Integrative Cancer Research
INTEGRATE
Through Innovative Biomedical Infrastructures

SEVENTH FRAMEWORK PROGRAMME

ICT-2010-270253

INTEGRATE

Driving excellence in Integrative Cancer Research through Innovative Biomedical Infrastructures

STREP
Contract Nr: 270253

Deliverable: D6.5: Scenarios for validation and detailed metrics for the validation of tools

Due date of deliverable: (02-28-2014)
Actual submission date: (03-11-2014)

Start date of Project: 01 February 2011 Duration: 36 months

Responsible WP: FORTH

Revision: <outline, draft, proposed, **accepted**>

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination level		
PU	**Public**	**X**
PP	Restricted to other programme participants (including the Commission Service	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (excluding the Commission Services)	

0 DOCUMENT INFO

0.1 Author

Author	Company	E-mail
Alexandre Irrthum	BIG	alexandre.irrthum@bordet.be
Evaggelia Maniadi	FORTH	maniadi@ics.forth.gr
George Manikis	FORTH	gmanikis@ics.forth.gr
Ioannis Karatzanis	FORTH	karatzan@ics.forth.gr
Keyur Mehta	GBG	keyur.mehta@germanbreastgroup.de
Lefteris Koumakis	FORTH	koumakis@ics.forth.gr
Maxime Steisel	IJB	maxime.steisel@bordet.be
Njin-Zu Chen	Philips Research	njin-zu.chen@philips.com
Thanasis Iliopoulos	FORTH	ailiop@ics.forth.gr

0.2 Documents history

Document version #	Date	Change
V0.1	22/01/2014	Starting version, template
V0.2	03/02/2014	Definition of ToC
V0.3	17/02/2014	First complete draft
V0.4	19/02/2014	Integrated version (send to WP members)
V0.5	20/02/2014	Updated version (send PCP)
V0.6	20/02/2014	Updated version (send to project internal reviewers)
Sign off	11/03/2014	Signed off version (for approval to PMT members)
V1.0	11/03/2014	Approved Version to be submitted to EU

0.3 Document data

Keywords	
Editor Address data	Name: Ioannis Karatzanis Partner: FORTH Address: Nikolaou Plastira 100, Vassilika Vouton GR - 711 10, Heraklion, Crete, Greece Phone: +30 2810 39 1618 Fax: E-mail: karatzan@ics.forth.gr
Delivery date	

0.4 Distribution list

Date	Issue	E-mailer
		fp7-integrate@listas.fi.upm.es

Table of Contents

0 DOCUMENT INFO 2
0.1 Author 2
0.2 Documents history 2
0.3 Document data 2
0.4 Distribution list 2
1 INTRODUCTION 5
2 EVALUATION METHODOLOGY 6
2.1 ISO 25000 series and Validation procedures 6
2.2 Evaluation procedures 8
2.3 Evaluation and validation for INTEGRATE 11
3 TEST BEDS 13
4 DATA SETS 14
4.1 GBG 14
4.2 The IJB Breast MDT 14
4.2.1 MULTIDISCIPLINARY TEAM MEETING 14
4.2.2 GENERIC INFORMED CONSENT 14
4.2.3 DATA STRUCTURE 15
4.3 Maastro 15
4.4 SAGE 15
4.5 TOP 15
4.6 Image Data for the Central Pathology Review 16
5 VALIDATION SITES 17
6 VALIDATION OF THE PATIENT SCREENING TOOL DECIMA 19
6.1 Overview 19
6.2 Qualitative evaluation 19
6.3 Quantitative evaluation 19
6.3.1 EVALUATION SCENARIO 20
6.3.2 TIMED SCREENING TASK 21
6.3.3 METRICS 22
7 VALIDATION OF THE COHORT SELECTION TOOL NONA 23
7.1 Overview 23
7.2 Qualitative evaluation 23
8 VALIDATION OF CENTRAL REVIEW FOR PATHOLOGY IMAGES TOOL 24
8.1 Overview 24
8.2 Qualitative evaluation 24
8.3 Quantitative evaluation 24

8.3.1 EVALUATION SCENARIOS 25
8.3.2 TIMED REVIEWING TASKS 25
8.3.2.1 Actors Involved 26
8.3.2.2 Pre-conditions 26
8.3.2.3 Work Steps 27
8.3.3 METRICS 29

9 VALIDATION OF THE ANALYSIS PLATFORM TOOL 30
9.1 Overview 30
9.2 Qualitative evaluation 30
9.3 Quantitative evaluation 30
9.3.1 EVALUATION SCENARIO 31
9.3.2 TIMED TASK 31
9.3.2.1 Actors involved 32
9.3.2.2 Pre-conditions 32
9.3.2.3 Work steps 32
9.3.3 METRICS 33

10 SUMMARY 34
11 REFERENCES 35
12 APPENDIX 36

1 Introduction

Evaluation is the systematic determination of the extent to which an entity meets its specified criteria. The evaluation of software product quality is vital to both the acquisition and development of software. The relative importance of the various characteristics of software quality depends on the intended usage or objectives of the system of which the software is a part; software products need to be evaluated to decide whether relevant quality characteristics meet the requirements of the system.

As the complexity and code size of the software increase, the risks of having a failure increase as well, and there is no effective general solution to the size, complexity, quality and other software engineering problems. However, by following standardized software development practices and by addressing the quality issues during the whole life cycle of the software, the likelihood of such defects and the cost incurred by them (both to users and to producers) may be greatly reduced.

The purpose of this document is to introduce the unified approach for ensuring the quality of the software products produced within the project, in accordance with the guidelines from the end users. So, in this document the procedures for the evaluation and validation activities will be established and qualitative measures of the benefits of the project as a whole will be developed.

In INTEGRATE, four tools will be evaluated, the cohort selection tool, Nona, the patient screening tool, Decima, the central pathology review (CPR) tool, and the analysis platform.

The implementation of this approach is adapted from various sources and mainly from the ISO/IEC 25000 series.

Abbreviations

BIG	Breast International Group
FORTH	Foundation for Research & Technology - Hellas
GBG	German Breast Group
IJB	Institut Jules Bordet
Maastro	Maastricht Radio-Oncology Clinic
SUS	Standardized Usability Score
UdS	Universitaetsklinikum des Saarlandes

2 Evaluation methodology

The goal of the validation and the evaluation is to ensure that the software produced in each technical WP is compliant with the end-user specifications. The relative importance of the various characteristics of software quality depends on the intended usage or objectives of the system.

Evaluation modules contain the specification of the quality model (i.e. characteristics, sub-characteristics and corresponding internal, external or quality in use measures), the associated data and information about the planned application of the model and the information about its actual application. Appropriate evaluation modules have been selected for the evaluation and validation based on the Software product Quality Requirements and Evaluation (SQuaRE) – Evaluation reference model and guide (1).

2.1 ISO 25000 series and Validation procedures

ISO and the International Electrotechnical Commission (2) form the specialized system for worldwide standardization. The ISO SQuaRE will be used as reference model. Joint Technical Committee ISO/IEC JTC 1, Information technology, Subcommittee SC 7, Software and systems engineering, prepared ISO/IEC 25010. ISO/IEC 25010 is a part of the SQuaRE series of International Standards, which consists of the following divisions:

- Quality Management Division ISO/IEC (3)
- Quality Model Division ISO/IEC (4)
- Quality Measurement Division ISO/IEC (5)
- Quality Requirements Division ISO/IEC (6)
- Quality Evaluation Division ISO/IEC (7)
- SQuaRE Extension Division ISO/IEC 25050 – ISO/IEC 25099

This first edition of ISO/IEC (4) cancels and replaces ISO/IEC 9126-1:2001, which has been technically revised.

ISO/IEC 9126:1991 was replaced by two related multipart standards: ISO/IEC 9126, *Software engineering — Product quality* and ISO/IEC 14598, *Software engineering — Product evaluation*. This International Standard revises ISO/IEC 9126-1:2001, and incorporates the same software quality characteristics with some amendments.

- The scope of the quality models has been extended to include computer systems, and quality in use from a system perspective.
- Context coverage has been added as a quality in use characteristic, with sub-characteristics *context completeness* and *flexibility*.
- *Security* has been added as a characteristic, rather than a sub-characteristic of functionality, with sub-characteristics *confidentiality*, *integrity*, *non-repudiation*, *accountability* and *authenticity*.
- C*ompatibility* (including *interoperability* and *co-existence)* has been added as a characteristic.

- The following sub-characteristics have been added: *functional completeness, capacity, user error protection, accessibility, availability, modularity* and *reusability*.
- The compliance sub-characteristics have been removed, as compliance with laws and regulations is part of overall system requirements, rather than specifically part of quality.
- The internal and external quality models have been combined as the product quality model.
- When appropriate, generic definitions have been adopted, rather than using software-specific definitions.
- Several characteristics and sub-characteristics have been given more accurate names.

ISO 25000 series International Standard defines:

- A software product quality model composed of eight characteristics (functional suitability, reliability, performance efficiency, operability, security, compatibility, maintainability and portability), which are further subdivided into sub-characteristics and relate to static properties of software and dynamic properties of the computer system (See Figure 1). The model is applicable to both computer systems and software products.
- A quality in use model composed of five characteristics (some of which are further subdivided into sub-characteristics) that relate to the outcome of interaction when a product is used in a particular context of use. This system model is applicable to the complete human-computer system, including both computer systems in use and software products in use.

