" displayName="History finding"/>
   <effectiveTime value="1970"/>
   <value xsi:type="CD" code="D3-15000"
   codeSystem="2.16.840.1.113883.6.5" codeSystemName="SNOMED"
   displayName="MI"/>
   <subject>
      <subjectRole classCode="PRS">
         <code code="FTH"/>
      </subjectRole>
   </subject>
   <sourceOf typeCode="CAUS">
      <Observation>
         <code code="F-0196A" codeSystem="2.16.840.1.113883.6.5"
         codeSystemName="SNOMED" displayName="History finding"/>
         <effectiveTime value="1970"/>
         <value xsi:type="CD" code="DF-D0000"
         codeSystem="2.16.840.1.113883.6.5" codeSystemName="SNOMED"
         displayName="death"/>
      </Observation>
   </sourceOf>
</Observation>
```

The Fallacy of "Self-Describing" XML

- Rules of encoding and annotating the data stream are usually placed in a separate document – xml schema
- The potential richness of the annotation of the data stream, together with adherence to some predefined set of rules (schema) has led to the misconception that:
  - Data streams expressed in XML are "self-describing"
    - Implies that meaning of the XML encoded data stream can always be discerned by receiving system

The Fallacy of "Self-Describing" XML

- Having an XML Schema does not always guarantee semantic interoperability
- One system may define patient demographic data as:
  - <name>
  - <Address_Line_1>
  - <Address_Line_2>
  - <Address_Line_3>
  - <Address_Line_4>
  - <Address_Line_5>

Another system as:

```
<Name>
    <First_Name>
    <Middle_Name>
    <Family_Name>
</Name>
<Address>
    <Street>
    <Suburb>
    <…….>
</Address>
```

The Fallacy of "Self-Describing" XML

- Worse, still – a third system may use the following encoding rule:
  - <kzjaqwm>Stephen</kzjaqwm>
  - <*c*a*t>Popular</*c*a*t>
  - <mtcypzb>Chu</mtcypzb>
- How "self-describing" are these tags?
- They are "human readable" in the sense that human can read and try to guess/infer that the data stream relates to a person name
- BUT XML is NOT intended to be read primarily by human

The Fallacy of "Self-Describing" XML

- What would a machine do if it receives those data streams?
  - Attempts to parse it
    - Possible as the parsing software understands the XML syntax
  - Attempts to process it
    - But semantics of the encoding may not be understood leading to
      - Confusion
      - Inability/difficulty in validating the data received

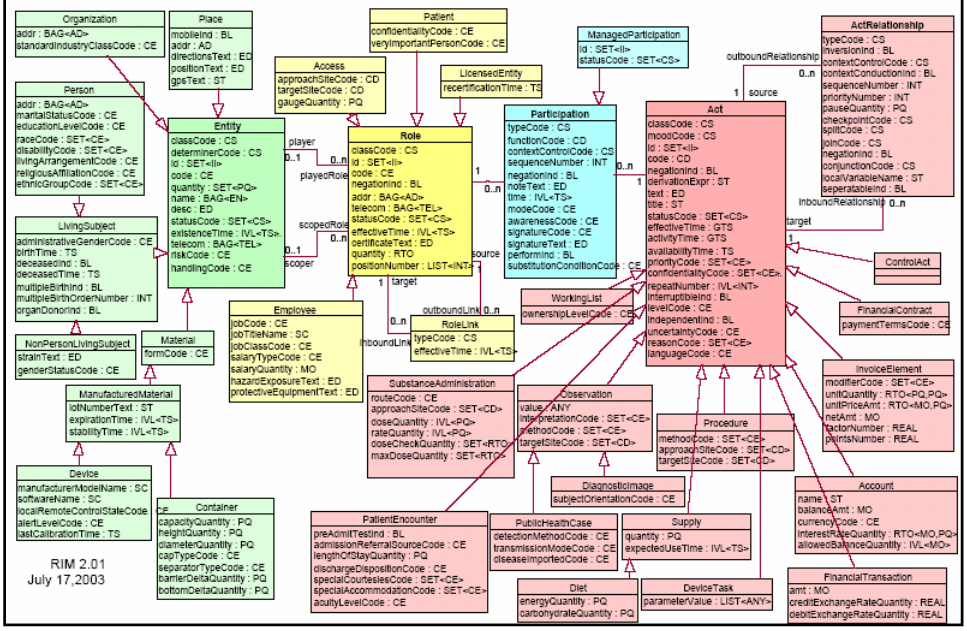
The Fallacy of "Self-Describing" XML

- Even if processing is made possible by use of XML schema, which supposedly "describes" the content of the data stream
  - Each system needs to use multiple schema to process and validate data from multiple systems
  - This "impedance mismatch" is no less problematic than having dedicate interfaces for different communicating systems
- **The need for a common conceptual model to which all communicating system conform cannot be eliminated simply by the use of XML**

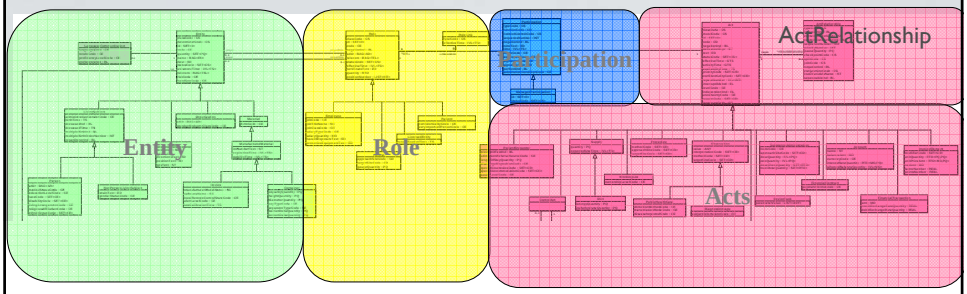
What is this common conceptual model?

