CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM

The CDISC vision is to inform patient care & safety through higher quality medical research.

Strength *through Collaboration*

The CDISC Moonshot - Can we do it?

Presented by FH-Prof. Dr. Jozef Aerts
University of Applied Sciences FH Joanneum
Graz, Austria

FH | JOANNEUM
University of Applied Sciences

XML 4Pharma

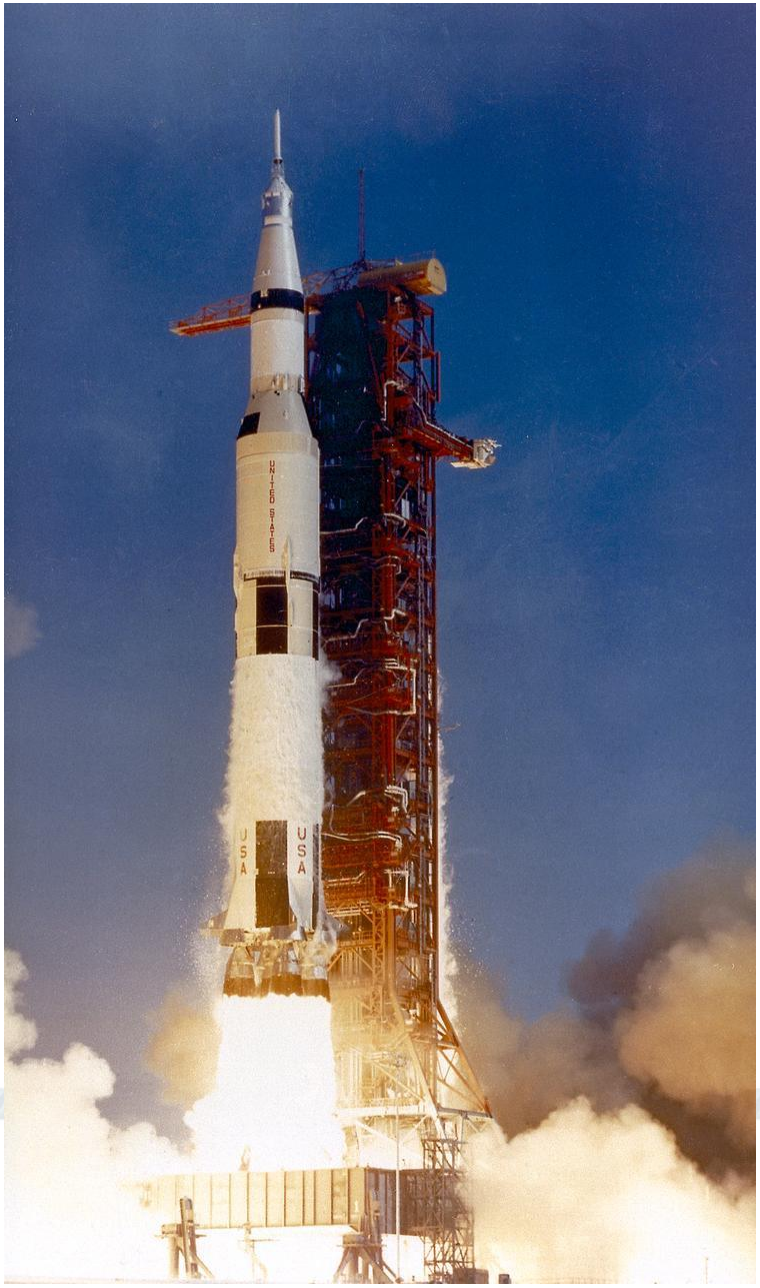
Strength *through Collaboration*

CDISC has grown …

But it is time to take the next step …

The CDISC Moonshot

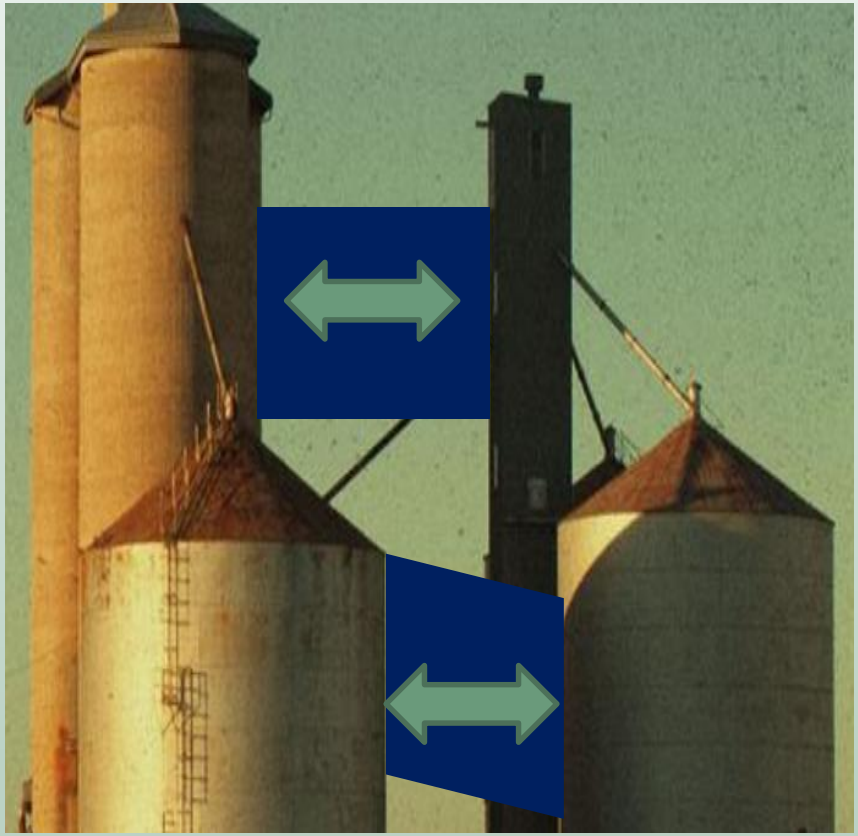
The next great step …



From the CDISC website …

Are these stages of a CDISC rocket?

Oh no … these are CDISC silos …



From silos to rocket stages

- Most of our standards have their origin in the last century …
- Are not according to "current state of the art" in (bio)medical informatics
- Have developed along the same lines and traditions we had almost 20 years ago
- Don't we need a modernization?

The CDISC ODM Standard

- 20 years old, only minor updates
  - And some extensions (Define-XML, SDM-XML, Dataset-XML, CTR-XML)
- No support for:
  - RESTful web services
  - Flexible and complicated study designs
  - Wearables, insideables
- Little support for electronic health records
- Based on the concept of "files"
  - Who still needs data files? We need "data services"