A quality model is a set of requirements, entities and relationships that must be fulfilled to assess good quality. The model should be structured in three main levels:

- Characteristic
- Sub-characteristic
- Attribute

We can refer to two models of quality:

- the internal and external quality
- the quality in use

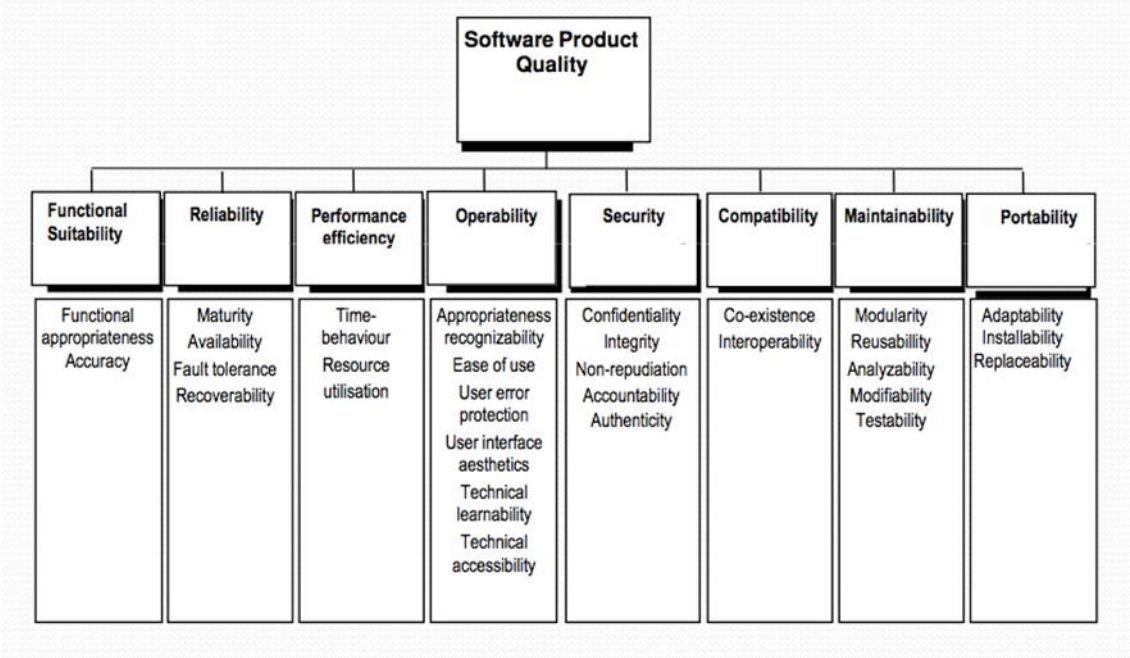

Figure 1: Software product quality categories and characteristics (source ISO/IEC 25040)

The selected software product quality measures shall be applied to the software product and components, according to the evaluation plan, resulting in values on the measurement scales.

For each tool in INTEGRATE the developers have described the functional and non-functional requirements and identify how will measure those characteristics categorized according to the ISO/IEC 25023.

2.2 Evaluation procedures

None of the quality characteristics discussed above can be measured directly, but must be assessed in terms of objective sub-characteristics. ISO/IEC 25000 series does not prescribe specific quality requirements for software, but instead describes a quality model, which can be applied to any software.

End-user evaluation of the infrastructure will be conducted through a number of selected scenarios covering the anticipated usage of the infrastructure, from administration of the software components to specific clinical trials. For each step in the scenario, the required input data are enumerated and a description of the expected results will be given. The steps listed for the execution of the scenarios correspond to criteria which will help objectively rating the degree of success of the modules addressed therein. The end users who will participate to the evaluation phase will fill in an evaluation form for each component. The evaluation form will cover all the appropriate quality characteristics from the product quality model of the ISO/IEC 25000 series (Figure 1).

At the evaluation phase different type of users, such as physicians, system developers and bioinformaticians will participate. Having such a diverse target group of evaluators, the evaluation forms must be:

- simple
- accurate
- easy to understand (especially for non IT experts)
- non time consuming
- without loss of functionality/quality

For that reason, we have translated the crucial sub-characteristics of software quality measures into simple questions (in natural language). The evaluation form will be a list of such questions where the evaluator will answer with a degree of satisfaction with scale 5 (from 1 to 5).

The selected sub-characteristics for the evaluation form and their translation into simple questions can be found in the table below (table 1).

Functionality	Suitability	Can software perform the tasks required?
	Accurateness	Is the result as expected?
	Interoperability	Can the system interact with another system?
	Compliance	Is the system compliant with standards?
Efficiency	Time Behaviour	How quickly does the system respond?
	Resource utilization	Does the system utilize resources efficiently?
Compatibility	Co-existence	Can the system share resources without loss of its functionality?
	Interoperability	Can the system share information/data with other components?
Usability	Understandability	Does the user comprehend how to use the system easily?
	Learnability	Can the user learn to use the system easily?

	Operability	Can the user use the system without much effort?
	Attractiveness	Does the interface look good?
Reliability	Maturity	Have most of the faults in the software been eliminated over time?
	Fault tolerance	Is the software capable of handling errors?
	Recoverability	Can the software resume working & restore lost data after failure?
Security	Authenticity	Does the system provide identification access wherever is needed?
	Confidentiality	Are data accessible only to authorized users?
	Accountability	Can the system trace actions uniquely?
	Integrity	Does the system prevent unauthorized access?
Maintainability	Analysability	Can faults be easily diagnosed?
	Changeability	Can the software be easily modified?
	Stability	Can the software continue functioning if changes are made?
	Testability	Can the software be tested easily?
Portability	Adaptability	Can the software be moved to other environments?
	Installability	Can the software be installed easily?
	Conformance	Does the software comply with portability standards?
	Replaceability	Can the software easily replace other software?
Quality of use	Effectiveness	How accurate and complete is the software for the intended use?

	Efficiency	Does the software improve the time or reduce resources for the intended goal?
	Satisfaction	Does the software satisfy the perceived achievements of pragmatic goals?
	Health and safety risk	Can the software harm people in the intended contexts of use?

Table 1: From software quality characteristics to NL questions

We also use the System Usability Scale (SUS) as a generic tool for measuring the usability. SUS has become an industry standard and allows the evaluation of a wide variety of products and services, including hardware, software, mobile devices, websites and applications. When a SUS is used, participants are asked to score the following 10 items with one of five responses that range from Strongly Agree to Strongly disagree:

- I think that I would like to use this system frequently.
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very cumbersome to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

2.3 Evaluation and validation for INTEGRATE

The validation plan within INTEGRATE is based on the user requirements. We built special tools/scenarios in order to fulfill specific requirements. Each developer has identified the initial user requirements. The validation procedure identifies the specifications which conform or fail to conform to the user needs and the intended uses.

For each component, the developers have described the functional and non-functional requirements. At the evaluation phase different type of users, such as physicians, system developers and biostatisticians will participate. Having such a diverse target group of evaluators, the evaluation forms must be simple. For that reason we have translate the crucial sub-characteristics of software quality measures into simple questions (in natural language). The evaluation form will be a list of questions where the evaluator will answer with a degree of satisfaction on a Likert scale. The Likert scale is based on forced-choice questions, where a statement is made and the

respondent then indicates the degree of agreement or disagreement with the statement on a 5 point scale. The ***Evaluation Questionnaire*** consists of two forms:

- **Form A**: The selected sub-characteristics, for the evaluation form of the scenarios, and its translation into a simple question for the end user can be found in the Generic Evaluation Questionnaire.
- **Form B**: We use the System Usability Scale (SUS) for global assessment of systems usability. The SUS can be found at the Generic Evaluation Questionnaire.

Having identified both functional & non-functional requirements (Validation step) we identified and adopted the evaluation questionnaire (Table 1) according to the needs of each component. The form B measures the usability and will be the same for all the components.

In some cases, we decided to perform an extensive evaluation, separate from the formal evaluation, which will prove the impact of the INTEGRATE project. The idea is that measurable parameters will be established in cooperation with the responsible clinical partners (e.g. recruitment rate, the number of SAE/SUSAR avoided etc.) for each tool offered within INTEGRATE. Those measurable parameters will be first monitored during a pre-defined time frame without using the INTEGRATE infrastructure. Then, INTEGRATE tools will be used and the same parameters will be monitored. In this way we will be able to demonstrate the real impact of the infrastructure.

The following sections describe in detail the evaluation and validation methodologies for each tool.

3 Test beds

We will be using two test beds for the INTEGRATE evaluation:

- INTEGRATE evaluation test bed: The distributed INTEGRATE evaluation test bed
- Local test bed at Institut Jules Bordet (IJB)

For the IJB test bed almost all components will be locally deployed at IJB in Brussels. Tiling Service, Imaging Service, R server and Report Generator Service will be deployed remotely, at the FORTH's infrastructure. The components that make up the distributed test bed are run at the sites of the collaborators. Table 2 lists the maintainer and location of each of the components.

	Component	Maintainer	Physical server location
Nona, Decima	Clinical Data Warehouse	UPM	Madrid, Spain
	Semantic Layer	UPM	Madrid, Spain
	Evaluation Engine	Custodix	Gent, Belgium
	Security	Custodix	Gent, Belgium
	Patient ID service	Custodix	Gent, Belgium
	Trial Metadata Database	Philips Research	Eindhoven, Netherlands
	Locker	Philips Research	Eindhoven, Netherlands
Pathology review, Analysis Platform	Common Information Access service	UPM	Madrid, Spain
	Data Push service	UPM	Madrid, Spain
	Semantic Layer	UPM	Madrid, Spain
	Security	Custodix	Gent, Belgium
	Tiling Service	FORTH	Heraklion, Greece
	Imaging Service	FORTH	Heraklion, Greece
	R Server	FORTH	Heraklion, Greece
	Report Generator Service	FORTH	Heraklion, Greece

Table 2: Maintainers and physical location of the components of the distributed test bed

The patient screening tool, Decima, and the cohort selection tool, Nona, are native Windows applications. A laptop with a complete installation will be provided by Philips Research for all evaluation sessions. The central review for pathology images tool and the analysis platform tool have web-based interfaces, and will run on a pc (laptop or workstation) with a browser capable to handle html 5 content.

4 Data sets

4.1 GBG

The GBG Forschungs GmbH is going to deliver data of the following trials:

- TBP (Metastatic study)
- GAIN (Adjuvant study)
- GeparQuattro (NeoAdjuvant study)

These trials are completed, their data were analysed and the results were published.

The data is available in the csv format and is about:

- Baseline data
- Adverse Events during Therapy

The data does not contain free text fields, the following terms present a part of the column headers of the csv file:

- Thromboembolic events
- Eye disorders
- Anaemia, grade 3-4
- Neutropenia, grade 3-4
- nausea
- vomiting
- mucositis
- constipation

4.2 The IJB Breast MDT

The dataset that we will use for the validation at IJB is the structured medical data collected to prepare the breast Multidisciplinary Team Meeting for patients that have signed the generic informed consent.