- HL7 provides a common conceptual model:
- The HL7 Reference Information Model (RIM)
  - A model of clinical information content
  - Provides a form way of expressing
    - Data items and their definitions as standard structures
    - Data types
    - Relationships between data items
    - Contexts which govern the meaning or interpretation of the data values

The HL7 Reference Information Model

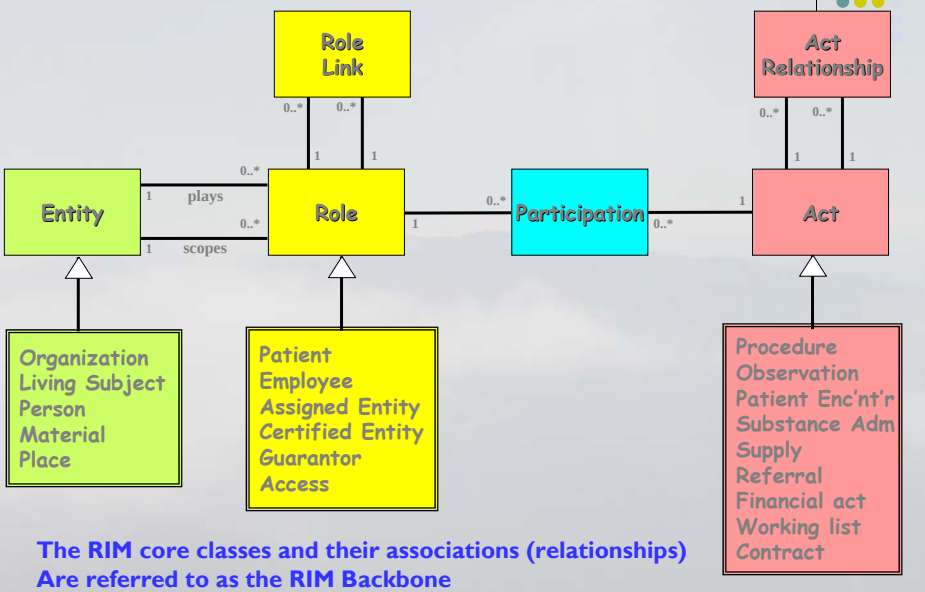


The HL7 Reference Information Model (RIM)



- 4 **Primary Subject Areas**
- 35 **Classes**
- 181 **Attributes**
- 9 **Associations**
- 28 **Generalizations**

The HL7 RIM



The RIM core classes and their associations (relationships)
Are referred to as the RIM Backbone

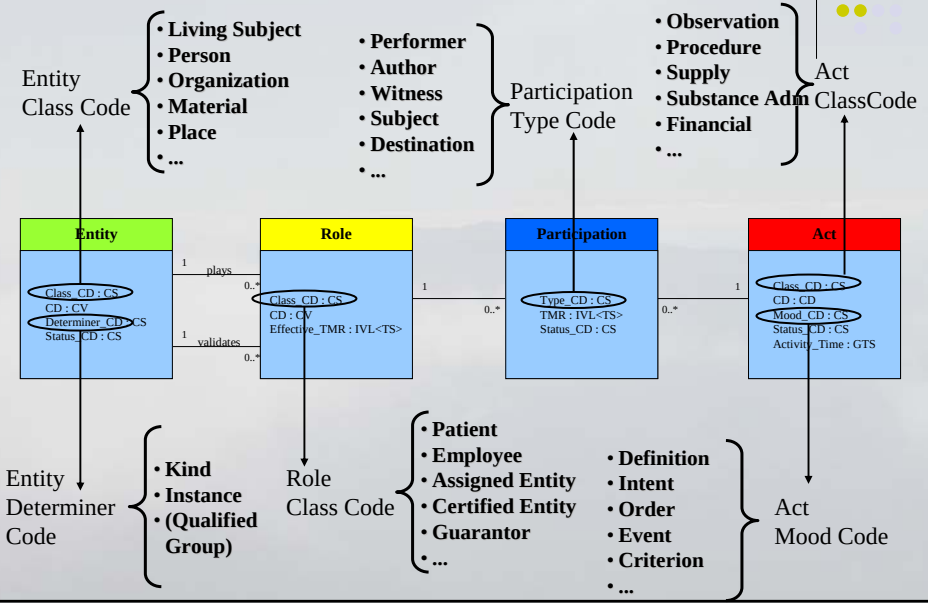
The HL7 Reference Information Model (RIM)

- All healthcare events/activities are **ACT**s
  - Observations, medication prescriptions and administrations, procedures, supply, etc
- **ACT**s are related through **Act_relationship**
  - Composition, preconditions, revisions
- **Participation** defines the context for an Act
  - Author, performer, subject, location,etc
- **Roles** (as participants) engage in participations
  - Patient, provider, practitioner, specimen
- **Roles** are played by **Entities**
  - Person, organization, material, place, devices

The HL7 RIM

- The RIM provides a set of standard structures
  - Entity, Role, Participation, Act, Act_Relationship
- These standard structures are supported by
  - Standard HL7 (coded) names/vocabularies which provide coded value for their attributes (examples – next slide)
- Coded attributes in the RIM must be associated with one Vocabulary Domain (value) prior to being used in a message specification
- A vocabulary domain may be specified as an enumerated list of
  - coded concepts (HL7 defined) or
  - as a reference to an externally maintained list of coded concepts (e.g., SNOMED, LOINC, CPT)

RIM Core Attribute Value Set



Combing RIM and Standard Vocabulary

- When standard (reference) terminologies are mapped to a standard (reference) information model
  - we have a basis for interoperability of information across all applications and all users
- The Reference Model and vocabulary domain provide clear meaning to the information in clinical context, e.g.