ODM 2.0 - the first rocket stage

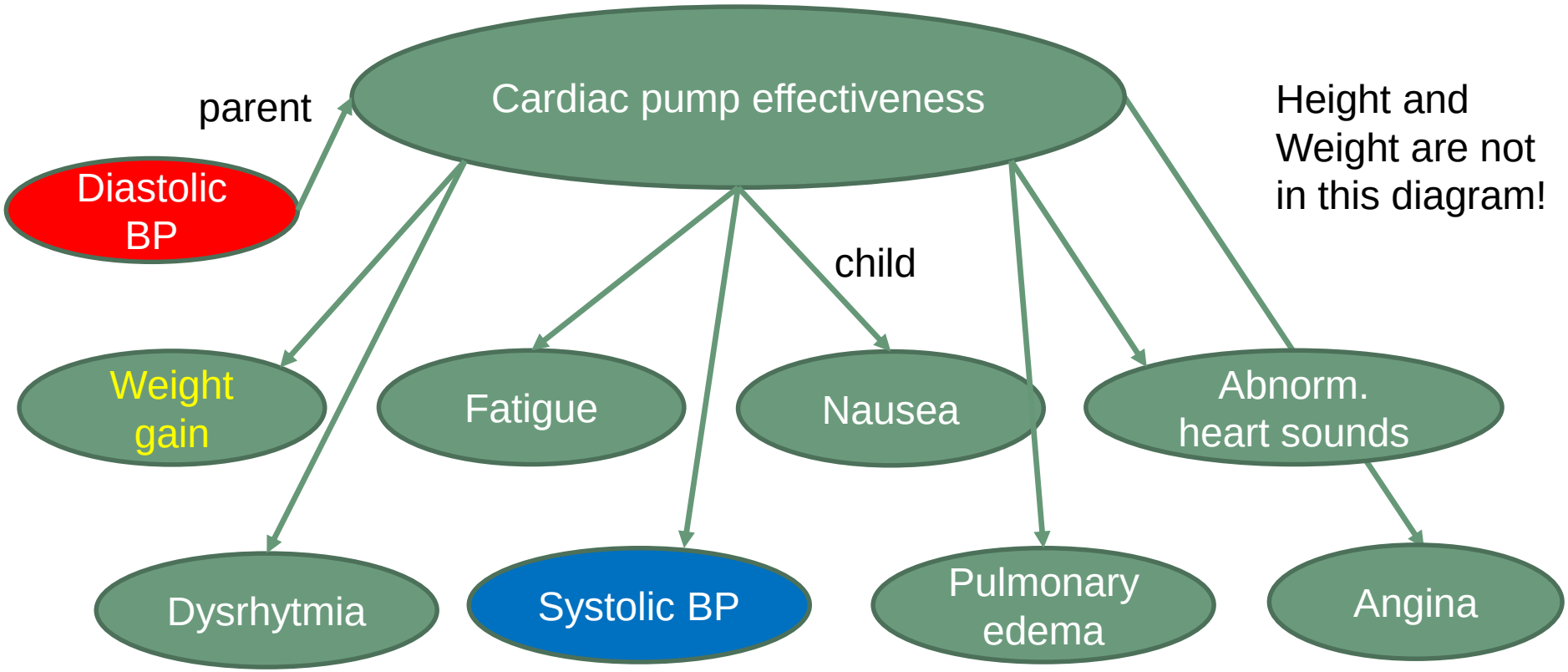
- Support for RESTful web services
  - The data can be anyware - not necessarily in files
- Full integration of electronic health records
- We need some concepts that are also used by FHIR
  - But clinical research is more complicated (protocol based)
- Downward compatible
- Can ODM and FHIR be one standard in 10 years from now?

CDISC controlled terminology

- "Lists" of terms with little of no links between them
- DIABP and SYSBP are as much related as DIABP with FRMSIZE - really?
- Refusal to use controlled terminology from the medical world
  - LOINC
  - UCUM
  - ICDx
  - SNOMED
  - …