4.2.1 Multidisciplinary Team Meeting

Due to the inherent multidisciplinary aspect of oncology, treatment options for patients (including a possible inclusion in a clinical trial) are discussed during a Multidisciplinary Team Meeting (MDT) held every week bringing together experts in different speciality, including surgeons, pathologist, radiologist, radiotherapist and oncologists. To prepare this meeting, a data manager mines the file (which consist mainly of textual reports) of each patient and extracts all relevant information to a different database, allowing experts to have a normalized executive summary of the patient state.

4.2.2 Generic Informed Consent

Since 2011, each patient is given a generic informed consent for residual material utilisation and secondary data use for any scientific project that IJB's ethics committee considers useful. As of writing, 889 patients have signed this informed consent, of

which 317 have been discussed in the multidisciplinary breast cancer conference. Because the same patient can be discussed in several MDT, we have 347 records of discussed patient in total.

4.2.3 **Data Structure**

The data consist of a database dump in TSV (tab separated values). The columns include most relevant information in a structured way, including the patient and his family history; gynaecologic information, the TNM classifications of the tumor and a selection of important laboratory results. See Appendix D for details.

4.3 Maastro

This is a set of approx. anonymized 3000 records of historical patients from the Maastro Radiation Oncology clinic in Maastricht, the Netherlands. It goes together with a list of ten trials that Maastro has been recruiting for during the time of treatment of these patients.

4.4 SAGE

METABRIC is a public dataset and is available through the Synapse Commons data repository (synapse.sagebase.org), an IRB-approved data repository. Data contributed to this repository is made available under a tiered system, designed to protect the privacy and confidentiality of study participants.

The METABRIC dataset contains SNP genotypes, expression profiles, Copy Number Variant (CNV) profiles and clinical traits derived from 1981 breast cancer tumors collected from participants of the METABRIC trial. Gene expression data are performed on the Illumina HT 12v3 platform. Copy number data are performed on the Affymetrix SNP 6.0 platform. Both data are normalized as described in (8). Clinical covariates and survival data are also provided and fully described in (9).

4.5 TOP

The Trial of Principle (TOP) study was started for the prospective evaluation of Topoisomerase II Alpha (*TOP2A*) gene amplification and protein overexpression as markers predicting the efficacy of Epirubicin in the primary treatment of breast cancer patients. The TOP trial included 149 patients, 139 of whom were evaluable for response prediction analyses. The primary end point was pathologic complete response (pCR). TOP2A and gene expression profiles were evaluated using pre-epirubicin biopsies. Gene expression data from ER-negative samples of the EORTC (European Organisation for Research and Treatment of Cancer) 10994/BIG (Breast International Group) 00-01 and MDACC (MD Anderson Cancer Center) 2003-0321 neoadjuvant trials were used for validation purposes.

The structure of the data collected in the TOP trial has been described in details in deliverable (10).

4.6 Image Data for the Central Pathology Review

The slides which are used in the Central Review for Pathology Images platform, were provided from the BCTL laboratory at IJB. They are biopsies from breast cancer tumors, either haematoxylin and eosin stained (H&E) or immunohistochemistry-stained for markers HER2, ER and PgR. In total there are a dozen images of digital pathology slides, which are in ndpi file type format and they were scanned by a Hamamatsu scanner. The files of the acquired images have sizes from 150 megabytes up to 840 megabytes, and contain image scans at a 40x magnification.

5 Validation sites

Table 3 summarizes the potential evaluation sessions and sites for the tools, as well as the intended user audience. The details may change depending on the availability and accessibility of the sites. Each tool will be tested in at least one site.

	Validation Sites						
	IJB	BIG	Maastro	GBG	Evaluation Workshop	UdS	Target user
Decima - Patient Screening - qualitative	X					X	Oncologists
Decima - Patient screening - quantitative	X		X	X	X		Trial nurses
Nona – Cohort Selection - qualitative		X		X	X		Oncologists
Pathology Review - qualitative	X				X		Pathologists
Pathology Review - quantitative	X				X		Pathologists
Analysis Platform- qualitative	X						Oncologists, Biostatisticians
Analysis Platform- quantitative	X				X		Oncologists, Biostatisticians

Table 3: Overview of tools and potential test sites, and the intended user audience. The colours indicate sessions.

These evaluation sessions for the INTEGRATE tools have been performed already in 2012:

- IJB
- UdS
    - These two sessions cover a qualitative evaluation of the early patient screening prototype tool, running with a limited number of test patients and trials.

The following evaluation sessions for the INTEGRATE tools will be performed in 2014:

- Maastro
  - This is a qualitative test of the patient screening tool, including performance measurements that can be related to historical data. The evaluation will run on the INTEGRATE evaluation bed. The participating users are Maastro staff, very familiar with the tasks to be evaluated.
- IJB/BIG
  - This is a comprehensive setting to evaluate all tools. The entire INTEGRATE infrastructure will also be deployed locally at IJB, to allow for use of local data, except four services (Tiling Service, Imaging Service, R server and Report Generator Service) that will be deployed remotely, in FORTH's infrastructure The users participating here are local IJB staff, familiar with the tasks to be evaluated. Reference tests will be conducted to benchmark the tool performance where possible.
- Evaluation Workshop
  - This is a comprehensive setting to evaluate all tools. This evaluation will run on the INTEGRATE evaluation bed. The users participating are invited from outside the INTEGRATE consortium.
- GBG
  - This is a comprehensive setting to evaluate Nona and Decima tools. There will be at least two GBG staff members involved. This evaluation will run on the INTEGRATE evaluation bed.

6 Validation of the Patient Screening Tool Decima

6.1 Overview

The purpose of the Decima tool and the underlying referenced services is to evaluate the potential enrollment of a patient for a clinical trial, starting from an individual patient. This use case is described in detail in (11), Section 3.6.2. The elements relating to the ordering of molecular testing and informed consent are left out, since we do not test with live patient data.

6.2 Qualitative evaluation

The qualitative evaluation consists of a set of structured interviews, and a tool review. Two sessions are conducted: one at IJB with four oncologists, and one at UdS with one oncologist. The interview consisted of the following parts:

1. Current situation and approach
    - The larger treatment workflow
    - The patient screening workflow
    - The patient record as it relates to the screening process

2. Paper prototype: (only performed at IJB): a test with a paper criterion cards to investigate the preferred information presentation structure.

3. Review of the Decima prototype with a 22” touchscreen. This was done with dummy data in an early implementation of the prototype tool. We asked the doctors to think aloud while performing tasks, to investigate the preferred workflow.

4. Discussion. How will this tool affect the work of the oncologists and how can we optimize the concept?

There are no metrics at this stage. The outcome is a report summarizing the findings. This is used to refine the tool, and prepare it to go into quantitative testing. If required, multiple qualitative tests can be performed.

6.3 Quantitative evaluation

The evaluation is conducted according to ISO/IEC 25023. The questionnaires used are reproduced in the appendix (section 12). Additionally, we measure the SUS (also in the appendix), and metrics specific to the patient screening process (see Section 6.3.3).

There are four potential quantitative evaluations, at the following sites:

- Maastro
    - This evaluation involves at least two users. These will be local staff who are performing the screening task as part of their normal daily work. Timing will be compared to the historical data from the users. The test will run with

historical patient data from Maastro, and a set of 10 trials that have been used in the timeframe from which the patient data is taken. From these 10, five concern breast cancer, and five concern lung cancer.

- IJB
  - This evaluation involves at least three users. These will be local staff who are performing the screening task as part of their normal daily work. Timing will be compared with a reference test in which the current way of screening is used. The test will be run with historical local data, and trials that run or have run at IJB.
- Evaluation Workshop
  - This evaluation involves at least three users. These will be representative users recruited through the INTEGRATE network, but who are not part of the consortium.
- GBG
  - This evaluation involves at least three users. These users will be recruited from GBG staff. The setup is the same as for the Evaluation Workshop.

6.3.1 **Evaluation scenario**

As trial nurses are the users most often performing the patient screening task, the qualitative evaluation is aimed at them primarily.

For each participant, the test scenario is as follows:

- Consent Form
- Introduction to the system
- Instruction on how to operate the patient screening tool Decima, and time to get familiar with the tool
- Guided test: (see test case described in Section 6.3.2)
  - Take a set of test patients. We use all patients[1] registered on a single date. We take a random date with at least 15 patients, to get enough test cases, but at most 20 patients, so that the duration of the test is limited.
  - For each patient in the test set, let the investigator judge the eligibility status. Time this. Count the number of correctly assigned patients, number of missed patients and number of erroneously assigned patients.
  - Performance measures:
    - Time taken
    - Assignment status per patient
- Usability questionnaire and software quality questionnaire afterwards (Questionnaire A and B as in 12).
- Exit interview and rounding off

[1] At Maastro, this number is on average 18 patients per day, but it varies greatly.

The test should fit in a 90 minute time window.

After the test, the system is reset. This brings the database back to its starting state. Any decisions on enrolment, annotations, validation of criteria etc. are erased.

6.3.2 **Timed screening task**

The evaluation task for Decima is derived from UC.1: Patient trial screening (11) . Elements related to consent and ordering extra molecular testing are excluded. The aim is to simulate as much as possible a typical screening session as it would occur in the normal daily work setting of the trial nurse. The instruction reads as follows:

"As a trial nurse at Maastro you prescreen intake patients for eligibility in the clinical trials running at Maastro. Your task is to use the Decima patient screening tool to systematically go through all the patients in today's worklist. One by one you should determine the patient's eligibility for current trials. If the patient is eligible mark the patient as eligible in the appropriate trial. Once you have gone through all the patients in today's worklist, you should export the results. "

(Note: this sample text is localized for the Maastro evaluation, and will be slightly altered for the other evaluation sites.)

Actors Involved

- Investigator

Pre-conditions

- The services for running Decima are available.
    - Authentication service
    - Patient Info service
    - Trial metadata service
    - Criterion Matcher service
    - Persistence service
- An investigator can authenticate himself to the system and is authorized to use all services.

Work steps

1. The investigator opens Decima, and logs in with his credentials.
2. The investigator sees a list of patients. For each patient, the list of available trials can be expanded.
    - Per trial, the additional information (like the trial enrollment closure date, number of required patients, etc.) is shown. Overall criterion matching status is shown (how many criteria are eligible, ineligible and unknown).

