Requirements analysis and consolidation of the NeoBIG scenario

Project Number:	FP6-2005-IST-026996
Deliverable id:	D 5.6
Deliverable name:	Requirements analysis and consolidation of the NeoBIG scenario
Submission Date:	25/11/2009

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	ACGT
Project Full Name:	Advancing Clinico-Genomic Clinical Trials on Cancer: Open Grid Services for improving Medical Knowledge Discovery
Document id:	D 5.6
Document name:	Requirements analysis and consolidation of the NeoBIG scenario
Document type (PU, INT, RE)	RE
Version:	0.13
Submission date:	NN/NN/NNNN
Editor: Organisation: Email:	Anca Bucur Philips Research anca.bucur@philips.com

Document type PU = public, INT = internal, RE = restricted

ABSTRACT: The NeoBIG program is a research program led by Breast International Group, and aims to accelerate drug and biomarker development in early breast cancer, recognizing that the current drug development process is suboptimal and aims to improve the results of clinical trials. . A durable, multidimensional translational research structure supporting neo-adjuvant trials will be build, sharing strategies, expertise, technologies, methodologies and protocols. In addition this will provide a strong foundation for future adjuvant trials in breast cancer (and research in other cancers). The program should result in a lasting bioinformatics platform for collaboration between cancer research institutes in Europe. This document carries out an initial requirements collection and analysis for the data sharing platform meant to support the NeoBIG research program of the Breast International Group. We have also briefly described several scenarios concerning the expected uses of the platform.

KEYWORD LIST: NeoBIG, clinical trials, databases, data integration, data access services

MODIFICATION CONTROL			
Version	Date	Status	Author
0.13	23/11/2009	Draft	A. Bucur J. van Leeuwen

List of Contributors

- Anca Bucur, Philips Research
- Christine Desmedt, Institut Jules Bordet
- Jasper van Leeuwen, Philips Research

Contents

1 Introduction

This investigation has been triggered by a discussion in the ACGT consortium of whether and how the expertise, and potentially also the tools, developed in ACGT could be used to support a large real-life multi-centric clinical trails programme, such as NeoBIG, the new research programme of the Breast International Group. Our focus was on the IT needs of such a research programme, specifically with respect to secure privacy-preserving data sharing as these are issues at the core of ACGT. We have tried to answer these questions by first collecting and analyzing the requirements of BIG concerning the data sharing platform needed to support their future clinical trials, and based on that briefly evaluating potential alternatives in which ACGT could support this programme.

1.1 Scope

This report addresses the requirements collection and the scenario refinement concerning the building of a data sharing platform to support the NeoBIG programme of the Breast International Group. Part of this investigation we have also considered the suitability of the ACGT tools and infrastructure for this task, as well as of relevant existing caBIG tools.

Based on the requirements for the data sharing platform collected during our discussions with the users we also propose several clinical use scenarios. Due to the limited scope of this investigation, these scenarios are only preliminary and they need to be further elaborated and refined in future work.

1.2 Structure

This document is structured as follows. We first briefly explain our approach and the context of this investigation. Next we describe the clinical programme in the context of which the need for the development of a data sharing platform has been identified.

Section 2 focuses on the main requirements for the neoBIG platform identified together with BIG and the BrEAST data centre. We summarize our main learning points and the future perspectives of this work in the conclusions section.

1.3 Approach

The requirements collection was carried out based on discussions and interviews with the end users from the Breast International Group. The aspects considered relevant are presented in Section 2. In our interviews we have covered all the questions presented in the diagram in Section2, addressing technical and clinical aspects, but also practical issues related to maintenance and sustainability.

Based on the interviews and discussions we have also elaborated some generic use scenarios which can be further used for refining the requirements and elaborating the detailed use cases.

Next to ACGT solutions we have also considered the tools and services available in the wider research community, specifically those emerging from the caBIG efforts.

1.4 NeoBIG

The NeoBIG program is a research program led by Breast International Group (BIG, see [2]). BIG is an international non-profit organisation for academic breast cancer research groups from around the world. BIG comprises 45 members and spans the world (Europe, Canada, Latin America, Asia and Australasia). BIG coordinates the BIG trials, with 3000 specialised hospitals and research centres contributing around the globe. Currently, 76000 patients have been recruited for BIG clinical trials. BIG mission is to facilitate breast cancer research internationally in order to reduce wasteful duplication of effort, advance knowledge in the field and optimally serve those affected by the disease. In line of this mission BIG has defined the NeoBIG program. NeoBIG aims to accelerate drug and biomarker development in early breast cancer, recognizing that the current drug development process is suboptimal and aims to improve the results of clinical trials in various ways.

NeoBIG pursues an integral approach to realise its aims. A durable, multidimensional translational research structure supporting neo-adjuvant trials will be build by sharing strategies, expertise, technologies, methodologies and protocols. In addition this will provide a strong foundation for future adjuvant trials in breast cancer (and research in other cancers).

