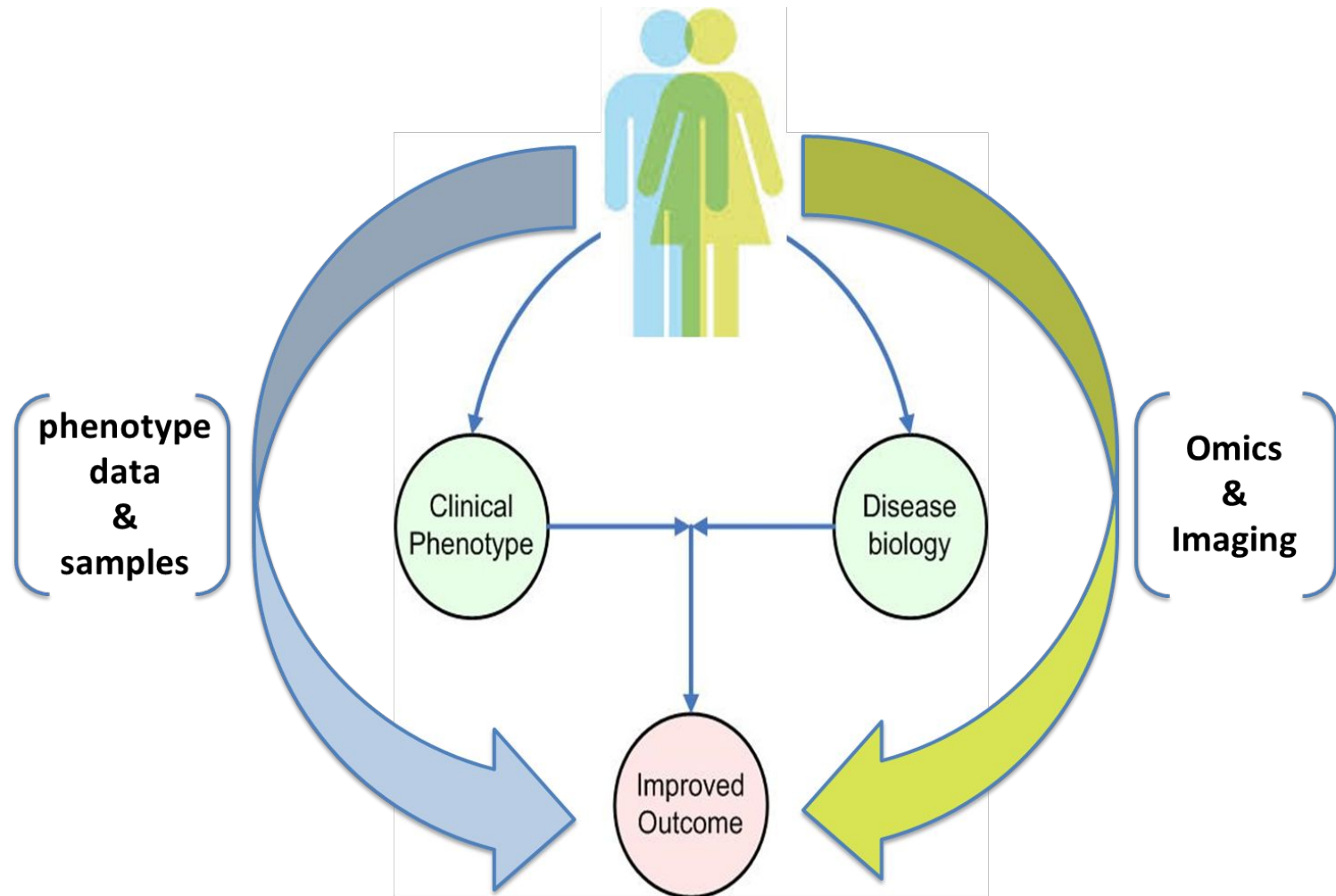




WP3.4 – Multi-modal data integration

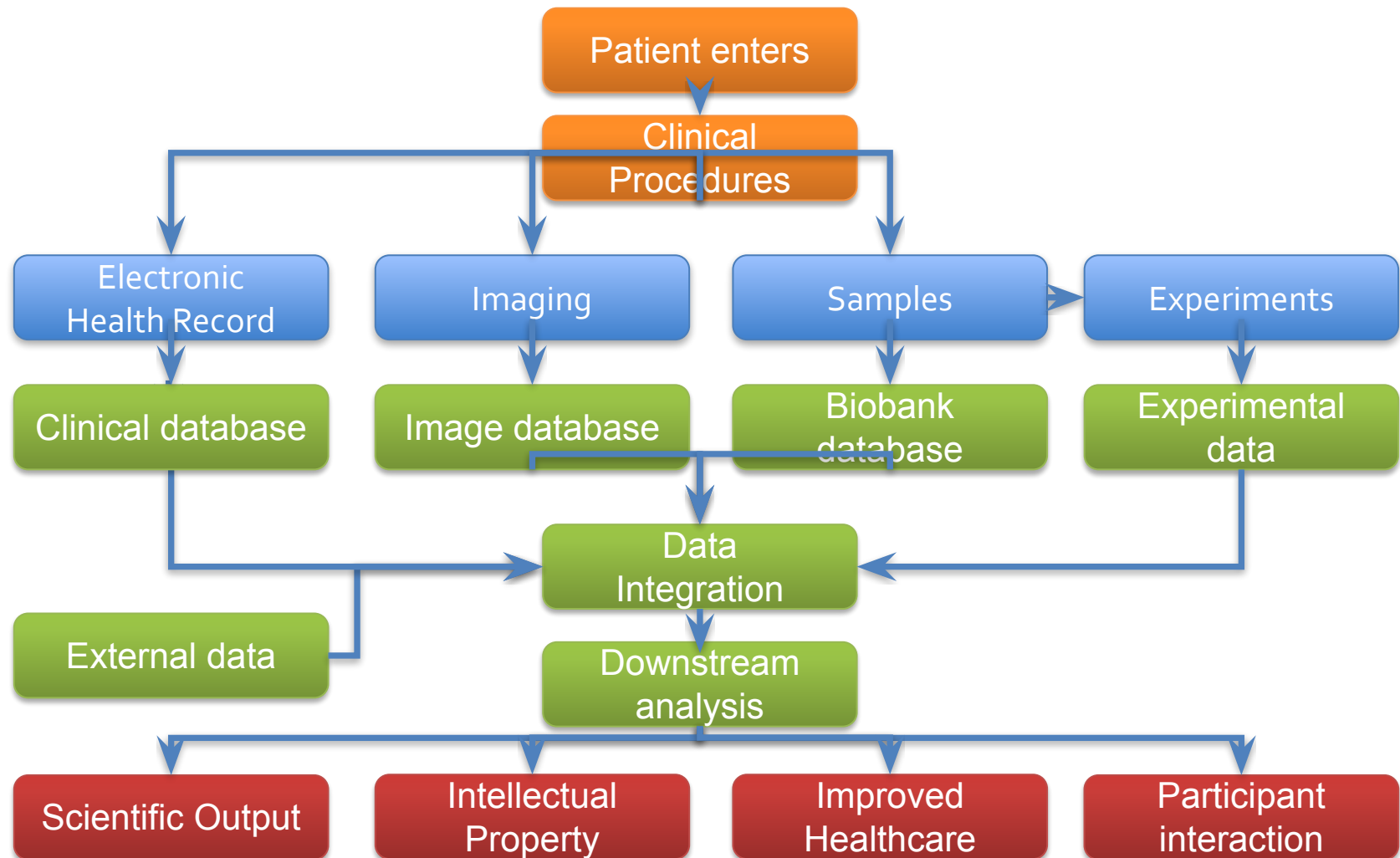


High-level design for biomarker research projects



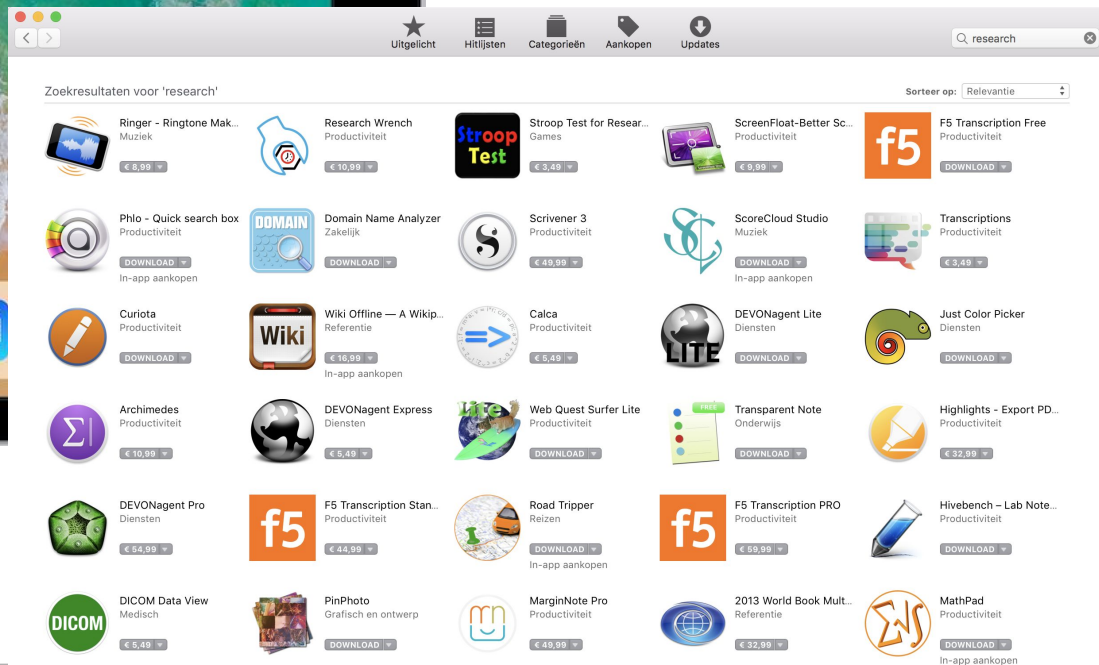


The underlying process is very multimodal and usually also multi-centric





"App-store" of open-source tools & procedures for multi-modal translational research?





Lots of complexity, but also many commonalities

Use case 3.4. Data integration and management services for image-driven and genomics-driven biomarker studies

Objective(s)

- To develop a common IT framework to support data handling and analysis for both clinical and preclinical biomarker research
- Applicable in any disease area
- Multimodal (clinical, imaging, omics, etc.)



Overview of use cases in task 3.4

- Two clinical use cases
 - Cancer Core Europe (NKI/Mariska Bierkens)
 - Osteoarthritis
 - IMI APPROACH use case (Lygature/Sjaak Peelen; UMCU/Paco Welsing; Rizzoli/Maria Mordenti)
 - Imaging biomarkers; PROOF study (Erasmus MC/Matthias Hansson, Stefan Klein)
- Tooling for clinical use cases
 - Data integration: tranSMART & cBioPortal
 - Image archive: XNAT
- Pre-clinical use case
 - Imaging based (Unito/Dario Longo)



tranSMART: hypothesis-free browsing of clinical and genomics data

Comparison

Summary Statistics

Grid View

Advanced Workflow

Analysis

aCGH Survival Analysis

Box Plot with ANOVA

Correlation Analysis

Frequency Plot for aCGH

Geneprint

Group Test for aCGH

Group Test for RNASeq

Heatmap

Hierarchical Clustering

K-Means Clustering

Line Graph

Logistic Regression

Marker Selection

PCA

RNA-Seq Differential Gene Expression (EdgeR)

Scatter Plot with Linear Regression

Survival Analysis

Table with Fisher Test

Waterfall

GenePrint

Customize

AURKA 50%

TP53 31%

MAL 100%

Gain Loss mRNA Downregulation

Copy number alterations are putative.