This is the added result of the computed matching plus any manual overrides.

3. By selecting a specific trial, the investigator can see the detailed information per criterion (inclusion/exclusion, evidence, matching status as suggested, etc.).
    a. Per criterion, the investigator can either validate the suggested matching status or override to set an explicit matching status.
    b. This step can be repeated for each trial.
4. The investigator can return to the patient – trial overview, and there choose what to do with the patient:
    a. Indicate that the patient is selected for enrolment in a particular trial.
    b. Remove the patient from the worklist, if the patient cannot match any trial
    c. Leave the patient on the list for reevaluation later on.

    This step is repeated for all patients.

5. If all patients have been analyzed, the investigator can hit the 'export' button to create a summary of the patients that are eligible, plus the trials for which they are eligible. This ends the test case.

The timings and the exported list are used for determining metrics. The test environment is reset for the next test.

6.3.3 **Metrics**

We track the following metrics in this evaluation:

- The outcome of the questionnaires as mentioned in the appendix (section 12). The adjusted ISO/IEC questionnaire is split into an end-user part and a technical part. The technical part (Part C) is only evaluated once by the development team. The remaining part for the end-user is Questionnaire A (the remaining ISO/IEC questions) and Questionnaire B, which is the standard SUS.
- The time performance of the users will be recorded, per patient. We derive the overall time per patient, time per assigned patient, and time per not-assigned patient.
  The target is to have a shorter time per patient on all three measures.
- Per patient, the assignment status is recorded. We track:
    - The rate of false positives
    - The rate of false negatives and
    - The overall enrolment rate.

  We should not see false negatives or positives, and the overall enrolment rate should be at least equal to the reference measurement.

7 Validation of the Cohort Selection Tool Nona

7.1 Overview

The purpose of the cohort selection tool, Nona and the underlying referenced services is to select cohorts from larger patient databases. This use case is described in detail in [3], Section 3.7. Nona utilizes a wide range of services to assist the user in this task. As such, Nona lags the other tooling regarding the state of development. Only qualitative evaluations are planned for this tool.

7.2 Qualitative evaluation

The qualitative evaluation consists of a set of structured interviews, and a tool review. Three potential sessions are identified: IJB, GBG, and the Evaluation Workshop.

- Current work situation and approach
  - The larger treatment workflow
  - Cohort selection workflow
- Review of the Nona prototype
- Discussion: how will this tool affect work, and how can we optimize this concept

There are no metrics at this stage. The outcome is a report summarizing the findings. This is used to refine the tool, and prepare it to go into qualitative testing. If required, multiple qualitative tests can be performed.

8 Validation of Central Review for Pathology Images tool

8.1 Overview

The purpose of the Central Pathology Review (CPR) platform and its services is to evaluate the process of Reviewing Digital Pathology Images from one or more groups of Specialists (physicians). Related use cases are described in detail in [D 1.5] section 3.9.

8.2 Qualitative evaluation

The qualitative evaluation consists of a set of structured interviews and a tool review. Two sessions are conducted:

- One in Crete with one local specialist (pathologist)
- One at IJB (remotely) with one specialist (pathologist)

These sessions are scheduled as follows:

- Demonstration of the CPR platform from the Reviewer and the Moderator perspective
- Documentation about the platform
- Review of the platform from the Reviewer and the Moderator perspective
- Discussion about:
    - Could CPR platform be used for reviewing digital pathology images?
    - Are CPR modules well defined?
    - How could we optimize CPR platforms modules? (e.g. Image Annotator)

There are no metrics at this stage. The outcome is a report summarizing the findings. This is used to refine the tool, and prepare it to go into quantitative testing. If required, multiple qualitative tests can be performed.

8.3 Quantitative evaluation

The evaluation is conducted according to ISO/IEC 25023. The questionnaires used are listed in Appendix B. Additionally, the SUS (also in Appendix B) is measured and metrics specific to the CPR tool are used (Section 6.8.3). There are two quantitative evaluation scenarios, at the following sites:

- Evaluation Workshop
    - This quantitative evaluation involves at least three users (Pathologists). These will be representative users recruited through the BIG network, but who are not part of the consortium. This test will use the digital pathology imaging data provided by BCTL laboratory at IJB (Section 4.6).
- IJB
    - This quantitative evaluation involves at least three users (Pathologists). These will be representative users recruited locally (IJB). This test will use

the digital pathology imaging data provided by BCTL laboratory at IJB (Section 4.6).

8.3.1 **Evaluation scenarios**

The quantitative evaluation is aimed at pathologists and is different for each of the two major roles involved in the platform.

For each reviewer the test scenario is as follows:

- Consent Form
- Introduction to CPR from reviewer's perspective
- Instructions on how to operate the CPR platform, and time for the user to get familiar with the environment (reviewer's perspective)
- Timed test (test cased described in following section):
- Usability questionnaire and software quality questionnaire afterwards (Questionnaire A.1 and B as in Appendix B)

The test should fit in a 90 minute time window.

For each moderator the test scenario is as follows:

- Consent Form
- Introduction to CPR from moderator's perspective
- Instructions on how to operate the CPR platform, and time for the user to get familiar with the environment (moderator's perspective)
- Guided test (test cased described in following section):
- Usability questionnaire and software quality questionnaire afterwards (Questionnaire A.2 and B as in Appendix B)

The test should fit in a 120 minute time window.

8.3.2 **Timed reviewing tasks**

The evaluation tasks for CPR are derived from UC.CR.3: Create/define a new task, UC.CR.5: Review and annotation process, UC.CR.6: Comparison of the images which were reviewed and resolution of potential conflicts, UC.CR.7: History of the images that have been reviewed. All these Use Cases are listed in [D 1.5]. Use Cases UC.CR.1, UC.CR.2 are excluded since they are part of the security INTEGRATE component. Also UC.CR.4 is merged with UC.CR.3 [D 1.5].

The aim is to simulate, as much as possible, the process of:

a. The definition of a Review Protocol of digital Images by a specialist pathologist, the moderator.
b. The conduction of a Review Protocol where a group of pathologists review and annotate the Review Protocol's images and finally submit their annotations to the platform. In parallel a moderator should administer and manage the whole procedure aiming to the successful completion of its lifecycle.

Since there are two major roles in this procedure, the instructions should be as follows per role:

Reviewer: "As a member of a group of specialist pathologists, you will participate in a Review Protocol for digital pathology images. The objective is for the people in the group to annotate the same set of images and to try to resolve in the platform the annotation discrepancies that will be detected".

Moderator: "As a specialist pathologist, you will be responsible for registering and conducting a Review Protocol for digital pathology images. Your main tasks will be to: A) Register a new protocol that will address specific types of annotations for several pathology images. You will also define the group of specialists that will participate in the review. B) Manage the protocol conduct by checking the reviewers' annotations and the resolution (or lack of resolution) of discrepancies between the different reviewers.**"**

8.3.2.1 Actors Involved

- Reviewer
- Moderator

8.3.2.2 Pre-conditions

- CPR Platform is pre-configured with the appropriate Image Types/Protocol Type Templates
- Authentication Service
- Imaging Service
- Both Reviewer & Moderator can authenticate to the platform and they are authorized to use per role the following services:
    - Moderator: All Services
    - Reviewer: Imaging Service, Notification Services, Protocol Task/Review Services

8.3.2.3 Work Steps

A) Reviewer

Outlined below are the major application tasks from the perspective of a reviewer. A reviewer is the user who is notified of and completes the review tasks generated by a protocol's registration.

- Reviewer is notified with email that a new protocol task has been assigned to him and is pending for Review
    - Reviewer logs into the platform and accesses his personal Inbox (Notification Center). There, he can have an overall view of all tasks assigned to him and he can find the specific task he desires.
    - He then has two options for accessing them:
        - Via the Image Browser, he can view the images (in thumbnail pictures) that are assigned to him for review; and per Image, the tasks and their corresponding statuses.
        - Via the Task Center, he has an aggregated view of all tasks assigned to him and their statuses. These statuses are either:
            - Open/Pending: awaiting assignee actions. Assignee should access this task, review the pathology image and submit annotations to the system.
            - Submitted: assignee annotations are submitted and the moderator's acceptance is pending. No further actions are required.
            - Conflicting: annotations from different reviewers are in conflict with one another. The reviewer has to review and resubmit the conflicting annotations (which can of course stay the same).
            - Closed: no further actions required. Annotations are accepted and the corresponding task is closed.

              According to a task status, the reviewer continues with the required actions.

            - (*) Per task, review process on images consists of two parts:
                - Reviewer must fill out a pre-configured diagnosis form (Required)
                - Reviewer can use the Image Annotator for marking areas of interest on Digital Pathology Image (Optional)
- Reviewer is notified by email that a previously submitted annotation is in conflict with annotation(s) from other reviewer(s).
    - Reviewer logs into the platform and accesses his personal Inbox (Notification Center). He can have an overall view about his tasks and find the corresponding task. He then proceeds as described in the previous paragraph, according to this specific task status. (Conflict).

B) Moderator

A moderator is the user who administers the registration and conduct of a review protocol. Outlined below are the major application tasks that the moderator needs to perform to register and complete a review protocol.

Since the platform is pre-configured with templates for image types (e.g. H&E, various types of immunohistochemistry staining etc.) and Review Protocols, the moderator can:

- Select Digital Pathology Images from the Imaging Repository for use in Review Protocols. During selection process, he assigns a Review Protocol Template/Image Type to each image synchronized/imported in the system.
- Next, register new Review Protocols by selecting:
    - The parameter sets (e.g. ER analysis, PgR Analysis) that needs to be included in the protocol
    - The digital pathology images for which the review process will take place
    - The pathologists/reviewers that will participate in this particular study
    - Since a new review protocol is registered, CPR platform generates relevant tasks and notifies all the participating reviewers
- Administers registered Review Protocols, using the Review Manager module of the CPR platform. Via this module, he can:
    - View all registered Review Protocols and their corresponding statuses (Open/Closed). He can also archive or delete closed Review Protocols.
    - He can select a specific Review Protocol to administer and check reviewers' answers per protocol Image.
    - If so, he can have an aggregated view of all reviewers' answers per image and also the statuses of all Images participating in a Protocol. Depending on reviewers' answers, he can:
        - Mark their tasks (per protocol image) as:
            - Closed (if there are no disagreements between reviewers)
            - Conflicting (if there are disagreements between reviewers)
        - If there are Conflicting answers he can create and send a message to the reviewers with instructions on how to try to resolve the conflict.
        - He can also send reminders (email and internal notifications) for delayed "open" tasks.
        - Finally if all tasks per Protocol images are be closed or all conflicts are considered of no importance, the moderator can mark the entire Protocol as closed", completing is lifecycle.