Scientifically, the platform is lead by the Core Institutions, which are leading academic centres with neoadjuvant expertise. The focus of the platform is on controlled sharing of data obtained by running clinical trials. It is a multi-partner collaboration with various partners contributing to the data collection and contributing to the data analysis. (e.g. pharmaceutical, imaging and bio-diagnostic companies). Data will be securely shared between organizations and the platform provides access control. For instance, consortium partners can access data of the control arms of clinical trials. The platform supports tissue bio-banking, which will leverage future translational research. Finally the platform will harmonise technical, legal and contractual procedures, streamlining the clinical trial process.

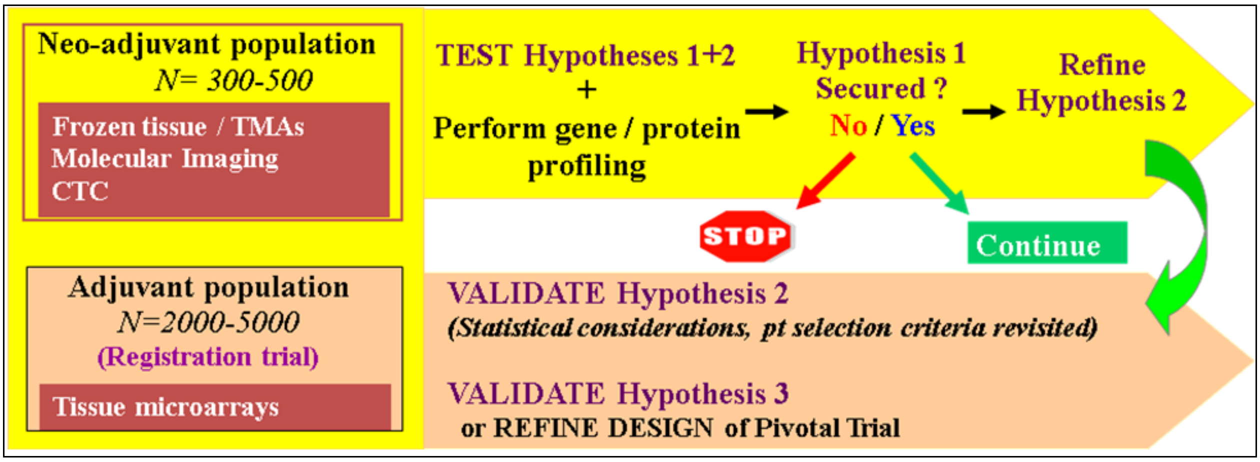


Figure 1 - New model of early breast cancer trials

A new model trial design (see [1]) is proposed to speed up the evaluation of potential new drugs and biomarkers. NeoBIG trials will use selected patient populations (based on molecular subtype) in a Neo-adjuvant setting, such that surrogate endpoints will enable quick go-no-go decisions. Trials will be based on an integrated biomarker program (using gene expression signatures, circulating tumour cells, proteomics, genetics, and functional imaging), and will contain a central pathology review; all resulting in a pipeline of targeted agents. In the new model of early breast cancer trials, the trial (see Figure 1) will consist of two strategies comparing the standard of care (strategy A) versus the alternative strategy (strategy B). It will to try to secure hypothesis 1 "*Benefit of at least 1 surrogate endpoint in B versus A in neoadjuvant setting*" for a Neo-adjuvant population (300 to 500 patients). If hypothesis 1 is secured, the trial will focus on an adjuvant population (2000 to 5000 patients) in order to validate two hypothesis; "*a molecular markers profile triggers this benefit (hypothesis 2)"*, and *"There is a clinical benefit of strategy B over strategy A in early breast cancer"*. The overall result of the new model should lead to traditional "adjuvant" trials - where new drugs need to be validated – with a greater chance of success due to the strong biological hypothesis.