Intermediate Result - Job Name: geneprint-aCGHgroup-test-182723

chromosome	cytband	start	end	probab	is
1	1p16.32	3613512	3718612	0.0288	0.07450265125
1	1p16.32	3786155	4086118	0.0288	0.0707587437
1	1p16.32	4102310	4218027	0.0288	0.0707587437
1	1p16.32	4234488	4305880	0.0288	0.07369126102
1	1p16.32	4417012	4541570	0.0593	0.1144073094
1	1p16.32	4552352	5103861	0.0593	0.1144073094
1	1p16.32-1p16.31	5114962	5211863	0.0554	0.1136352941
1	1p16.31	5222615	5414265	0.0558	0.11554647405

Kaplan-Meier estimator

Subset/Target

Frequency Plot of Losses with FDR

Human GRCh37hg19 chr4:7,17,102,45,718,257

Selected Cohort

Clinical & Low Dimensional Biomarker Data
(Drag and drop low dimensional nodes here to filter the exported data.)

ACGH data
(Drag and drop high dimensional nodes here to filter the exported data.)

Messenger RNA data (Microarray)
(Drag and drop high dimensional nodes here to filter the exported data.)

68 patients

TSV

68 patients

BED

68 patients

TSV

Subset 2

(Private Studies|DeCoDe|Molecular profiling|Non-highthroughput molecular profiling|Microsatellite instab...