An example using UMLS RESTful web services

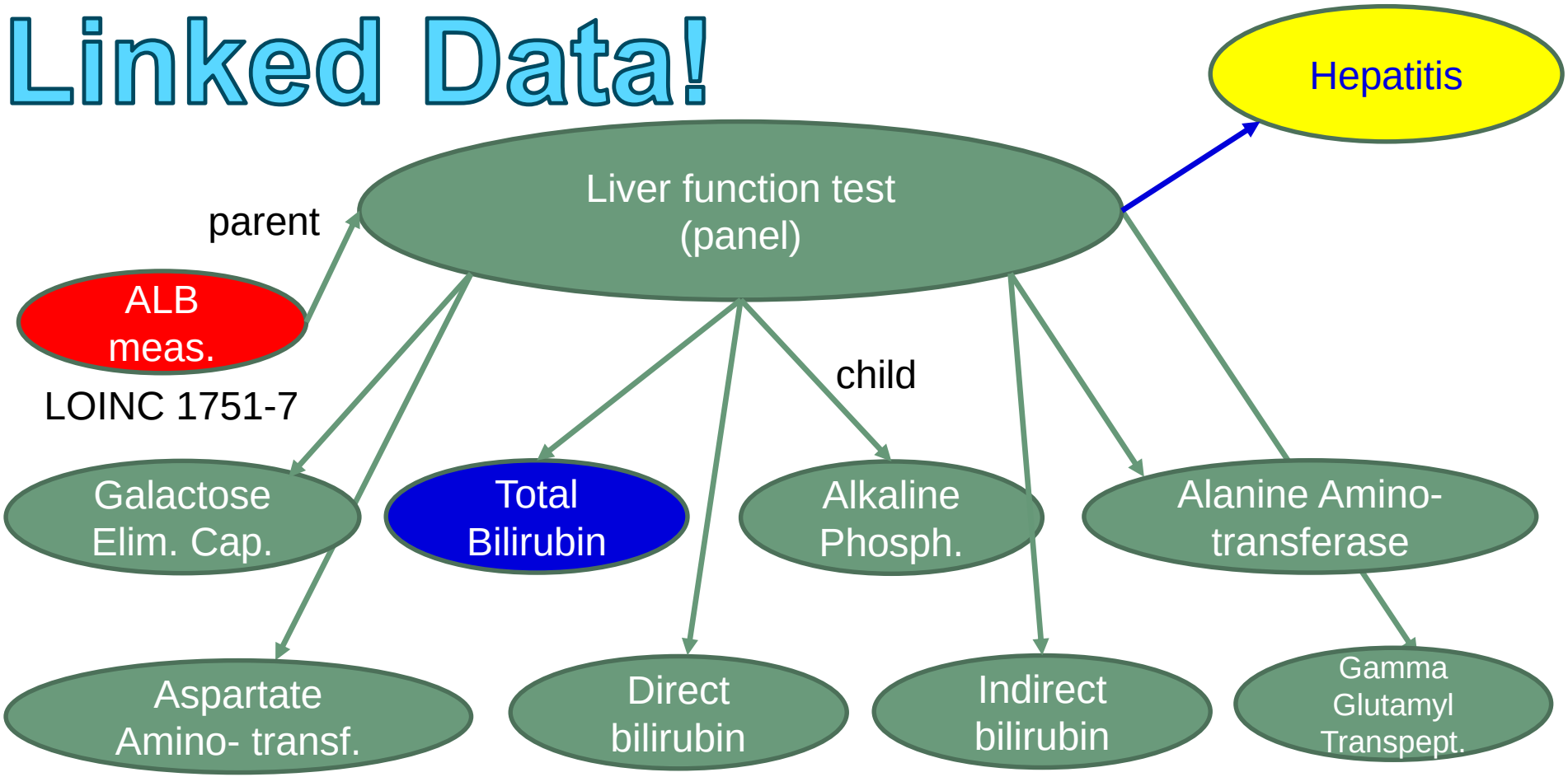
Linked Data!



"DIABP" and "SYSBP" are related through e.g. "*cardiac pump effectiveness*"
There is no way ever to find out using CDISC-CT alone…

A LOINC example using UMLS RESTful web services

Linked Data!



"ALB" and "BILI" are related through e.g. "Liver function test"
There is no way ever to find out using CDISC-CT alone…

Units

- How much cm is 1 inch?
  - Does CDISC-CT know?

Answer:

http://www.xml4pharmaserver.com:8080/CDISCCTService/rest/ucumtransform/[in_i]/to/cm

UCUM Transformation Web Services

```
- <XML4PharmaServerWebServiceResponse ServerDateTime="2017-02-08T16:13:28">
  - <WebServiceRequest>
      http://www.xml4pharmaserver.com:8080/CDISCCTService/rest/ucumtransform/1/from/%5Bin_i%5D/to/cm
    </WebServiceRequest>
  - <Response>
      <SourceQuantity>1.0</SourceQuantity>
      <SourceUnit>[in_i]</SourceUnit>
      <TargetUnit>cm</TargetUnit>
      <ResultQuantity>2.54</ResultQuantity>
    </Response>
  </XML4PharmaServerWebServiceResponse>
```