8.3.3 **Metrics**

We track the following metrics in this evaluation:

- The outcome of the questionnaires as mentioned in Appendix B. The adjusted ISO/IEC questionnaire is split in an end-user questionnaire per user role and a technical part.
    - The technical part (Part C) is only evaluated once by the development team.
    - The remaining parts for the end-user – remaining ISO/IEC questions - are Questionnaires A.1 (Reviewer user role), A.2 (Moderator User role) and Questionnaire B, which is the standard SUS.
- The time performance of the users (reviewers) will be recorded, per task. We derive the overall time per task submission and time per all protocol's tasks submission for each reviewer.
- The time performance of the users (moderators) will be recorded. We record the overall time per Protocol's Image administration (check reviewers' answers and define task statuses for one pathology image) as also the overall time per Protocol administration (check reviewers answers for all Images in the Protocol).

9 Validation of the Analysis Platform Tool

9.1 Overview

The main objective of the Analysis Platform is to provide users with a web-based access to a collaborative, multi-functional and easy-to-use environment for exploiting, analyzing and assessing the quality of large multi-level data. The main goal is to empower the clinician to analyze with ease clinico-genomic data in order to get simple statistics on selected parameters, perform survival analyses, compare regiments in selected cohorts of patient, obtain genomic analysis results, and build a framework enabling the development of multi-scale predictive models of response to therapy and drug efficacy. The platform doesn't require any expertise on using such analysis tools or any software or libraries installed on the user's computer. These requirements are described in detail in (11), Section 3.12.

9.2 Qualitative evaluation

The qualitative evaluation consists of a set of structured interview and a tool review and will be held at IJB. The schedule is as follows:

- Current work situation and approach
- Work with the Analysis platform
- Review of the Analysis platform
- Discussion: how will this tool affect work and how can we optimize this?

The qualitative evaluation process contains no metrics. The outcome is a report summarizing the findings. This is used to refine and improve the tool. If required more qualitative tests can be performed.

9.3 Quantitative evaluation

The Analysis Platform is designed to be used by oncologists with at least little experience in clinical trial data analysis and by researchers/biostatisticians. Quantitative evaluation will be carried out by them according to standard ISO/IEC 25023. There are two different quantitative evaluation sessions for the Analysis platform:

- The Evaluation Workshop, in Crete. The users have been chosen to be outside of the consortium.
- Evaluation sessions at Institut Jules Bordet (IJB). The users will be local staff.

It should be mentioned that the comparison of the Analysis Platform with other, quite similar tools is out-of-scope, since the platform is developed for the specific research questions addressed in the user needs and requirements in (12).

For both the evaluation sessions, TOP trial and SAGE dataset will be used. Additional clinical datasets could also be used.

9.3.1 **Evaluation scenario**

For each participant, the testing scenario is as follows:

1. Consent Form
2. Introduction to the system: describing the platform and its main functionalities, instructions on how to operate with the platform will also be given on this session using e.g., hardcopy manuals, presentations, informational videos by at least one person of the development team. This task will take approximately **15 minutes**.
3. Familiarization with the system: the user will have **15 minutes** to familiarize himself with the platform**.**
4. Guided test: the user will be asked to perform specific tasks, following a specific workflow. Briefly, these tasks include user authentication, data retrieval from the Common Data Model (CDM), performing statistical and predictive analysis, and getting access to the metadata information. During the test, the user screen will be recorded in order to get both quantitative measures (e.g. elapsed time to complete the tasks) and qualitative measures (e.g. tasks for which the user has difficulties to complete or tasks that are completed relatively easily), verifying somehow if the platform "moves to the right direction". This task will take **45 minutes.**
5. Completion of the questionnaires (Appendix C, questionnaire A and B). This task will take **20 minutes**.
6. Discussion: user's overall impression about the tool, remarks, possible weaknesses/ deficiencies, improvements proposals. This task will take **10 minutes.**

After the test the system is reset and the database gets back to its starting state. Any pending time-consuming analysis will not be interrupted, since it is preferable to test the platform under real conditions (i.e. multiple users run multiple type of analysis simultaneously).

9.3.2 **Timed task**

The evaluation task is aimed to check if the user can easily navigate to the platform and perform all the necessary steps. These steps cover almost the entire functionality of the platform. The below task contains only the general steps. Obviously, during the real evaluation sessions, users will be requested to perform *specific* analyses on *specific* cohorts.

The instruction reads as follows:

"As a researcher focusing on the analysis of large, multi-scale clinico-genomic data, your task is to use the INTEGRATE Analysis Platform to get simple statistics on selected parameters, perform survival analyses, compare regiments in selected trials, obtain genomic analysis results, develop multi-scale models for predicting drug response and assess candidate biomarkers in specified cohorts of patients".

9.3.2.1 Actors involved

Researcher (Oncologist, Biostatistician)

9.3.2.2 Pre-conditions

These necessary services for running the INTEGRATE Analysis platform are available:

- Authentication service
- Common Information Access service
- R server
- Report generation service

The user can authenticate himself to the system and is authorized to use all services.

9.3.2.3 Work steps

1. **User authentication and data retrieval**
   - 1.1. The researcher browses to the INTEGRATE Analysis platform site and logs in using his credentials.
   - 1.2. The researcher browses to the “Data Sources” page.
   - 1.3. The researcher interacts and retrieves directly the analysis data from the Common Data Model (CDM).
2. **Statistical Analysis using TOP trial dataset**
   - 2.1. The researcher browses to the “Analytical Tools” page.
   - 2.2. The researcher selects the TOPTRIAL dataset that has been previously retrieved from the CDM and is suitable for statistical analysis. Then he proceeds to the next step where a specified cohort should be build.
   - 2.3. In the next step, the researcher selects the specified analyses.
   - 2.4. In the final step he triggers the execution.
   - 2.5. The researcher views the results using the produced overall diagram.
3. **Statistical Analysis using SAGE dataset**
   - 3.1. The researcher goes back at the appropriate step and selects the SAGE dataset which is eligible for statistical analysis. Then he proceeds to the next step where a specified cohort should be build.
   - 3.2. In the next step, the researcher selects the specified analyses.
   - 3.3. In the final step he triggers the execution.
   - 3.4. The researcher views the results and the selected cohort using the produced overall diagram.
4. **Predictive Modelling using SAGE dataset (Testing process)**
   - 4.1. The researcher browses to the “Predictive Modelling Tools” page.
   - 4.2. The researcher selects the SAGE which is eligible for predictive analysis. Then he proceeds to the next step where a specified cohort should be build.
   - 4.3. In the next step, the researcher selects to predict the clinical outcome of the selected cohort based on an already trained model.
   - 4.4. In the final step he triggers the execution.
5. **Predictive Modelling using SAGE dataset (Complete Study process)**
   - 5.1. Without waiting for the previous analysis to be completed, the researcher goes back at the appropriate step and selects another specified cohort.

5.2. Then he selects to perform a complete predictive analysis study using the selected cohort.
5.3. In the final step he triggers the execution.

6. History navigation

6.1. Without waiting for the previous analysis to be completed, the researcher browses to the “History” page and views the metadata information for all completed and pending analyses.
6.2. The researcher views the cohorts for which the specified analyses have been performed and the generated reports in both html and pdf format.
6.3. He also compares the results from different executed specified models, edits the reports using the basic editing toolbar and saves the changes back to the server.
6.4. Researcher views and edits a specified edited report.
6.5. Finally, as soon as the execution of the predictive models is completed, he views the results and the selected cohort.

9.3.3 **Metrics**

For the validation of the INTEGRATE Analysis platform we track the following outputs:

- The outcome of the questionnaires as mentioned in Appendix C. Questionnaire A and C are the adjusted ISO/IEC questionnaires while the Questionnaire B corresponds to the System Usability Scale (SUS) for global assessment of systems usability and is the same for all the tools. Questionnaires A and B will be filled out by the end-users and the questionnaire C by the development team. Specifically, the questionnaire C tries to measure the degree to which the system is compliant with the initial user needs and requirements while Questionnaire A and B try to measure the added value and the usability of the tool. These three questionnaires are common for the different user groups of the Analysis platform (i.e., oncologists, biostatisticians), since they have similar needs and requirements.
- A small report based on the discussions at the end of test scenarios, reporting the users’ comments and their overall impression about the tool, possible weaknesses/deficiencies, improvements proposals etc. This document could also report the measures extracted from the screen cast, i.e., quantitative measures (e.g. elapsed time to complete the tasks) and qualitative measures (e.g. tasks for which user has difficulties to complete or tasks that completed relatively easily).
- Number of analyses the user managed to execute successfully in the specific time frame.

It should be noticed that the validation/evaluation procedure relies mainly on the usability of the platform. The evaluation of R serve and the evaluation of the analysis algorithms is out-of-scope, since they have already been evaluated and validated.

10 SUMMARY

This document established the requirements of the evaluation, identified the INTEGRATE products to be evaluated, and identified the measures and models for the evaluation. Qualitative evaluation based on structured interviews and quantitative evaluation based on ISO/IEC 25023 will be performed at five different validation sites. The tools that will be evaluated are the Patient Screening tool “Decima”, the Cohort Selection tool “Nona”, the Central Review for Pathology Images tool and the Analysis Platform tool.

11 REFERENCES

1. **SQuaRE, ISO/IEC 25040.** Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) – Evaluation reference model and guide.

2. **IEC.** http://www.iec.ch/. [Online]

3. **2500n, ISO/IEC.** Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Guide to SQuaRE.

4. **25010, ISO/IEC.** Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models.

5. **25020, ISO/IEC.** Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Measurement reference model and guide.

6. **25030, ISO/IEC.** Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Quality requirements.

