Multi-modal data management: clinical, multi-omics, and imaging data

Clinical trial days - May 20\textsuperscript{th} 2019

Jan-Willem Boiten
Lygature – Utrecht & Health-RI, janwillem.boiten@lygature.org
Why Lygature?
Multilateral partnerships are challenging

‘The art of the (bilateral) deal’

But multilateral partnerships...
Lygature is the driving force behind >20 large partnerships in four main areas – examples

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Details</th>
<th>Funding</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data infrastructure</td>
<td>‘Office suite’ for translational research</td>
<td>EUR 19 m TTI transition, KWF</td>
<td>&gt;3,500 users &gt;500 (multicenter) studies</td>
</tr>
<tr>
<td>Strategic asset sharing</td>
<td>Pan-European collaborative drug discovery</td>
<td>EUR 196 m EU-IMI, private</td>
<td>&gt;80 public drug discovery programmes First partnering deals</td>
</tr>
<tr>
<td>Regulatory innovation</td>
<td>Neutral platform for regulatory innovation</td>
<td>EUR 1 m Private (unrestricted)</td>
<td>Regulators draft new guidelines for this special class of drugs</td>
</tr>
<tr>
<td>Neglected tropical diseases</td>
<td>Pediatric formulation for a Neglected Tropical Disease</td>
<td>EUR 38 m GHIT, EDCTP, private</td>
<td>Start of phase 3 trial, planning for access in endemic countries</td>
</tr>
</tbody>
</table>

For a total overview and more information, please visit https://www.lygature.org/projects
Data fuels science for society – but what about the health domain?
Underlying problem in health domain: fragmentation

Knowledge institutes
- UMC Utrecht
- LUMC
- Radboud MC
- Erasmus MC
- VUmc
- AMC
- UMCG
- MUMC
- UU
- LU
- Radboud
- 3TU
- TNO
- NKI
- RIVM
- Hubrecht Institute
- Sanquin
- CBS
- private institutes

Knowledge institutes: VSNU Beta

International community
- Many different parties

Ministries
- OC&W
- EZ
- VWS

Funding other than ESFRI
- ZonMW
- NWO
- Various charities
- Top-sectors
- eScience Center
- Horizon 2020

Patient organisations
- Various patient organisations

Sector/cross-sector platforms
- BMS.NL
- DTL
- Health-Holland

Standardisation institutes
- NEN
- NICTIZ

Partners in PPP
- CRO’s
- Pharma
- MKB / Spin off

Hospitals
- Diverse andere ziekenhuizen

Northwestern Foundation
- Clinics
- Partnerships
- Projects

NFU projects
- PSI/Life Lines
- Registratie aan de Bron
- Platform Clinical Research

Suppliers
- EPD’s suppliers
- SURF/VANCIS
- IT companies (IBM, Oracle, SAS etc.)
- Largescale devices (Philips e.a.)

Topinstitutes
- NGI/INRIC
- CTMM / TraIT
- TI Pharma / Mondriaan
- BMM

NL Roadmap largescale research infrastructure/ ESFRI
- EATRIS-NL/EATRIS
- NeCEN/INSTRUCT
- BBMRI-NL / BBMRI
- DTL/ELIXIR-NL/ELIXIR

IT companies
- IBM, Oracle, SAS etc.

Largescale devices
- (Philips e.a.)

Underlying problem in health domain: fragmentation

27/05/2019

Slide: NFU Data4lifesciences
Health-RI organizes data services and expertise, and bundles forces from existing infrastructures.
Example of Health-RI services: Support for ethical, legal and societal issues: the ELSI servicedesk

Frequently asked questions

Possibility to ask questions to expert team

Link to relevant resources

elsi.health-ri.nl
Once upon a time in The Netherlands.....

The Center for Translational Molecular Medicine (CTMM)

From Symptomatic to Preventive Medicine

Organ level
Detection, Tumor burden

Tissue level
Angiogenesis, Growth kinetics, Drug delivery

Cellular level
Tumor markers, Drug targeting

Molecular level
Genetic mutations, Biomarker testing, Gene therapy

Budget
300 M€ Government (50%), Public (25%), Private (25%)
Common IT issues found across translational projects in the CTMM program

Across the program there was a lack of.....