  - WBC (white blood count) may be used to name an order (intent), or a result (observation), or a goal (target value), and a problem (extreme value)

RIM and Vocabulary Domain Example

3 instances of *ACT*, specialization as *observation*, act.code attribute = "wbc"
(NOTE: "wbc" should be replaced with the appropriate LOINC code for WBC)

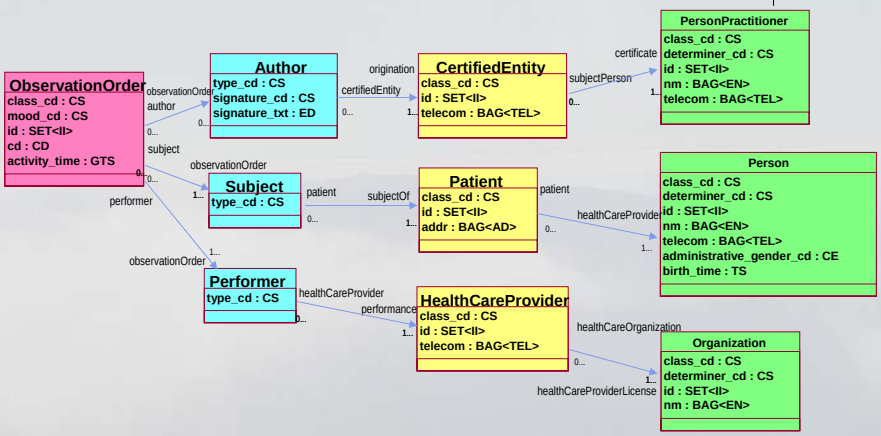
Mood=RQO	Mood=EVN	Mood=Goal
effectiveTime ="12/12/04@ 2pm PST"	effectiveTime ="10/1/04@ 1pm PST"	effectiveTime ="11/1/04"
(No Value Attribute in this model)	(Observation) Value="xxx"	(Goal) Value="yyy"
Wbc ***Order*** to occur on 12/12/04@2pmPST	Wbc ***Obs***ervation of 'xxx' made on 10/1/04@1pmPST	Wbc ***goal*** of specific value of 'yyy' to be attained by 11/1/04