86 patients

TSV

86 patients

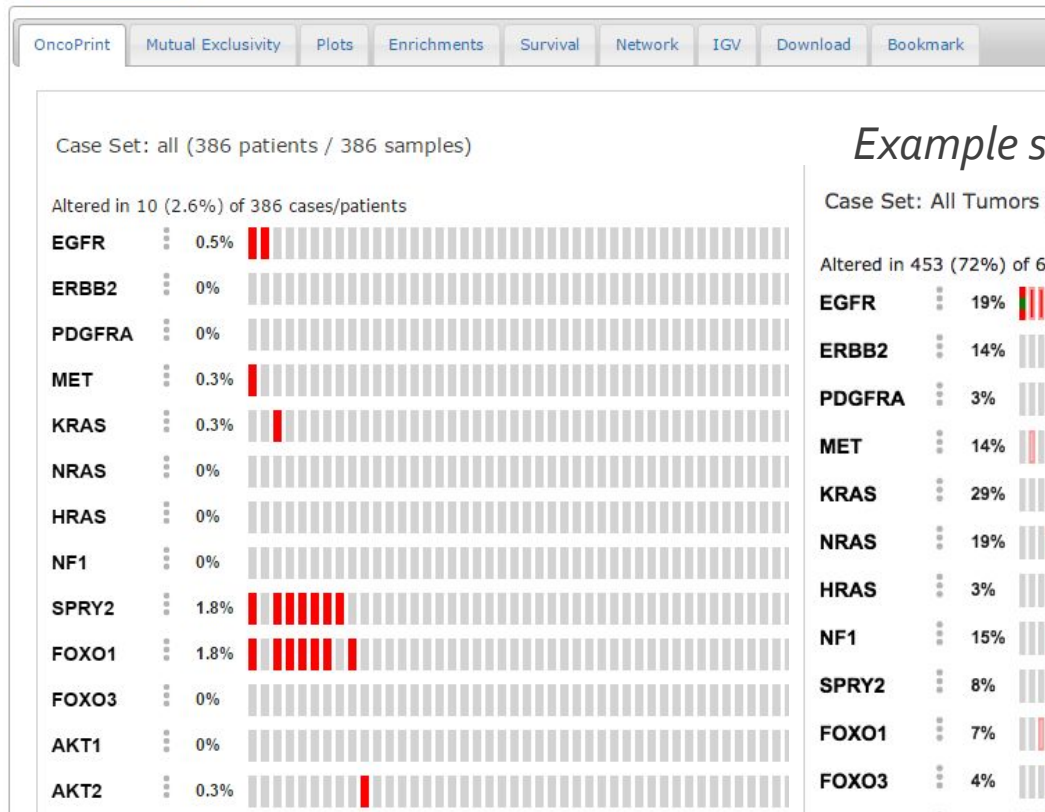
BED

86 patients

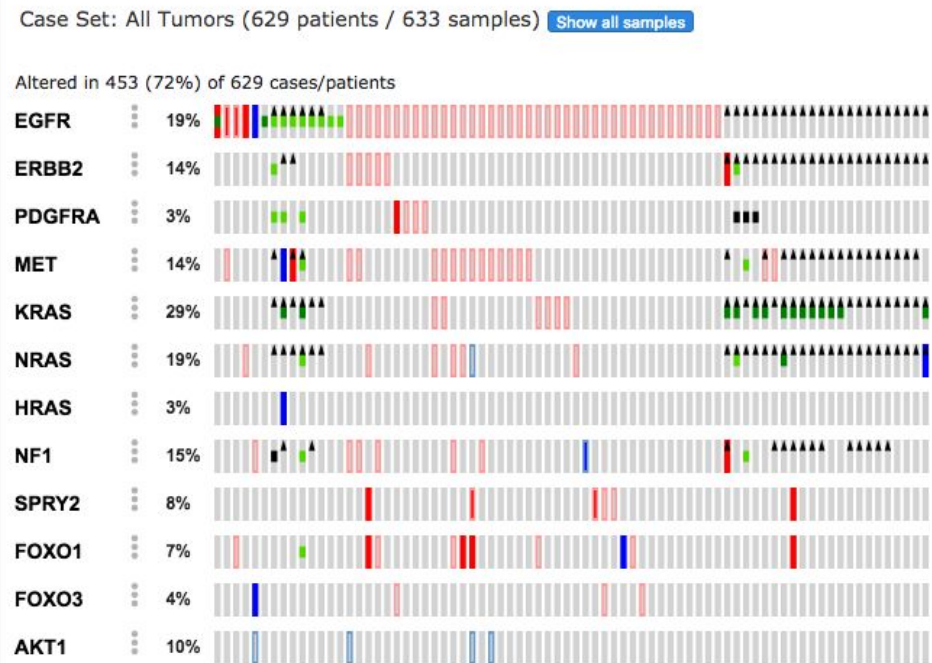
TSV



cBioPortal: gene-centric data integration



Example study with more omics data types



Screenshot cBioPortal.org



XNAT: open source imaging archive

- Web service: <https://xnat.bmia.nl>
- Store, organize, and share both imaging data and analysis results
- Open source platform

The screenshot displays the XNAT web application interface. On the left, a sidebar shows project navigation options. The main content area