• Collaboration – only local solutions available
• Culture of data sharing
• Multi-modal solutions – data integration
• Sustainable and reliable services
• Integrated solutions supporting the workflow of the scientists
• Etc

That's why CTMM started the TraIT project

CTMM: multi-modal research programs
What’s actually needed is a standardized ecosystem
"App store" for multimodal translational research - The TraIT suite of tools

Clinical data research
- OpenClinica
- L.

Clinical imaging
- XNAT
- KEOSYS
- RSNA

Digital pathology
- EPIS

Biosample data
- workflow
- molgenis catalogue

Experimental data
- Phenotype database
- Galaxy

Data integration and analysis
- cBioPortal

Deployment and user support
- ServiceDesk
### Some examples of TraIT services

<table>
<thead>
<tr>
<th>Clinical data research</th>
<th>Clinical imaging</th>
<th>Digital pathology</th>
<th>Biosample data</th>
<th>Experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="OpenClinica" /></td>
<td><img src="image" alt="XNAT" /></td>
<td><img src="image" alt="EPIS" /></td>
<td><img src="image" alt="BBMRI - PODIUM" /></td>
<td><img src="image" alt="Phenotype database" /></td>
</tr>
<tr>
<td><img src="image" alt="keosys" /></td>
<td><img src="image" alt="rsna" /></td>
<td><img src="image" alt="MOvES" /></td>
<td><img src="image" alt="molgenis catalogue" /></td>
<td><img src="image" alt="Galaxy" /></td>
</tr>
<tr>
<td><img src="image" alt="100 years" /></td>
<td><img src="image" alt="100 years" /></td>
<td><img src="image" alt="HOVON HOP" /></td>
<td><img src="image" alt="OPENSPCEMEN" /></td>
<td><img src="image" alt="Galaxy" /></td>
</tr>
<tr>
<td><img src="image" alt="100 years" /></td>
<td><img src="image" alt="100 years" /></td>
<td><img src="image" alt="HOVON HOP" /></td>
<td><img src="image" alt="OPENSPCEMEN" /></td>
<td><img src="image" alt="Galaxy" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><img src="image" alt="BBMRI - PODIUM" /></td>
<td><img src="image" alt="Phenotype database" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><img src="image" alt="molgenis catalogue" /></td>
<td><img src="image" alt="Galaxy" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><img src="image" alt="OPENSPCEMEN" /></td>
<td><img src="image" alt="Galaxy" /></td>
</tr>
</tbody>
</table>

#### Data integration and analysis
- ![cBioPortal](image)

#### Deployment and user support
- ![ELSI](image)
- ![ServiceDesk](image)
- ![PIA tool](image)
OpenClinica - web-based clinical data capture

Login to TraIT OpenClinica webserver

Enter data in eCRF

Export study data in desired format

Build your study/eCRF

TraIT OpenClinica is easy to access via your internet browser

It is easy to build a simple CRF and with some learning you can build sophisticated functions
Multi-model data management requires interoperable tools – OpenClinica integrations as an example

- **Ldot**: Clinical trial logistics
- **XNAT**: Image archive
- **tEPIS**: Pathology image sharing
- **tranSMART**: Clinical-genomics data integration

Open interfaces – plug-and-play: No lock-in with all tools!
Medical imaging data in multimodal studies

Clinical data research
- OpenClinica
- eXtensible Neuroimaging Archives (XNAT)
- KEOSYS Medical Imaging

Clinical imaging
- NBIA

Digital pathology
- eEPIS

Biosample data
- BBMRI - PODIUM
- molgenis catalogue
- HOVON HOP
- OPENSECIMEN

Experimental data
- Phenotype database
- Galaxy

Data integration and analysis
- cBioPortal

Deployment and user support
- ELSI
- ServiceDesk

PIA tool

27/05/2019

trait.health-ri.nl
How to store and share medical imaging data for research?
Image anonymization pipeline with CTP

- DICOM anonymization
- Configurable per study, per site, per instrument

XNAT – heart of the Health-RI/TraIT imaging platform

XNAT.BMIA.NL currently contains 83 projects, 9689 subjects, and 29672 imaging sessions.
Health-RI/TraIT multimodal data services – pivotal role for tranSMART