Units

- How much is 1 "international unit" pro mg "supercompound" pro g pro kg of "chicken" pro hour in: "international unit" pro ounce "supercompound" pro g pro pound of "chicken" per day ?

```
− <XML4PharmaServerWebServiceResponse ServerDateTime="2017-02-08T16:22:02">
  − <WebServiceRequest>
      http://www.xml4pharmaserver.com:8080/CDISCCTService/rest/ucumtransform/1/from/%5BiU%
      %5D%7Bchicken%7D/d)
    </WebServiceRequest>
  − <Response>
      <SourceQuantity>1.0</SourceQuantity>
      <SourceUnit>[iU]/mg{supercompound}/(g/kg{chicken}/h)</SourceUnit>
    − <TargetUnit>
        m[iU]/[oz_av]{supercompound}/(g/[lb_av]{chicken}/d)
      </TargetUnit>
      <ResultQuantity>2604166.7</ResultQuantity>
    </Response>
  </XML4PharmaServerWebServiceResponse>
```

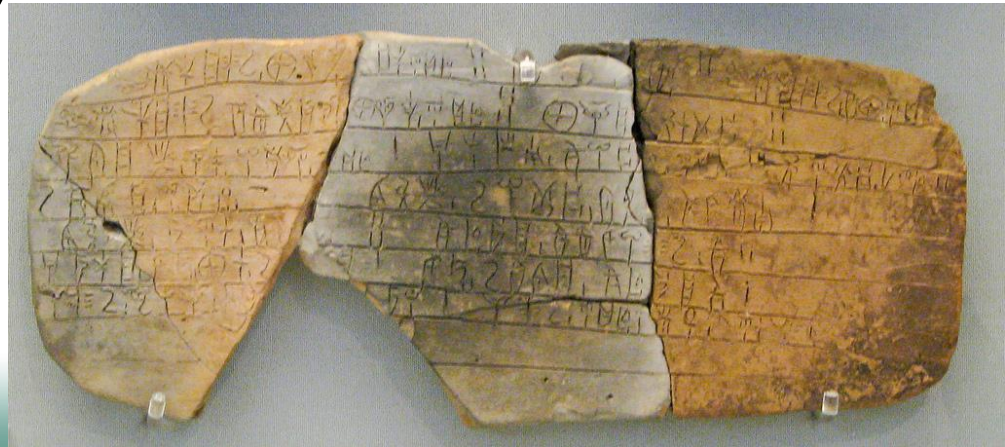
Forget about doing this when using CDISC "units"

CDISC Controlled Terminology

- CT is our "rocket fuel"
- And we can't afford not to get the highest quality fuel
- If there is better fuel than "our own invented one", we need to use that
- Some of our "fuel" is OK, but the most important part isn't

SDTM / SEND / ADaM

- Two dimensional tables
- Breaking almost any rule of "good database design"
- Massive amount of data redundancy
  - 30-40% of the variables could be removed
- Very limited amount of traceability (only through define.xml)
- Pretty bad (outdated) data model



Smart tools can replace data redundancy

DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC
DM	01-701-1015	1015	2014-01-02	2014-07-02
DM	01-701-1023	1023	2012-08-05	2012-09-0a
DM	01-701-1028	1028	2013-07-19	2014-01-14
DM	01-701-1033	1033	2014-03-18	2014-04-14
DM	01-701-1034	1034	2014-07-01	2014-12-30
DM	01-701-			
DM	01-701-			
DM	01-701-			
DM	01-701-1111	1111	2012-09-07	2012-09-17

01-701-1033 (USUBJID)
First date of study treatment exposure = Tue Mar 18 2014
Last date of study treatment exposure = Mon Mar 31 2014

COUNTRY	DMDTC
USA	2013-12-26
USA	2012-07-22
USA	2013-07-11
USA	2014-03-10
USA	
USA	

2013-07-11 (DMDTC)
DMDY = -8

VS

USUBJID	VSSEQ	VSTESTCD	VSPOS	VSO
01-701-1015	7	DIABP	SUPINE	56
01-701-1015	8	DIABP	STANDING	51
01-701-1015	9	DIABP	STANDING	61
01-701-1015	10	DIABP		
01-701-1015	11	DIABP		
01-701-1015	12	DIABP		