Using the RIM to Model Reality

Person A → Practitioner → Performer → Listen to Lung/Chest

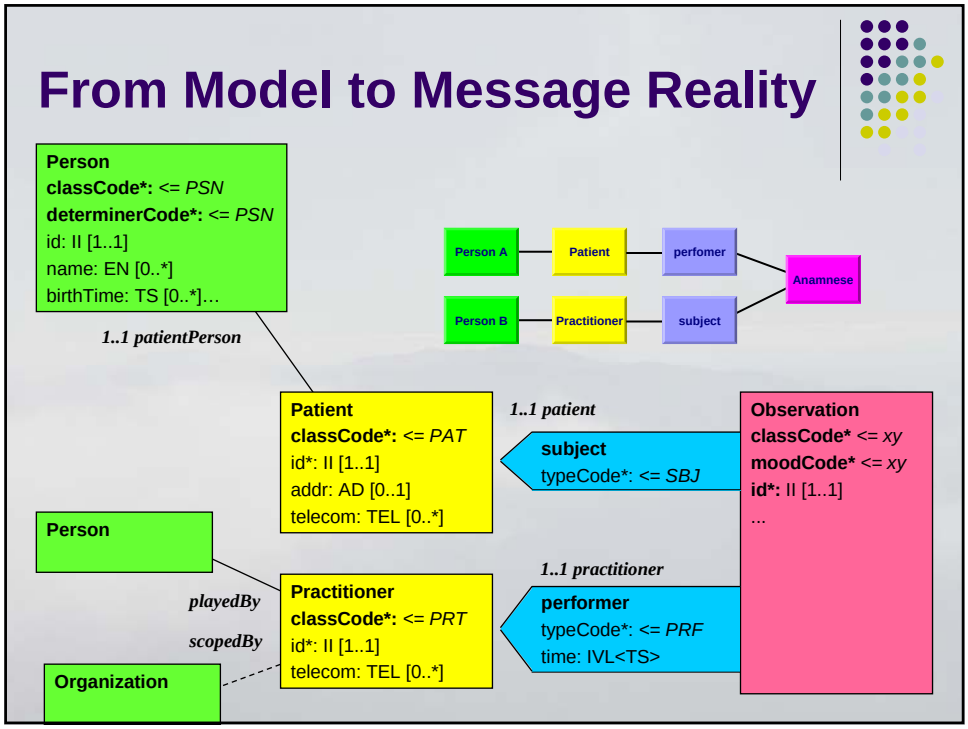
Person B → Patient → "Subject" → Listen to Lung/Chest

Using the RIM to Model Reality

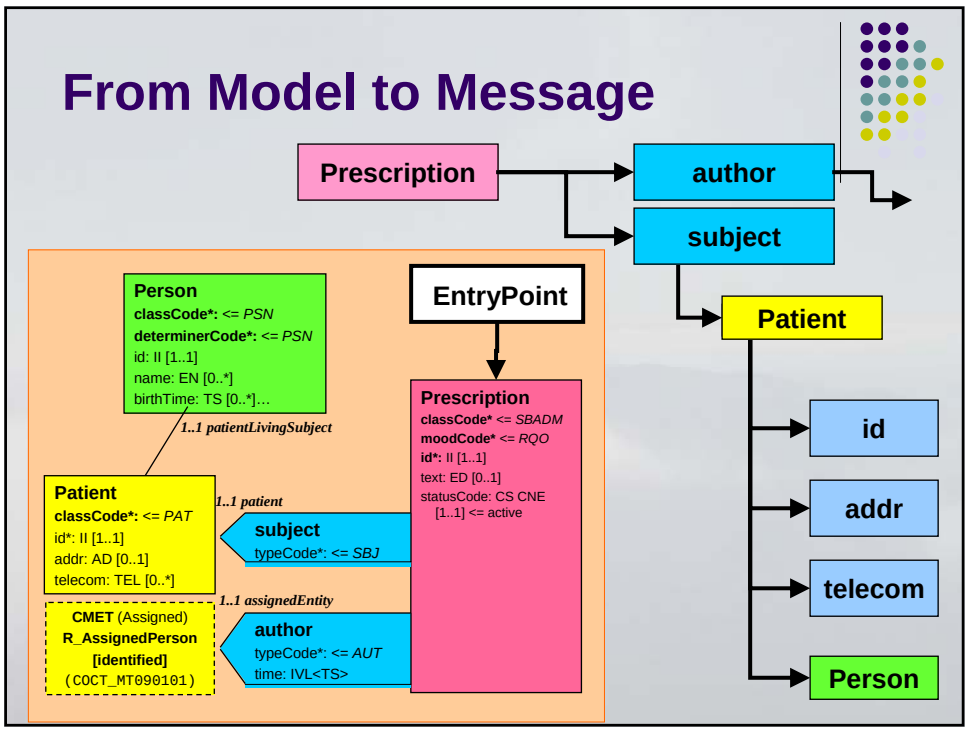


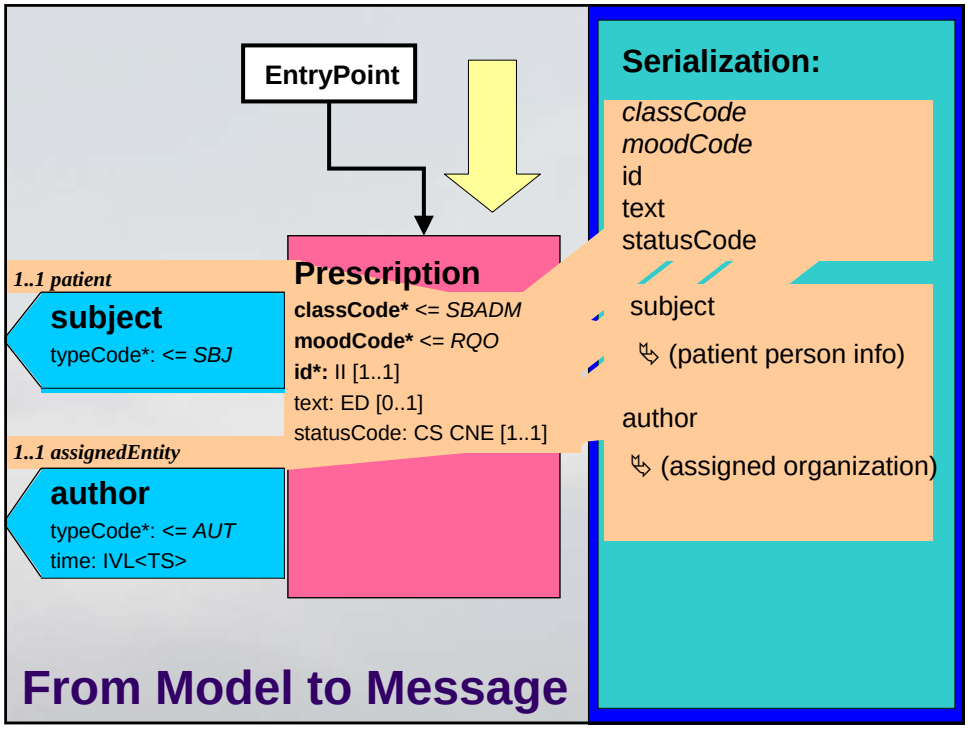
RIM ACT Class — **PARTICIPATION Class** — **ROLE Class** — **ENTITY Class**

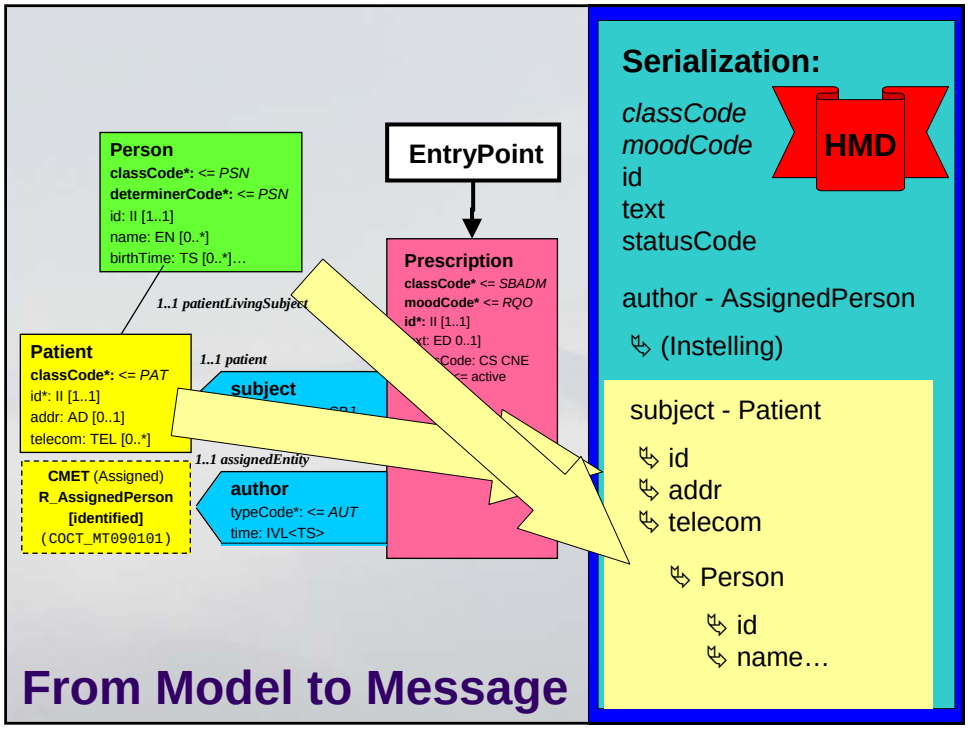
From Model to Message Reality



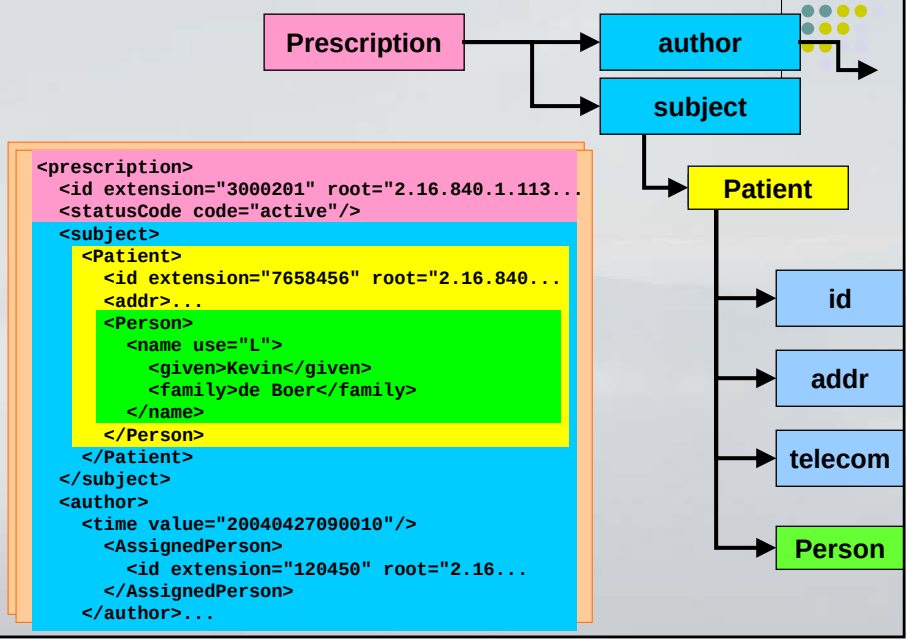
From Model to Message




EntryPoint
Prescription
classCode* <= SBADM
moodCode* <= RQO
id*: II [1..1]
text: ED [0..1]
statusCode: CS CNE [1..1]
1..1 patient
subject
typeCode*: <= SBJ
1..1 assignedEntity
author
typeCode*: <= AUT
time: IVL<TS>
Serialization:
classCode
moodCode
id
text
statusCode
subject
(patient person info)
author
(assigned organization)
From Model to Message


Person
classCode*: <= PSN
determinerCode*: <= PSN
id: II [1..1]
name: EN [0..*]
birthTime: TS [0..*]…
EntryPoint
Prescription
classCode* <= SBADM
moodCode* <= RQO
id*: II [1..1]
1..1 patientLivingSubject
Patient
classCode*: <= PAT
id*: II [1..1]
addr: AD [0..1]
telecom: TEL [0..*]
1..1 patient
subject
1..1 assignedEntity
CMET (Assigned)
R_AssignedPerson
[identified]
(COCT_MT090101)
author
typeCode*: <= AUT
time: IVL<TS>
Serialization:
classCode
moodCode
id
text
statusCode
HMD
author - AssignedPerson
(Instelling)
subject - Patient
id
addr
telecom
Person
id
name…
From Model to Message

From Model to Message



Where HL7 and XML Meet

- HL7 RIM defines the fields.