displays details for a specific MR session (subject0001), including accession number, date, and scanner information. Below this, a table lists tissue volumes. At the bottom, a table shows scan details. On the right, a large window displays a multi-panel MRI viewer with four panels showing different views of a brain scan. A separate window on the far right shows a detailed view of a specific MRI slice.

Tissue Volumes

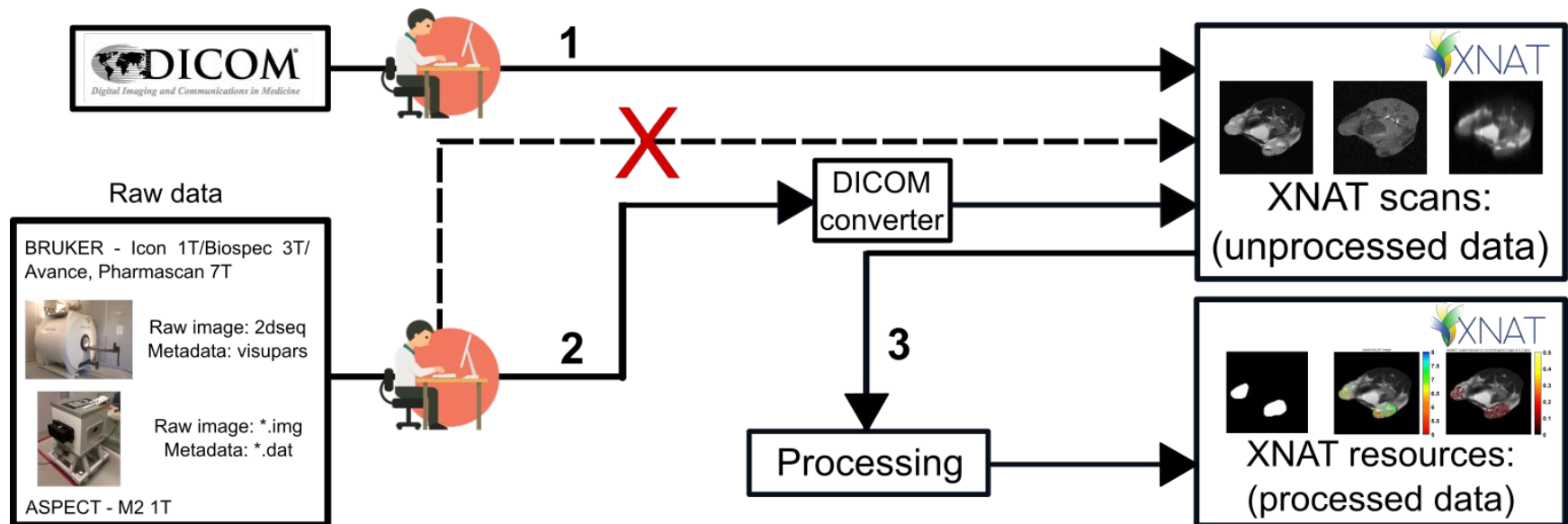
Volume	Value
WB_pvcv	83.41
WB_p	90
WM_pvcv	36.09
WM_p	89
GM_pvcv	47.32
GM_p	63
WML_pvcv	1.90
WML_p	72