DIABP (VSTESTCD)
Diastolic Blood Pressure
NCI: C25299

VISITNUM	VSDTC	VSTPT
6	2014-02-01	AFTER LYIN
6	2014-02-01	AFTER STAN
6	2014-02-01	AFTER STAN
7		AFTER LYIN
7		AFTER STAN
7		AFTER STAN

6 (VISITNUM)
Visit: AMBUL ECG REMOVAL
Planned Study Day of Visit: 30

VISITNUM	LBDTC
9	2014-03-26T15:15
10	2014-05-07T11:21
11	2014-06-21T10:58

2014-05-07T11:21 (LBDTC)
LBDY = 125
Element = PBO (Placebo) - Epoch = Treatment

Reviewers should start using these tools…

Smart Dataset-XML Viewer

SDTM / SEND / ADaM

- Due to data redundancy and 2-D tables
  our submission standards have too much weight …
- Like that, we will never reach the moon …
- CDISC should never have given up first principles
  because of (silly) requests from reviewers who use completely outdated tools
- It is not too late to slim down SDTM/SEND/ADaM
- And we should start thinking about alternatives
  - Some people do, but their work is mostly ignored

Is SDTM ready for the future?

- More and more trials with "wearables" and "insideables"
  - "CRF free"
- Remote clinical trials without any "visits"
  - But SDTM still requires us to submit VISITNUM
- Clinical trials with virtual sites
- Read Doug Bain's and Kai Langel's blogs (ClinPal)

- Shouldn't we rethink a good number of SDTM (outdated) assumptions?

Better / smarter validation tools

- Current process:

External company

Validation Software

- Result:

Better / smarter validation tools

The "Open Rules for CDISC Standards" Initiative

FDA,PMDA,CDISC rules implemented as:

- Transparent rules
- Human-readable and machine-executable
- Use RESTful web services
  - In future also using the SHARE API
- Updates/corrections over a RESTful web service
  - No more "waiting for the next release"
- Anyone can develop his own validation system

"Open Rules for CDISC Standards"

Example - Rule FDAC017

```
(: Rule FDAC017-FDAC018: SDTM Required variable not found - Variables described in SDTM as Required
 must be included in the dataset :)
(: The following Query relies on that the define.xml is complete and
 that Mandatory='Yes' is set for each required variable :)
xquery version "3.0";
declare namespace def = "http://www.cdisc.org/ns/def/v2.0";
declare namespace odm="http://www.cdisc.org/ns/odm/v1.3";
declare namespace data="http://www.cdisc.org/ns/Dataset-XML/v1.0";
declare namespace xlink="http://www.w3.org/1999/xlink";
declare namespace request="http://exist-db.org/xquery/request";
(: "declare variable ... external" allows to pass $base and $define from an external programm :)
(: declare variable $base external; :)
(: declare variable $define external; :)
let $base := '/db/fda_submissions/cdisc01/'
let $define := 'define2-0-0-example-sdtm.xml'
(: iterate over all datasets mentioned in the define.xml :)
for $itemgroup in doc(concat($base,$define))//odm:ItemGroupDef
    (: get all the ItemRef-OIDs which have 'Mandatory="Yes" :)
    let $mandatory := $itemgroup/odm:ItemRef[@Mandatory='Yes']/@ItemOID
    (: get the dataset itself :)
    let $datasetfilename := $itemgroup/def:leaf/@xlink:href
    let $dataset := doc(concat($base,$datasetfilename))
    let $datasetname := $itemgroup/@Name
    (: iterate over all the records :)
    for $record in $dataset//odm:ItemGroupData
        let $recnum := $record/@data:ItemGroupDataSeq
         (: iterate over the 'mandatory' OIDs  :)
         for $m in $mandatory
             (: and give an error when there is no such ItemData/@ItemOID  :)
             let $varname := doc(concat($base,$define))//odm:ItemDef[@OID=$m]/@Name
             where not($record/odm:ItemData[@ItemOID=$m])
             return <error rule="FDAC017" datasetname="{$datasetname}" variable="{data($varname)}"
             rulelastupdate="2015-08-31" recordnumber="{$recnum}">
             No data found for required variable {data($varname)} in record number {data($recnum)}
             in dataset {data($datasetname)}</error>
```