7. **25040, ISO/IEC.** Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Evaluation process.

8. *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.* **Curtis, Christina, et al.** s.l. : Nature, 2012, Nature.

9. **the-dream-project.org.** [Online] http://www.the-dream-project.org/challenges/sage-bionetworks-dream-breast-cancer-prognosis-challenge.

10. **Deliverable: 6.1 Report on the development environment and on the available test data.**

11. Deliverable 1.5: Consolidation of the User Needs, Use Case Development and Requirements Analysis (final).

12. Deliverable 1.2: Definition of relevant user scenarios based on input form the users.

12 APPENDIX

Appendix A Questionnaires for the qualitative evaluation of the Decima tool

For analysis purposes, all results will be normalised to a scale from 1 to 5. The questions were selected so that the common response to half of them was strong agreement, and to the other half, strong disagreement. This was done in order to prevent response biases caused by respondents not having to think about each statement; by alternating positive and negative items, the respondent has to read each statement and make an effort to think whether they agree or disagree with it. Questionnaire B corresponds to the general SUS score.

Questionnaires A and C correspond to the ISO/IEC standard, and are split in questions that are aimed at the end users, and questions for the development team.

Questionnaire A	**Rating**				
	Strongly Disagree		**Neutral**		**Strongly Agree**
	1	**2**	**3**	**4**	**5**
Using the system, I can determine whether there are suitable trials for a patient.					
The system responds quickly enough.					
The delays for processing of data seemed reasonable.					
I was not distracted or interrupted in my task due to the system communication with external components (3rd party).					
The interaction with the system was pleasant.					
The look of the application seemed professional and appropriate					

Few or no errors occurred while I used the system					
If an error occurred the system recovered and informed me appropriately					
The login and authentication seems to be used at the appropriate level for access to the clinical data					
The login process is secure					
The software allows me to do everything I would expect or want to do to determine which trials are suitable for patients					
The software reduces the amount of time I need to spend to find suitable trials for a patient					
I think this tool would be effective in practice to find suitable clinical trials for a patient					

Questionnaire B	Rating				
	Strongly Disagree		Neutral		Strongly Agree
	1	2	3	4	5
I found the system unnecessarily complex					
I thought the system was easy to use					
I think that I would need the support of a technical person to be able to use this system					
I found the various functions in this system were well integrated					
I thought there was too much inconsistency in this system					
I would imagine that most people would learn to use this system very quickly					
I found the system very cumbersome to use					
I felt very confident using the system					
I needed to learn a lot of things before I could get going with this system					

Questionnaire C

Functionality

Can the software find suitable trials for a patient?

- ☐ Yes
- ☐ No
- ☐ Not Applicable

Is the result as expected?

- ☐ Yes
- ☐ No
- ☐ Not Applicable

Can the system interact with the trial metadata repository?

- ☐ Yes
- ☐ No
- ☐ Not Applicable

Can the system access the relevant patient data in the CDM?

- ☐ Yes
- ☐ No
- ☐ Not Applicable

Efficiency

How quickly does the system respond?

- ☐ **1** Delays are substantially and require explicit planning of issued commands
- ☐ **2** Delay breaks the person's normal workflow
- ☐ **3** Noticeable delay, forcing explicit pauses in a person's workflow
- ☐ **4** Noticeable delay, but not disturbing
- ☐ **5** No perceived delay

Compatibility

Can the software run in a desktop environment?

- ☐ **1** Application can only run by itself in its environment
- ☐ **2**
- ☐ **3** Application can only run when there are a limited number of other apps running
- ☐ **4**
- ☐ **5** Application can run under any other parallel use

Usability

Does the user comprehend how to use the system easily?

- ☐ **1** Training needed beforehand
- ☐ **2**
- ☐ **3** Can readily find help online or elsewhere
- ☐ **4**
- ☐ **5** No explanation needed

Can the user learn to use the system easily?

- ☐ **1** Serious investment of time needed to learn
- ☐ **2** Some time needed to learn where to find everything and conventions of the software
- ☐ **3** A few minor or less used features take a bit of time to learn
- ☐ **4** Shown once and then user has learnt how to use
- ☐ **5** Walk up and use, and no learning needed

Can the user use the system without much effort?

- ☐ **1** Seriously investment of time to get any result
- ☐ **2**
- ☐ **3** Some time and effort needed
- ☐ **4**
- ☐ **5** Minimal effort

Reliability

Have most of the faults in the software been eliminated over time?	☐ **1**	User continuously needs to work around issues
	☐ **2**	
	☐ **3**	User can use the system but on occasion a fault might occur
	☐ **4**	
	☐ **5**	System appears perfect to the user
Is the software capable of handling errors?	☐ **1**	Crashes and no feedback to user
	☐ **2**	
	☐ **3**	Might crash occasionally, but in general gives the user appropriate feedback
	☐ **4**	
	☐ **5**	System never crashes
Can the software resume working & restore lost data after failure?	☐ **1**	After failure end-user support is necessary
	☐ **2**	
	☐ **3**	After failure a restart is required, but then works
	☐ **4**	
	☐ **5**	System will not stop on failures, but handles them.

Security

Does the system provide identification access wherever is needed?

- ☐ Yes
- ☐ No
- ☐ Not Applicable

Is data accessible only to authorized users?

- ☐ Yes
- ☐ No
- ☐ Not Applicable

Maintainability

Can faults be easily diagnosed?

- ☐ **1** Within days
- ☐ **2**
- ☐ **3** Within hours
- ☐ **4**
- ☐ **5** Within half an hour

Can the software be easily modified?	☐ **1**	Time lost on average to redesign is on average 200% or more
	☐ **2**	
	☐ **3**	Time lost on average to redesign is on average below 50%
	☐ **4**	
	☐ **5**	Time lost on average to redesign is on average below 10%
Can the software be tested easily?	☐ **1**	Manual test
	☐ **2**	
	☐ **3**	Partially Automated
	☐ **4**	
	☐ **5**	Fully automated test

Portability

Can the software be moved to other environments?	☐ **1**	Not at all
	☐ **2**	Limited to a single environment
	☐ **3**	Multiple Environments
	☐ **4**	Runs multiple devices, environments
	☐ **5**	Runs on any device, anywhere, and adjusts to its new context

Can the software be installed easily?

- ☐ **1** Needs an expert
- ☐ **2** Needs a person familiar with the system
- ☐ **3** Needs generic technical support
- ☐ **4** Can be done by end user + manual / skilled end user
- ☐ **5** Can be done by an unskilled end user

Quality in Use

How accurate and complete is the software for the intended use?

- ☐ **1** Not at all
- ☐ **2**
- ☐ **3** Somewhat
- ☐ **4**
- ☐ **5** Completely

Does the software reduce the amount of time needed for the intended goal?

- ☐ **1** Not at all, or increases time needed
- ☐ **2** Minor time savings
- ☐ **3** Some time savings, enough to be beneficial
- ☐ **4** Moderate time savings
- ☐ **5** Significant time savings

Is there any chance of harming people by using the software in the intended context of use?	☐ **1**	Highly likely that people will be harmed
	☐ **2**	
	☐ **3**	Moderate chance of harm
	☐ **4**	
	☐ **5**	Negligible chance of harm

Appendix B Questionnaires for the qualitative evaluation of the Central Review for Pathology Images tool

Questionnaire A.1	Rating				
	Strongly Disagree		Neutral		Strongly Agree
	1	2	3	4	5
CPR platform corresponds in a timely manner (quickly enough)					
I was not distracted or interrupted in my task due to the system communication with external components (3rd party).					
The look of CPR platform is professional and clean					
Your experience with the platform was satisfying					
I did not experience, or I experience few, bugs and system errors during CPR platform usage					
If an error occurred during system usage, CPR platform recovered and I manage to finish actions with no corruptions					
User roles are set and working, providing access to the appropriate views of the application					
Login process is secure and sufficient					
Using the platform I can review Digital Pathology Images and submit my answers to the system					
Using the Image Annotator I can easily annotate areas on Digital Pathology Images					
Image Annotator provides enough and easy to use UI tools for annotating digital images					
The platform reduces the time for conducting Review Studies on Clinical					

Images					
The platform could enhance the overall process of conducting Review of Clinical Images for patient trial selection					

Questionnaire A.2	Rating				
	Strongly Disagree		Neutral		Strongly Agree
	1	2	3	4	5
CPR platform corresponds in a timely manner (quickly enough)					
I was not distracted or interrupted in my task due to the system communication with external components (3rd party).					
The look of CPR platform is professional and clean					
Your experience with the platform was satisfying					
I did not experience, or I experience few, bugs and system errors during CPR platform usage					
If an error occurred during system usage, CPR platform recovered and I manage to finish actions with no corruptions					
User roles are set and working, providing access to the appropriate views of the application					
Login process is secure and sufficient					
CPR can access Image DW and fetch available images for use					
Using the platform I was able to register a new review protocol					
Registration of new review protocol is an easy process for Moderator					
Using the platform I was able to administer existing review protocols					

Protocol Manager is suitable, easy to use for managing existing protocols					
The platform reduces the time for conducting Review Studies on Clinical Images					
The platform could enhance the overall process of conducting Review of Clinical Images for patient trial selection					

Questionnaire B	**Rating**				
	Strongly Disagree		**Neutral**		**Strongly Agree**
	1	**2**	**3**	**4**	**5**
I found the system unnecessarily complex					
I thought the system was easy to use					
I think that I would need the support of a technical person to be able to use this system					
I found the various functions in this system were well integrated					
I thought there was too much inconsistency in this system					
I would imagine that most people would learn to use this system very quickly					
I found the system very cumbersome to use					
I felt very confident using the system					
I needed to learn a lot of things before I could get going with this system					

Questionnaire C

Functionality

System is able to communicate with DW, access available images, select & sync them internally and assign image type	☐	Yes
	☐	No
	☐	Not Applicable
Moderator is able to register new review protocols	☐	Yes
	☐	No
	☐	Not Applicable
Moderator is able to administer registered review protocols (check tasks, change tasks statuses, close protocols, mark as conflicting)	☐	Yes
	☐	No
	☐	Not Applicable
Reviewer is able to participate in Review Protocols. Acquire his tasks, submit back his answers, Check tasks statuses	☐	Yes
	☐	No
	☐	Not Applicable
Reviewer is able to annotate pathology image	☐	Yes
	☐	No
	☐	Not Applicable