Scans

Scan	Type	Series Desc	Usability	Files	Note
8:3	Sag FSPGR 3D	Sag FSPGR 3D	usable	Show Counts	
8:5	Ax T2 FLAIR	Ax T2 FLAIR	usable	Show Counts	
Total Counts					

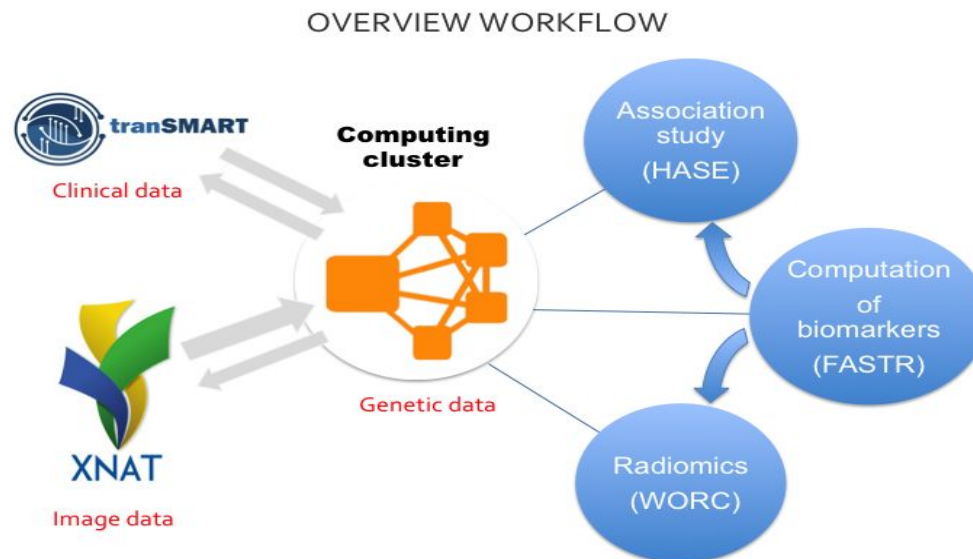
Preclinical use case: integration of XNAT

- Traditionally standard tools for preclinical (animal) imaging are lacking
- Overcome in CORBEL by upload pipelines to standard clinical tooling (XNAT) using a clinical standard (DICOM)
- Extra DICOM custom tags have been introduced



PROOF use case for osteoarthritis supported by a set of open-source tools ("apps")