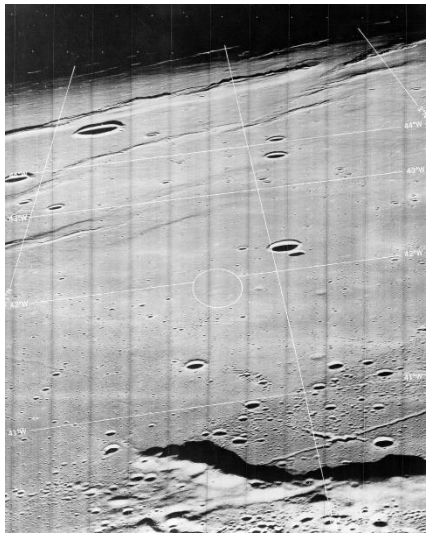
Already implemented in the "Smart Dataset-XML Viewer"

http://xml4pharmaserver.com/RulesXQuery/index.html

The final stage: landing on the moon

- On average, our spacecraft is circling 6 months (the review time) around the moon before landing

- The crew is told they are free to use any tools they brought with them,
  even when they are completely outdated …



How can review be shortened

without any loss of quality - even with quality increase

- Reviewers should not be allowed to use files anymore
- All submissions in a single (set of) database(s)
  - E.g. native XML databases
  - The database is the only allowed source of information
- CDISC volunteers can provide modern tools (e.g. open-source) to regulators that reviewers must at least use

STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LI
CDISCPILOT01	LB	01-701-1015	233	BASO	264
CDISCPILOT01	LB	01-701-1015	263	BASO	264
CDISCPILOT01	LB	01-701-1015	298	BASO	264
CDISCPILOT01	LB	01-701-1015	7	BILI	197
CDISCPILOT01	LB	01-701-1015	44	BILI	
CDISCPILOT01	LB	01-701			
CDISCPILOT01	LB	01-701			
CDISCPILOT01	LB	01-701			
CDISCPILOT01	LB	01-701			

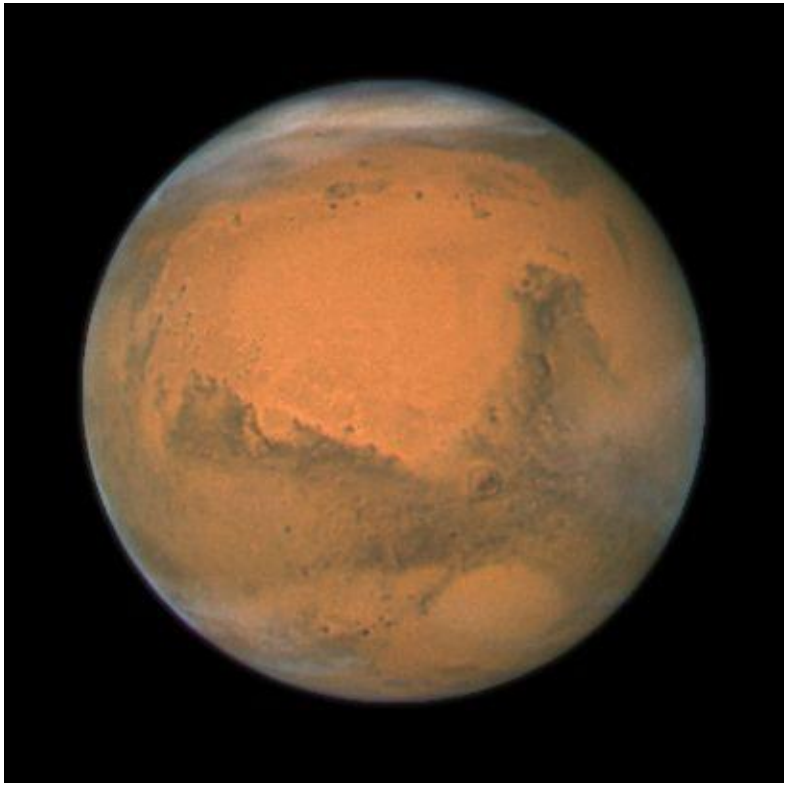
01-701-1015 (USUBJID)
ACTARMCD: Pbo
AGE: 61 years
SEX: Female