<table>
<thead>
<tr>
<th>Clinical research data</th>
<th>Clinical imaging</th>
<th>Digital pathology</th>
<th>Biosample data</th>
<th>Experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenClinica</td>
<td>NBIA</td>
<td>IBIS</td>
<td>workflow</td>
<td>Phenotype database</td>
</tr>
<tr>
<td>ImageHub</td>
<td>XNAT</td>
<td>KEOSYS</td>
<td>molgenis</td>
<td>Galaxy</td>
</tr>
<tr>
<td></td>
<td>KEOSYS Medical Imaging</td>
<td></td>
<td>catalogue</td>
<td>Chipster</td>
</tr>
<tr>
<td></td>
<td>RSNA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data integration and analysis

- cBioPortal

Deployment and user support

- ServiceDesk

 tranSMART
tranSMART: hypothesis-free browsing of clinical & genomics data
Simple subset selection example:
For which patients do we have mutation data?
Simple subset selection example:
For which patients do we have DNA samples?

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sex</th>
<th>Age</th>
<th>2_Catalogue_link__Subject_a</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeCoDe_WP6_CRC_LM_0412</td>
<td>male</td>
<td>53</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0156</td>
<td>male</td>
<td>69</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0157</td>
<td>male</td>
<td>75</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0158</td>
<td>female</td>
<td>47</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0159</td>
<td>female</td>
<td>54</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0160</td>
<td>female</td>
<td>61</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0161</td>
<td>male</td>
<td>68</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0162</td>
<td>male</td>
<td>77</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0163</td>
<td>male</td>
<td>56</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0164</td>
<td>female</td>
<td>70</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0165</td>
<td>female</td>
<td>69</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0166</td>
<td>male</td>
<td>61</td>
<td>[Direct hyperlink]</td>
</tr>
<tr>
<td>DeCoDe_WP6_CRC_LM_0167</td>
<td>male</td>
<td>62</td>
<td>[Direct hyperlink]</td>
</tr>
</tbody>
</table>

Integration between open source tools:
Direct hyperlink to corresponding sample in MOLGENIS sample catalogue
On-the-fly analyses within tranSMART

Exploring and comparing subgroups in your data set
Can I reproduce the results of my own PhD students if I have their processed data?

How FAIR is my data?
With tranSMART you can explore the data and reproduce the figures from the paper.

The data is in! We can:
- Reproduce published results
- Perform additional analyses
Link to cBioPortal for more gene-centric data integration

Example study with more omics data types

Uptake of TraIT solutions – 4000 users from >400 projects

> 3750 users
> 625 locations worldwide

**Supported studies** 553

- Production Studies 392
- Oncology 42%
- Cardiovascular 12%
- Internal Medicine 11%
- Rheumatology 8%
- Neurology 8%
- Surgery 4%
- Other 15%

**Studies in preparation** 161
Professional support is a key success factor when setting up networked IT infrastructure: Health-RI Service Desk

1st line support (Service Desk)

2nd line support Service A
2nd line support Service B
2nd line support Service C

Assignment of calls by knowledgeable Service Desk Operator

Self Service Desk
http://trait.topdesk.net

Service Desk
servicedesk@ctmm-trait.nl
088 - 1167500

trait.health-ri.nl/service-desk
Breast cancers in young women behave more aggressively

Mortality is higher

Proper tools for prognostication are lacking

Most young women receive adjuvant systemic treatment

→ Several women would have been cured by locoregional treatment alone
Identify cases in national cancer registry and samples in national pathology archive PALGA

- Case selection NCR (n=3524)
- Diagnosed between 1989 and 2000
- Females ≤40 years

1. **Identify**
   - Link cancer registry to PALGA cases
   - Linkage through trusted third party
   - Useful samples received for 2310 patients

2. **Linkage & tissue acquisition**
   - 12 million tissue samples stored in pathology archives
   - 42 million records on almost 10 million patients stored in the central PALGA data-bank
Digitalize slides and distribute across Europe

Slides are collected throughout the country

Digitalise slides

Zoom through cases like Google Maps
Collect pathology revisions from across Europe in central OpenClinica database

1. Identify cases
   - Histologically proven breast cancer
   - Primary invasive
   - AST naive
   - ≤ 40 years
   - No previous malignancies
   - Diagnosed between 1989-2000
   - No metastases

2. Linkage

3. Collect tissues

4. Scan slides

6. Digital revision
Next step: Secure collaboration in the cloud – the anDREa project providing a “Digital Research Environment” based on Microsoft Azure.

Source: Arnoud van der Maas; Radboudumc