- Standard terminologies define the field values.

```
<Section templateId="10.233.64.98">
  <code code="1150x-y" codeSystem="LOINC">Plan</code>
  <Procedure moodCode="INT">
    <code code="P2-25010" codeSystem="SNOMED"
      displayName="Pulmonary function test"/>
  </Procedure>
  <Procedure moodCode="INT">
    <code code="P0-A022D" codeSystem="SNOMED"
      displayName="Teaching of skills">
      <qualifier>
        <name code="G-C0E7" displayName="has focus"/>
        <value code="P2-25214"
          displayName="Peak flow rate measurement"/>
      </qualifier>
    </code>
  </Procedure>
</Section>
```

Where HL7 and XML Meet

- HL7 RIM defines the fields.
- Standard terminologies define the field values.

```
<Section templateId="10.233.64.98">
  <code code="1150x-y" codeSystem="LOINC">Plan</code>
  <Procedure moodCode="INT">
    <code code="P2-25010" codeSystem="SNOMED"
      displayName="Pulmonary function test"/>
  </Procedure>
  <Procedure moodCode="INT">
    <code code="P0-A022D" codeSystem="SNOMED"
      displayName="Teaching of skills">
      <qualifier>
        <name code="G-C0E7" displayName="has focus"/>
        <value code="P2-25214"
          displayName="Peak flow rate measurement"/>
      </qualifier>
    </code>
  </Procedure>
</Section>
```

What About HL7 V2.x?

- HL7 V2.x does NOT have a reference information model behind them
- But each message type has an abstract message syntax that fully defines the message segments that make up the message type

HL7 V2.x Abstract Message Syntax

ORM^O01^ORM_O01	General Order Message	Chapter
MSH	Message Header	2
[{NTE}]	Notes and Comments (for Header)	2
[		
PID	Patient Identification	3
[PD1]	Additional Demographics	3
[{NTE}]	Notes and Comments (for Patient ID)	2
[		
PV1	Patient Visit	3
[PV2]]	Patient Visit- Additional Info	3
[{IN1	Insurance	6
[IN2]	Insurance Additional Info	6
[IN3]	Insurance Add'l Info - Cert.	6
}]		
[GT1]	Guarantor	6
[{AL1}]	Allergy Information	3
]		
{		
ORC	Common Order	4
[		
<OBR|	Order Detail Segment OBR, etc.	4
RQD|	(requisition details segment – for ordering medical, surgical & patient care supply)	
RQ1|	(paired with preceding RQD. Used for additional non-stock requisitions)	
RXO|	(pharmacy treatment order segment)	
ODS|	(for basic diet supply orders)	
ODT>	(diet tray specification, e.g. early tray, late tray, guest tray, etc)	
[{NTE}]	Notes and Comments (for Detail)	2
[CTD]	Contact Data	11
[{DG1}]	Diagnosis	6
[{		
OBX	Observation/Result	7
[{NTE}]	Notes and Comments (for Results)	2
}]		
]		
[{FT1}]	Financial Transaction	6
[{CTI}]	Clinical Trial Identification	7
[BLG]	Billing Segment	4
}		

HL7 V2.x PID message segment Spec

SEQ	LEN	DT	R/O	RP	TBL#	ITEM#	ELEMENT NAME
1	4	SI	O			00104	Set ID – Patient ID
2	20	CX	O			00105	Patient ID (External ID)
3	20	CX	R	Y		00106	Patient ID (Internal ID)
4	20	CX	O	Y		00107	Alternate Patient ID
5	48	XPN	R	Y		00108	Patient Name
6	48	XPN	O			00109	Mother's Maiden Name
7	26	TS	O			00110	Date of Birth
8	1	IS	O	Y	0001	00111	Sex
9	48	XPN	O	Y		00112	Patient Alias
10	1	IS	O		0005	00113	Race
11	106	XAD	O	Y		00114	Patient Address
12	4	IS	R			00115	Country Code
13	40	XTN	O	Y		00116	Phone number – home
14	40	XTN	O	Y		00117	Phone – business
15	60	CE	O		0296	00118	Language – Patient
16	1	IS	O		0002	00119	Marital Status

PID = Patient Identification

HL7 V2.x Segment Table

- The exact description of a segment and its attributes are defined in the segment table
- Structure of segment table:

SEQ	LEN	DT	R/O	RP	TBL#	ITEM#	ELEMENT NAME
1	4	SI	O			00104	Set ID – Patient ID
2	20	CX	O			00105	Patient ID (External ID)
3	20	CX	R	Y		00106	Patient ID (Internal ID)
…	…	…	…	…	…	…	…

Legends for the table headings:	
SEQ	Data component position (sequence in the segment)
LEN	Maximum length of data component
DT	Data type
R/O	Optionality (R = required; O = optional)
RP	Repetition
TBL#	HL7 table from which value of the data field is obtained