Performance

System corresponds in a timely manner	☐	Delays break user's normal workflow
	☐	Noticeable delays, forcing explicit pauses in user's normal workflow
	☐	Noticeable delays, but not disturbing
	☐	No delays
Did you notice any performance issues (overall) when simultaneous users used the platform	☐	Delays break user's normal workflow
	☐	Noticeable delays, forcing explicit pauses in user's normal workflow
	☐	Noticeable delays, not affecting user's experience
	☐	No delays

Compatibility

Image tiles stored in Imaging Server are fetch without losses or performance degradation	☐	No
	☐	In many cases image tiles seem to be missing
		In a few cases image tiles seem to be missing
	☐	Yes
System is able to fetch external information (sync images, fetch image tiles) without service corruption or performance degradation (SOAP, HTTPS)	☐	No
	☐	In many cases, communication with external resources seem to break down

	☐		In some cases, communication with external resources seem to break down
	☐		Yes
System can be accessed and used by a typical desktop web interface setup (browsers IE, Mozilla, Chrome)	☐		No
	☐		It can be accessed and used by limited web interface setups
	☐		It can be accessed and used by the most web interface setups
	☐		Yes

Usability

Reviewer can understand how to use the system easily	☐	**1**	Training needed beforehand
	☐	**2**	
	☐	**3**	Can readily find help online or elsewhere
	☐	**4**	
	☐	**5**	No explanation needed
Moderator can understand how to register new protocol easily	☐	**1**	Training needed beforehand
	☐	**2**	
	☐	**3**	Can readily find help online or elsewhere
	☐	**4**	
	☐	**5**	No explanation needed

Moderator can understand how to manage a registered review protocol easily	☐	**1**	Training needed beforehand
	☐	**2**	
	☐	**3**	Can readily find help online or elsewhere
	☐	**4**	
	☐	**5**	No explanation needed
Reviewer can learn to use CPR platform easily	☐	**1**	Serious investment of time needed to learn
	☐	**2**	Some time needed to learn where to find everything and conventions of the software
	☐	**3**	A few minor or less used features take a bit of time to learn
	☐	**4**	Shown once and then user has learnt how to use
	☐	**5**	Walk up and use, and no learning needed
Moderator can learn how to register new protocols easily	☐	**1**	Serious investment of time needed to learn
	☐	**2**	Some time needed to learn where to find everything and conventions of the software
	☐	**3**	A few minor or less used features take a bit of time to learn

	☐	**4**	Shown once and then user has learnt how to use
	☐	**5**	Walk up and use, and no learning needed
Moderator can learn how to manage existing protocols easily	☐	**1**	Serious investment of time needed to learn
	☐	**2**	Some time needed to learn where to find everything and conventions of the software
	☐	**3**	A few minor or less used features take a bit of time to learn
	☐	**4**	Shown once and then user has learnt how to use
	☐	**5**	Walk up and use, and no learning needed
Reviewer can use CPR platform without much effort	☐	**1**	Seriously investment of time to get any result
	☐	**2**	
	☐	**3**	Some time and effort needed
	☐	**4**	
	☐	**5**	Minimal effort
Moderator can use CPR platform without much effort	☐	**1**	Seriously investment of time to get any result
	☐	**2**	
	☐	**3**	Some time and effort needed

	☐	**4**	
	☐	**5**	Minimal effort
CPR platform protects user from making errors. Its UI is simple and its elements usage is clear	☐	**1**	Yes, interface does not permit any illegal actions. Also data input validation has been implemented so as the user can enter only valid data values
	☐	**2**	
	☐	**3**	In some cases there are elements that permit illegal operations. Data input validation has not been implemented in all possible cases
	☐	**4**	
	☐	**5**	No, interface permits user actions that can result in system breaks. Also data input validation does not work or has not been implemented

Reliability

Most of the faults in the software been eliminated over time	☐	**1**	User continuously needs to work around issues
	☐	**2**	
	☐	**3**	User can use the system but on occasion a fault might occur
	☐	**4**	
	☐	**5**	System appears perfect to the user

Is the software capable of handling errors?	☐	**1**	Crashes and no feedback to user
	☐	**2**	
	☐	**3**	Might crash occasionally, but in general gives the user appropriate feedback
	☐	**4**	
	☐	**5**	System never crashes
Can the software resume working & restore lost data after failure?	☐	**1**	After failure end-user support is necessary
	☐	**2**	
	☐	**3**	After failure a restart is required, but then works
	☐	**4**	
	☐	**5**	System will not stop on failures, but handles them.

Security

Access to the CPR is fully controllable. No unauthorized access is permitted	☐	Yes
	☐	No
	☐	Not Applicable
System deny any user actions resulting in data corruption or bad data insertion into the system	☐	Yes
	☐	No
	☐	Not Applicable

Maintainability

Can faults be easily diagnosed?	☐	**1**	Within days
	☐	**2**	
	☐	**3**	Within hours
	☐	**4**	
	☐	**5**	Within half an hour
Can the software be easily modified?	☐	**1**	Time lost on average to redesign is on average 200% or more
	☐	**2**	
	☐	**3**	Time lost on average to redesign is on average below 50%
	☐	**4**	
	☐	**5**	Time lost on average to redesign is on average below 10%
Can the software be tested easily	☐	**1**	Manual test
	☐	**2**	
	☐	**3**	Partially Automated
	☐	**4**	
	☐	**5**	Fully Automated test

Portability

Can the software be moved to other environments?	☐	**1**	Not at all
	☐	**2**	Limited to a single environment
	☐	**3**	Multiple Environments
	☐	**4**	Runs multiple devices, environments
	☐	**5**	Runs on any device, anywhere, and adjusts to its new context
Can the software be installed easily?	☐	**1**	Needs an expert
	☐	**2**	Needs a person familiar with the system
	☐	**3**	Needs generic technical support
	☐	**4**	Can be done by end user & manual / skilled end user
	☐	**5**	Can be done by an unskilled end user
Can the software or its modules be replaced easily?	☐		Yes
	☐		Some of its modules can be replaced
	☐		No

Quality in Use

How accurate and complete is the software for the intended use?	☐	**1**	Not at all
	☐	**2**	
	☐	**3**	Somewhat

	☐	**4**	
	☐	**5**	Completely
Does the platform reduces the time for conducting Review Studies on Clinical Images	☐	**1**	Not at all, or increases time needed
	☐	**2**	Minor time savings
	☐	**3**	Some time savings, enough to be beneficial
	☐	**4**	Moderate time savings
	☐	**5**	Significant time savings

Appendix C Questionnaires for the qualitative evaluation of the Analysis Platform tool

Questionnaire A	Rating				
	Strongly Disagree		Neutral		Strongly Agree
	1	2	3	4	5
Using the INTEGRATE Analysis platform I can run statistical analysis.					
Using the INTEGRATE Analysis platform I can run predictive analysis.					
The platform responds quickly enough. Please do not take into account analyses that are time/memory-consuming anyway.					
The delays for processing the data seemed reasonable.					
I was not distracted or interrupted in my task due to the system communication with external components (3rd party).					
The interaction with the platform was pleasant and friendly.					
The look of the INTEGRATE Analysis platform seamed professional and appropriate.					
I felt comfortable in using this platform.					
The system was learned easily.					
The procedures that must be followed to run statistical or predictive tools are clear enough.					
Few or no errors/bugs occurred while I used the system.					
If an error occurred the system recovered and informed me appropriately.					
The login and authentication seems to be used at the appropriate level for access to the data.					

The login process is secure.					
The platform is accurate and complete for intended use.					
The statistical tools provided by the platform are relevant to clinical trial data analysis.					
The predictive modelling tools provided by the platform are relevant to clinical trial data analysis.					
The platform meets my expectations for clinical trial data analysis.					
The platform meets my ideal for clinical trial data analysis.					
The platform helps user to finish their tasks quickly.					
The platform helps user to finish their tasks efficiently.					
I can run statistical analysis easily.					
I can run predictive analysis easily.					
The platform reduces the amount of time I need to spend to execute statistical analyses.					
The platform reduces the amount of time I need to spend to execute predictive analyses.					
I think this platform would be effective in practice.					
I think that I would like to use this system frequently.					
In general, I am very pleased with the ease with which I can use this tool.					
In general I am satisfied using this tool.					

Questionnaire B	Rating				
	Strongly Disagree		Neutral		Strongly Agree
	1	2	3	4	5
I think that I would like to use this system frequently.					
I found the system unnecessarily complex					
I thought the system was easy to use					
I think that I would need the support of a technical person to be able to use this system					
I found the various functions in this system were well integrated					
I thought there was too much inconsistency in this system					
I would imagine that most people would learn to use this system very quickly					
I found the system very cumbersome to use					
I felt very confident using the system					
I needed to learn a lot of things before I could get going with this system					

Questionnaire C

Functionality

Does the platform implement the research questions addressed by the initial user specifications?	☐	**1**	The platform is in totally another direction regarding the initial user specifications.
	☐	**2**	
	☐	**3**	The platform implements some of the research questions addressed by the initial user specifications.
	☐	**4**	
	☐	**5**	The platform is totally compatible with the initial user specifications.
Are the results from the execution of statistical analyses as expected?	☐		Yes
	☐		No
	☐		Not Applicable
Are the results from the execution of predictive analyses as expected?	☐		Yes
	☐		No
	☐		Not Applicable
Can the platform access the relevant analysis data from the CDM?	☐		Yes
	☐		No
	☐		Not Applicable
Can the platform interact with the	☐		Yes

local metadata repository?	☐		No
	☐		Not Applicable

Efficiency

How quickly does the system respond? Please do not take into account analyses that are time/memory-consuming anyway.	☐	**1**	Delays are substantially and require explicit planning of issued commands
	☐	**2**	Delay breaks the person's normal workflow
	☐	**3**	Noticeable delay, forcing explicit pauses in a person's workflow
	☐	**4**	Noticeable delay but not disturbing
	☐	**5**	No delays
Does the platform require many resources (CPU time, memory space) to perform? Please do not take into account analyses that are time/memory-consuming anyway.	☐	**1**	The platform requires too many resources to perform.
	☐	**2**	
	☐	**3**	The platform requires sufficient resources to perform.
	☐	**4**	
	☐	**5**	The platform requires acceptable resources to perform.
Can the platform be accessed by several users simultaneously without any impact on its efficiency?	☐		Yes
	☐		No
	☐		Not applicable