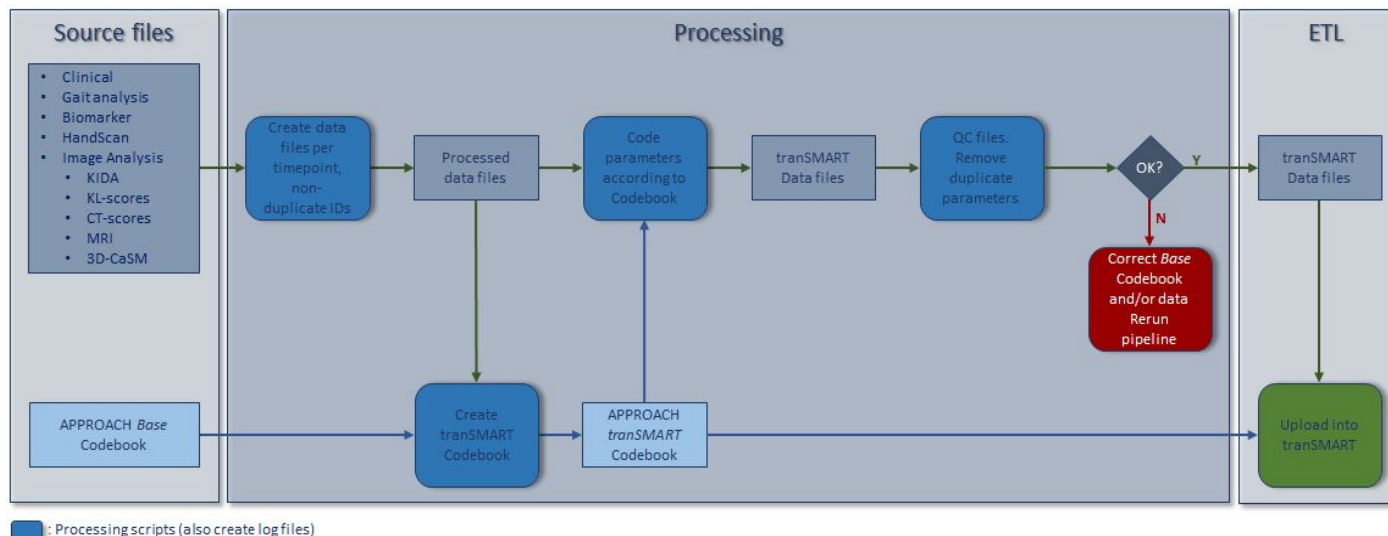
- Use case to test IT infrastructure for quantitative imaging
- Clinical study to develop biomarkers to predict knee cartilage damage and preventive strategies against this damage
- Generic workflow interacting with XNAT automatically:





IMI-APPROACH – cohort integration to detect new disease markers for osteoarthritis

- Integration of multiple existing cohorts to detect new disease stratification biomarkers
- Tooling: tranSMART and XNAT
- Data transformation and harmonization is the main challenge
 - Complex ETL pipelines highly automated to allow upload by non-experts





Cancer Core Europe: BRCA1 use case

- 7 Leading oncology centers in Europe wishing to share data at a common IT platform
- Selection made of common open-source tooling: OpenClinica, tranSMART, cBioPortal, Molgenis
- Experts in each of the centres trained for data loading
- Common standards set for data loading
- Data sharing between centres without a common legal contract proves to be very challenging: still finalizing the data processing agreements, etc.
 - Never to be underestimated!
- First results expected in the coming months



Tooling used: observations & conclusions

•tranSMART

- Progress in open-source community is slow; different forks
- Version used (16.2) lacks support for longitudinal data and incremental data support
- Data upload process requires very careful planning using standard pipelines
 - CORBEL data formatting recommendation document available

•cBioPortal

- Powerful tool highly appreciated; also potential use outside oncology
- Active open-source community with abundant on-line documentation at the cBioPortal website
- Standard upload pipelines and documentation available from CORBEL



Key conclusions about the tooling - XNAT

- Image sharing platform suitable for preclinical and clinical images; de facto standard in multi-center imaging projects
- Open-source community is active and vibrant
- Pseudonymization pipelines prove to be challenging: CTP is another open-source tool highly recommended for this purpose
- Built-in API structure makes it very suitable for image analysis
- For multi-center studies careful naming conventions need to be adopted across the project

Key conclusions



It has been demonstrated that a generic infrastructure based on open-source tools is broadly usable across disease areas



Many practical hurdles had to be overcome to realise working implementations in the use cases



Over-all recommendations

- Handling multimodal data should be implemented in a modular way rather than monolithic
- Fine-grained access control is a requirement since the data is often highly sensitive
- Container technology to deploy the entire infrastructure locally should be considered in order to simplify legal red tape & to cloudify the tooling
- Domain experts should be deeply involved in the data harmonisation process from the start
- Data harmonisation efforts should never be underestimated and should be avoided whenever possible (“FAIR by design”)
- Ethical/legal prerequisites require careful planning