ITEM#	Number of the item

HL7 V2.x Data Types

- Each field, component, and subcomponent within a HL7 V2.x message segment is constrained by its data type definition
- The data type of an element governs the following:
  - The format of information in the data element
  - The number of sub-elements that the element may contain, if any
  - Whether the element's content may be restricted to a controlled vocabulary

Sample HL7 V2.x Data Types

Figure 2-2. HL7 data types by category

Data Type Category/ Data type	Data Type Name	LEN	HL7 Section Reference	Notes/Format
Alphanumeric				
ST	String	199	2.9.43	
TX	Text data	65536	2.9.48	
FT	Formatted text	65536	2.9.20	
SRT	Sort order		2.9.42	<sort-by field/parameter (varies)> ^ <sequencing (ID)>

ST: used for elements that contain relatively short, left justified text strings
TX: used for elements that will contain text to be displayed to the user. It can also be used to accommodate long text strings
FT: used to allow special formatting commands (e.g. line break, indent; centre) to be embedded in long text strings

Sample HL7 V2.x Data Types

Figure 2-2. HL7 data types by category

Data Type Category/ Data type	Data Type Name	LEN	HL7 Section Reference	Notes/Format
Numerical				
CQ	Composite quantity with units		2.9.10	<quantity (NM)> ^ <units (CE)>
MO	Money		2.9.26	<quantity (NM)> ^ <denomination (ID)>
NM	Numeric		2.9.28	
SI	Sequence ID		2.9.40	
SN	Structured numeric		2.9.41	<comparator (ST)> ^ <num1 (NM)> ^ <separator/suffix (ST)> ^ <num2 (NM)>
Date/Time				
DT	Date		2.9.15	YYYY[MM[DD]]
TM	Time		2.9.44	HH[MM[SS[.S[S[S[S]]]]]][+/-ZZZZ]
TS	Time stamp		2.9.47	YYYY[MM[DD[HHMM[SS[.S[S[S[S]]]]]]][+/-ZZZZ] ^ <degree of precision>

Sample HL7 V2.x Data Types

Data Type Category/ Data type	Data Type Name	LEN	HL7 Section Reference	Notes/Format
Code Values				
CE	Coded element	250	2.9.3	<identifier (ST)> ^ <text (ST)> ^ <name of coding system (IS)> ^ <alternate identifier (ST)> ^ <alternate text (ST)> ^ <name of alternate coding system (IS)>
CNE	Coded with no exceptions	250	2.9.8	<identifier (ST)> ^ <text (ST)> ^ <name of coding system (IS)> ^ <alternate identifier (ST)> ^ <alternate text (ST)> ^ <name of alternate coding system (IS)> ^ <coding system version ID (ST)> ^ alternate coding system version ID (ST)> ^ <original text (ST) >
CWE	Coded with exceptions	250	2.9.11	<identifier (ST)> ^ <text (ST)> ^ <name of coding system (IS)> ^ <alternate identifier (ST)> ^ <alternate text (ST)> ^ <name of alternate coding system (IS)> ^ <coding system version ID (ST)> ^ alternate coding system version ID (ST)> ^ <original text (ST) >
CF	Coded element with formatted values		2.9.4	<identifier (ID)> ^ <formatted text (FT)> ^ <name of coding system (IS)> ^ <alternate identifier (ID)> ^ <alternate formatted text (FT)> ^ <name of alternate coding system (IS)>

Sample HL7 V2.x Data Types

XPN	Extended person name	250	2.9.54	In Version 2.3, replaces the PN data type. <family name (FN)> ^ <given name (ST)> ^ <second and further given names or initials thereof (ST)> ^ <suffix (e.g., JR or III) (ST)> ^ <prefix (e.g., DR) (ST)> ^ <degree (e.g., MD) (IS)> ^ <name type code (ID) > ^ <name representation code (ID)> ^ <name context (CE)> ^ <name validity range (DR)> ^ <name assembly order (ID)>
XON	Extended composite name and ID number for organizations	250	2.9.53	<organization name (ST)> ^ <organization name type code (IS)> ^ <ID number (NM)> ^ <check digit (NM)> ^ <code identifying the check digit scheme employed (ID)> ^ <assigning authority (HD)> ^ <identifier type code (IS)> ^ <assigning facility ID (HD)> ^ <name representation code (ID)>
XTN	Extended telecommunications number	250	2.9.55	In Version 2.3 and later, replaces the TN data type. [NNN] [(999)]999-9999 [X99999] [B99999] [C any text] ^ <telecommunication use code (ID)> ^ <telecommunication equipment type (ID)> ^ <email address (ST)> ^ <country code (NM)> ^ <area/city code (NM)> ^ <phone number (NM)> ^ <extension (NM)> ^ <any text (ST)>