Compatibility

Can the platform perform efficiently while sharing common resources with other products, without detrimental impacts on these products?	☐	**1**	Using the platform, no other product can be used.
	☐	**2**	
	☐	**3**	Using the platform, delays breaks are observed to other tools.
	☐	**4**	
	☐	**5**	Using the platform, no detrimental impacts are observed to other tools.
Can the platform interact with the CDM easily?	☐		Yes
	☐		No
	☐		Not applicable
Can the platform interact with the R serve easily?	☐		Yes
	☐		No
	☐		Not applicable
Can the platform interact with the Latex service easily?	☐		Yes
	☐		No
	☐		Not applicable
Can the platform interact with the cohort selection tool "Nona" easily?	☐		Yes
	☐		No
	☐		Not applicable

Can the platform be accessed by the most web browsers (e.g. Mozilla, Chrome, Explorer etc).	☐		Yes
	☐		No
	☐		Not applicable

Usability

Does the user comprehend how to use the system easily?	☐	**1**	Training needed beforehand.
	☐	**2**	
	☐	**3**	Can readily find help online or elsewhere
	☐	**4**	
	☐	**5**	No explanation needed
Can the user learn to use the system easily?	☐	**1**	Serious investment of time needed to learn
	☐	**2**	Some time needed to learn where to find everything and conventions of the software
	☐	**3**	A few minor or less used features take a bit of time to learn.
	☐	**4**	Shown once and then user has learnt how to use
	☐	**5**	No learning needed
Can the platform execute a statistical analysis tool easily?	☐	**1**	Serious investment of time needed to learn
	☐	**2**	
	☐	**3**	Some time and effort needed
	☐	**4**	

	☐	**5**	Minimal effort
Can the user execute a predictive modelling tool easily?	☐	**1**	Serious investment of time needed to learn
	☐	**2**	
	☐	**3**	Some time and effort needed
	☐	**4**	
	☐	**5**	Minimal effort
Are the similar operation of the platform carried out consistently (e.g. statistical analysis and predictive analysis)?	☐		Yes
	☐		No
	☐		Not applicable
Are the messages of the platform understood easily?	☐		Yes
	☐		No
	☐		Not applicable
Does the platform offer the possibility of customizing procedures when it is necessary?	☐	**1**	No
	☐	**2**	
	☐	**3**	Some of the procedures could be customized.
	☐	**4**	
	☐	**5**	All the necessary procedures are customized appropriately

Does the platform offer appropriate error avoidance capability?	☐	**1**	No
	☐	**2**	
	☐	**3**	The error avoidance capability is not complete
	☐	**4**	
	☐	**5**	Yes, the platform does not permit illegal actions
Does the platform offer appropriate input validation capability?	☐	**1**	No
	☐	**2**	
	☐	**3**	The input validation capability is not complete
	☐	**4**	
	☐	**5**	Yes, the platform does not permit illegal inputs

Reliability

Have most of the faults of the platform been eliminated over the time?	☐	**1**	User continuously needs to work around issues
	☐	**2**	
	☐	**3**	User can use the system but on occasion a fault might occur
	☐	**4**	
	☐	**5**	System appears perfect to the user
Is the software capable of handling errors?	☐	**1**	The platform crashes and no feedback is given to user
	☐	**2**	

	☐	**3**	Might crash occasionally, but in general gives user appropriate feedback
	☐	**4**	
	☐	**5**	System never crashes
Can the software resume working and restore lost data after failure?	☐	**1**	After failure end-user support is necessary
	☐	**2**	
	☐	**3**	After failure a restart is required but then works
	☐	**4**	
	☐	**5**	System will not stop on failures, but handles them

Security

Does the platform provide identification access wherever is needed?	☐	Yes
	☐	No
	☐	Not Applicable
Is data accessible only to authorized users?	☐	Yes
	☐	No
	☐	Not Applicable

Maintainability

Can faults be easily diagnosed?	☐	**1**	Within days

	☐	**2**	
	☐	**3**	Within hours
	☐	**4**	
	☐	**5**	Within half an hour
Can the platform be easily modified by the intended maintainers?	☐	**1**	Time lost to modify is on average 200% or more
	☐	**2**	
	☐	**3**	Time lost to modify is on average below 50%
	☐	**4**	
	☐	**5**	Time lost to modify is on average below 10%
Can the platform be easily tested?	☐	**1**	Manual test
	☐	**2**	
	☐	**3**	Partially automated
	☐	**4**	
	☐	**5**	Fully automated

Portability

Can the software be moved to other environments?	☐	**1**	Not at all
	☐	**2**	Limited to a single environment
	☐	**3**	Multiple environments

	☐	**4**	Runs on multiple devices, environments
	☐	**5**	Runs on any device, anywhere and adjust to its new context
Can the software be installed easily?	☐	**1**	Needs an expert
	☐	**2**	Needs a person familiar with the system
	☐	**3**	Needs generic technical support
	☐	**4**	Can be done by end-user & manual/skilled end-user
	☐	**5**	Can be done by an unskilled end-user

Quality in Use

How accurate and complete is the software for the intended use?	☐	**1**	Not at all
	☐	**2**	
	☐	**3**	Somewhat
	☐	**4**	
	☐	**5**	Completely
Does the platform reduce the amount of time needed for intended goal?	☐	**1**	Not at all, or increases time needed
	☐	**2**	Minor time savings
	☐	**3**	Some time savings, enough to be beneficial
	☐	**4**	Moderate time savings
	☐	**5**	Significant time savings

Appendix D IJB breast dataset structure

Name	Type	Description
DTTOUR	Date/Hour	Multidisciplinary date
NRDOSS	Integer	Patient Number (crypted)
DTNAPT	Date/Hour	Birth date (with day replaced by 1)
DEBHOSP	Date/Hour	Hospitalisation start date
FINHOSP	Date/Hour	Hospitalisation end date
DATECONS	Date/Hour	Consultation date
ATCSEIN	Boolean	Breast Cancer Antecedant
ATCOVAIR	Boolean	Ovarian Cancer Antecedant
ATCCANC	Boolean	Other Cancer Antecedant
ATCHTA	Boolean	Hypertension antecedant
ATCCAR	Boolean	Cardiac antecedant
ATCVASC	Boolean	Vascular problem antecedant
ATCDIAB	Boolean	Diabete antecedant
ATCDEPRESS	Boolean	Depression antecedant
ATCCATAR	Boolean	Cataracte antecedant
ATCAUTR	Boolean	Other Antecedant
ATCANNEXUNIT	Boolean	Unilateral annexetomy antecedant
ATCANNEXBILAT	Boolean	Bilateral annexetomy antecedant
ATCHYSTER	Boolean	Hysterectomy antecedant
ANTFAM	No/Yes/Unknown	Familial antecedant
FAMLIEN	Code	Familial link
MENARCH	Integer	Menarchal age
MENSTA	Integer	Menstrual age
AGMENOP	Integer	Menopausal age
TRTHORM	Boolean	Hormonal treatment
FSH	Real number	FSH level
LH	Real number	LH level
DATFSHLH	Date/Hour	FSH-LH level
GESTITE	Integer	Gestity (number of fœtus carried)
PARITE	Integer	Parity (number of baby delivered)

LATER	Code	laterality of the tumor
DIAGHISTO	Code	Trucut diagnosis
DIAGMAMM	Code	Mommotome diagnosis
DIAGBIOPS	Code	Biopsy diagnosis
CA153	Real number	Ca 15-3
DATECH	Date/Hour	Intervention date
TYPECH	Code	Intervention type
RECONS	Boolean	Reconstruction
VERIDEX	Code	Veridex metastasis PCR identification pos/neg
TOPOCH	Code	Topology of the tumor
HISTCH	Code	Histology of the tumor
NBFOYERS	Integer	Number of tumor sites
TAIL	Integer	Total diameter
TAIL1	Integer	1st lesion diameter
TAIL2	Integer	2nd lesion diameter
TAIL3	Integer	3rd lesion diameter
SBR	Code	SBR grade
GRAD	Code	Grade
IPVN	Code	Van Neuys pronostic index
EMBV	Boolean	Vascular emboly
EMBL	Boolean	Lymphatic emboly
EVAX	Code	Axilar voiding type
GGSPREL	No/Yes/Unknown	sentinel lymph node excised?
NGGSPOS	Integer	Number of sentinel lymph nodes positive
NGGSPREL	Integer	Number of sentinel lymph nodes excised
GGSRESU	Integer	sentinel lymph node result
GGNSPREL	No/Yes/Unknown	control lymph node excised?
NGGNSPOS	Integer	Number of control lymph node positive
NGGNSPREL	Integer	Number of control lymph node excised
GGNSRESU	Integer	control lymph node result
GGEFCAP	No/Yes/Unknown	capsular effraction?
SGGEFCAP	No/Yes/Unknown	capsular effraction of sentinel lymph nodes?

ER	Code	ER status
PgR	Code	PgR status
CERB	Code	c-erB2 (HER2) status
DTFISH	Date/Heure	Date of neu FISH
AJOUDAT	Date/Heure	Date of data collection
TRT1	Code	Treatment 1
TTTPL1	Code	Proposed treatment 1
TRT2	Code	Treatment 2
TTTPL2	Code	Proposed treatment 2
TRT3	Code	Treatment 3
TTTPL3	Code	Proposed treatment 3
TRT4	Code	Treatment 4
TTTPL4	Code	Proposed treatment 4
TRTM1	Code	Treatment 1 MINDACT
TTTPLM1	Code	Proposed treatment 1 MINDACT
TRTM2	Code	Treatment 2 MINDACT
TTTPLM2	Code	Proposed treatment 2 MINDACT
TRTM3	Code	Treatment 3 MINDACT
TTTPLM3	Code	Proposed treatment 3 MINDACT
TRTM4	Code	Treatment 4 MINDACT
TTTPLM4	Code	Proposed treatment 4 MINDACT