LBSEQ	LBTESTCD	LBLOINC	LBCAT	LBORRES	LBORRESU	VISITNUM	LBORNRLO
233	BASO	26444-0	HEMATOLOGY	0.03	THOU/uL	11	0
263	BASO	26444-0	HEMATOLOGY	0.04	THOU/uL	12	0
298	BASO	26444-0	HEMATOLOGY	0.03	THOU/uL	13	0
7	BILI	1975-2	CHEMISTRY	0.6	mg/dL	1	0.2
44	BILI	1975-2	CHEMISTRY	0.5	mg/dL	4	0.2
79	BILI	1975-2					
109	BILI	1975-2					
139	BILI	1975-2					

1975-2 (LBLOINC)
LOINC Common Name: Bilirubin.total [Mass/volume] in Serum or Plasma
Example UCUM Units: mg/dL

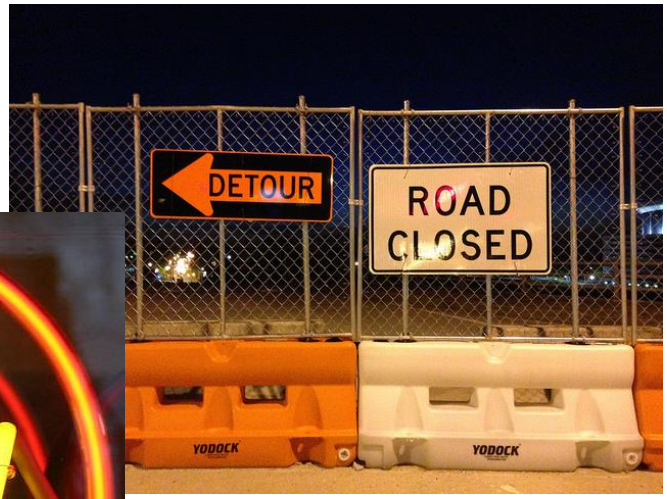
Some are even reaching further …

- Initiatives like:
  - RDF
  - Linked Data
  - SDTM-IG in XML
  - Protocol in XML
  - …
- Mostly not official CDISC activities
  (but shouldn't they?)
- But regulators are still extremely reluctant to even think about these



The Showstoppers … Actions required

- Get rid of SAS-XPT
- Start mandating LOINC and UCUM
- Get rid of "not invented here" syndrome
- Help / convince / force regulatory authorities to use modern technologies



Regulators and the "cancer moonshot"

- FDA is expected to receive US $xx million from the "cancer moonshot" programm
- If they would invest a few % of that in modern technology (computers, non-relational databases, tools) they would make a huge step forward
- And the result could be:

Acknowledgements

- All volunteers that fight
  - FOR getting things done better, simpler, more efficient
  - AGAINST "not invented here", "do it as we always have done" mentality

Photo credit: NASA/Kim Shiflett

Further Acknowledgements

- Anthony, Dave (2x), Frederic, Geoff, Kerstin, Lauren, Lex, Niels, Peter, Marcelina, Sally, Sam, Vojtech, …

- My eHealth students at the university of applied sciences FH Joanneum in Graz

Picture acknowledgements at:
www.XML4Pharma.com/London_Interchange/acknowledgements.html

The next step …?

Disclaimer

- None of the pictures in this presentation are owned by "PictureDesk", a Vienna-based "picture troll" company. PictureDesk has no rights to any of the pictures in this presentation.
- Almost all pictures come from either Flickr under a "Creative Commons License" or from NASA put in the public domain, or were reproduced with explicit permission.
- This presentation is in the public domain. You may use (parts of) it, except for legal action, or discredit the good name of CDISC, XML4Pharma, the University of Applied Science FH Joanneum, and/or Jozef Aerts
- By downloading this presentation, you agree to pay the sum of US$ 1,000,000 to XML4Pharma when using contents of this presentation for any legal action