V2.x ADT Sample Message

```
MSH|^~\&|REGADT|MCM|IFENG||199112311501||ADT^A04^ADT_A01|000001|P|2.4|||
EVN|A04|199901101500|199901101400|01||199901101410
PID|||191919^^^GENHOS^MR~371-66-9256^^^USSSA^SS
  |253763|MASSIE^JAMES^A||19560129|M|||171 ZOBERLEIN^^ISHPEMING^MI^49849^""^|
  |(900)485-5344|(900)485-5344||S|C|10199925^^^GENHOS^AN|371-66-9256||
NK1|1|MASSIE^ELLEN|SPOUSE|171 ZOBERLEIN^^ISHPEMING^MI^49849^""^|(900)485-5344
  |(900)545-1234~(900)545-1200|EC1^FIRST EMERGENCY CONTACT
NK1|2|MASSIE^MARYLOU|MOTHER|300 ZOBERLEIN^^ISHPEMING^MI^49849^""^|(900)485-5344
  |(900)545-1234~(900)545-1200|EC2^SECOND EMERGENCY CONTACT
NK1|3
NK1|4|||123 INDUSTRY WAY^^ISHPEMING^MI^49849^""^||(900)545-1200
  |EM^EMPLOYER|19940605||PROGRAMMER|||ACME SOFTWARE COMPANY
PV1||O|O/R||||0148^ADDISON,JAMES|0148^ADDISON,JAMES|0148^ADDISON,JAMES|AMB||||||
  |0148^ADDISON,JAMES|S|1400|A|||||||||||||||||GENHOS|||||199501101410|
PV2|||||||199901101400|||||||||||||||||||||199901101400
OBX||ST|1010.1^BODY WEIGHT||62|kg||||F
OBX||ST|1010.1^HEIGHT||190|cm||||F
DG1|1|19||BIOPSY||00|
GT1|1||MASSIE^JAMES^""^""^""^""^||171 ZOBERLEIN^^ISHPEMING^MI^49849^""^
  |(900)485-5344|(900)485-5344||||SE^SELF|371-66-925||||MOOSES AUTO   CLINIC
  |171 ZOBERLEIN^^ISHPEMING^MI^49849^""|(900)485-5344|
IN1|0|0|BC1|BLUE CROSS|171 ZOBERLEIN^^ISHPEMING^M149849^""^|
  |(900)485-5344|90||||||50 OK|
```

V2.x OBX Segment Syntax

HL7 Attribute Table – OBX – Observation/Result

SEQ	LEN	DT	OPT	RP/#	TBL#	ITEM#	ELEMENT NAME
1	4	SI	O			00569	Set ID - OBX
2	2	ID	C		0125	00570	Value Type
3	250	CE	R			00571	Observation Identifier
4	20	ST	C			00572	Observation Sub-ID
5	65536[1]	*	C	Y[2]		00573	Observation Value
6	250	CE	O			00574	Units
7	60	ST	O			00575	References Range
8	5	IS	O	Y/5	0078	00576	Abnormal Flags
9	5	NM	O			00577	Probability
10	2	ID	O	Y	0080	00578	Nature of Abnormal Test
11	1	ID	R		0085	00579	Observation Result Status
12	26	TS	O			00580	Date Last Observation Normal Value
13	20	ST	O			00581	User Defined Access Checks
14	26	TS	O			00582	Date/Time of the Observation
15	250	CE	O			00583	Producer's ID
16	250	XCN	O	Y		00584	Responsible Observer
17	250	CE	O	Y		00936	Observation Method
18	22	EI	O	Y		01479	Equipment Instance Identifier
19	26	TS	O			01480	Date/Time of the Analysis

```
OBX||ST|1010.1^BODY WEIGHT||62|kg||||F
OBX||ST|1010.1^HEIGHT||190|cm||||F
```

V2.x OBX Segment Syntax

```
OBX||ST|1010.1^BODY WEIGHT||62|kg|||||F
OBX||ST|1010.1^HEIGHT||190|cm|||||F
<OBX>
      <OBX.2>ST</OBX.2>
      <OBX.3>
          <CE.1>1010.1</CE.1>
          <CE.2>BODY WEIGHT</CE.2>
      </OBX.3>
      <OBX.5>62</OBX.5>
      <OBX.6>
          <CE.1>kg</CE.1>
      </OBX.6>
      <OBX.11>F</OBX.11>
</OBX>
```

V2.x OBX Segment Syntax

```
OBX||ST|1010.1^BODY WEIGHT||62|kg|||||F
OBX||ST|1010.1^HEIGHT||190|cm|||||F
<OBX>
      <OBX.2>ST</OBX.2>
      <OBX.3>
          <CE.1>1010.1</CE.1>
          <CE.2>HEIGHT</CE.2>
      </OBX.3>
      <OBX.5>190</OBX.5>
      <OBX.6>
          <CE.1>cm</CE.1>
      </OBX.6>
      <OBX.11>F</OBX.11>
</OBX>
```

Conclusion

- XML is the technology / language framework for serialising the data stream before it is transmitted from one system to another
- Health messaging is NOT just about squirting serialised XML data stream down the pipe
- It is about ensuring interoperability at
  - Functional and
  - Semantic levels

Conclusion

- interoperability
  - Ability of 2 or more systems or components to exchange information and to use the information that has been exchanged.

Functional interoperability

Semantic interoperability

- HL7 provides the rules for
  - encoding the data stream (functional interoperability)
  - Semantic basis for interoperability of information across all applications and all users (semantic interoperability)

The Risk of XML without HL7:

A consequence we can ill afford

Thank You