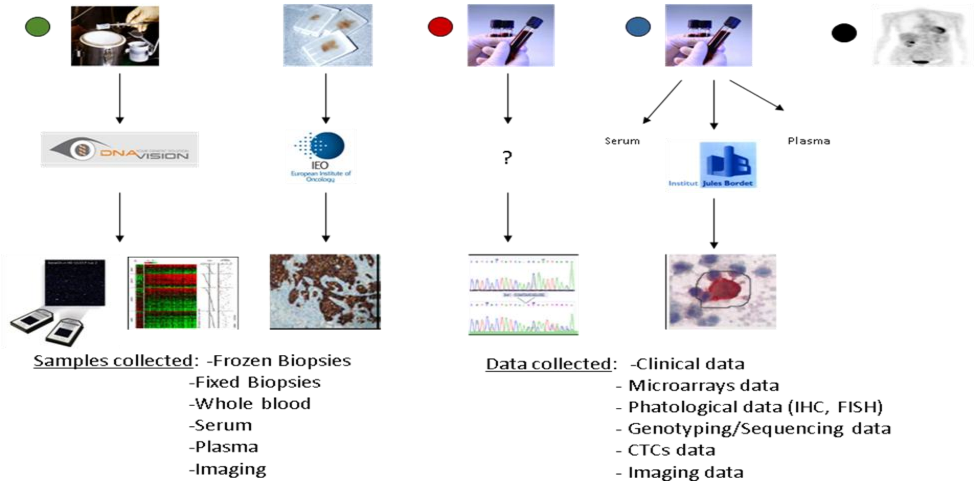

DNA VISION
IEO European Institute of Oncology
?
Serum
Plasma
Institut Jules Bordet
Samples collected: -Frozen Biopsies
-Fixed Biopsies
-Whole blood
-Serum
-Plasma
-Imaging
Data collected: -Clinical data
- Microarrays data
- Phatological data (IHC, FISH)
- Genotyping/Sequencing data
- CTCs data
- Imaging data

Figure 2 - Samples and data collection

In order to exploit the clinical trials to the fullest, it is apparent that access to the data generated by a clinical trial cannot be restricted to a single individual or institution. Therefore, the NeoBIG project envisions a platform in which data gathered by the clinical trials can be easily accessed, but at the same time carefully controlled. In this document, the requirements for such a platform are laid out. This platform is envisioned to provide secure and controlled access to the data gathered in the clinical trials for various parties. The platform thus will contain a wide variety of data types (see Figure 2), ranging from the Case Report Forms collected in the clinical trials to the latest technologies in the genomic and proteomic fields (e.g. molecular data, microarrays, imaging, etc.). The platform should result in a lasting bioinformatics platform for collaboration between cancer research institutes in Europe and have a strong focus on interoperability.

2 Requirements

We have collected the requirements by carrying out interviews and discussions with the main stakeholders of the data sharing platform, i.e. with representatives of BIG for the clinical aspects and with representatives of the Breast European Adjuvant Studies Team (BrEAST), the data centre of the BIG, for an IT perspective.

The programme will include several (five currently planned) neo-adjuvant trials as described in Section 2, that will be carried out together with various Pharma companies. A first trial is scheduled for 2010 and there will be a period of about 6 months between trial starts to allow time to phase in.

For the first trial the participants are university hospitals and it is expected to enrol around 300 patients per trial. There will be a consortium of trials, in each trial there will be 20-25 core partners (mainly EU: Belgium, Spain, Italy, the UK, but also Russia, India, South Africa, Singapore)

For the first trial a questionnaire has been sent out to assess the capabilities of the centres, including PET, MRI, to ask about the interest of the centres in the trials and their recruitment capabilities. Next to about 20 core partners, another 30 peripheral institutions will participate (each core partner may representing several hospitals).

With respect to the recruitment volume, for example a running trial joining around 120 centers recruited 450 patients.

Each trial will proceed as follows:

Year Zero: CRFs creation, sites initiation (contracts negotiation, budget, etc.).

Year One:

1. Biopsy samples are collected before treatment and sent to a single centre (IEO in Milan) for histopathology, to another single centre for gene expression profile (samples sent to an SME for processing and to be profiled at IJB in Brussels). Next step is to reconcile the two results (histopathology grading and genomic grading based on gene expression), before randomization, with respect to the eligibility criteria.

- If “yes&yes”, the patient is definitely admitted to the trial.
- The gene expression profiling will be generated at IJB. The patient data should satisfy at least minimal criteria from Milan, but definitely fit the Bordet criteria.
- Expected 7-10 days turnaround.
- Randomization in 4 arms will be done in Scotland, then the result will be sent to the centre (probably at IJB)
- Data elements:
    - Pathology (PA) (Result/Report) – digital pathology is out of the scope, maybe small image crops not full-resolution PA digital slides
    - Expression profile (30 MB/file)
    - SNP file (67MB/file)

- Imaging: PET/CT, MRI, Mammography, US (nice to have, but not clear how to transfer, “CD mail” can be considered as an option)
- Imaging use case not fully decided yet, “dual” check-up, centralized read use cases suggested
- Clinical data
- CRFs sent as pages (with barcode), filled in by the participating organizations and faxed back. The data entry will be done at IJB

2. All the above is repeated after 2 weeks of treatment (potentially also with MRI) to detect changes.

3. Repeat all again, at the time of the surgery and collect information on the surgery outcome. From the surgery biopsy samples are again tested in Milan (histopathology) and at Bordet (gene expression).

For the pharma company the trial is finished after surgery, but for research it would be interesting to follow the patients further. However there are too few patients. On average only data for 50% of the patients are available for follow up. Postcards are sent after the completion of the treatment to investigate the overall result (expected a large dropout rate – no answer).

It is harder to describe a clear use for the imaging data as it was not used before. PET/CT data should be centralized and perhaps also other types of imaging (MRI, US). For histopathology the image may also be retrieved (the data instead of only the interpretation).

Central reading of images would be desired as second opinion, to check the results from the local centres, and to reconcile the differences. Also reading MRI images may be challenging at some centres. The images should be transferred through the Internet, although in the first trial disks may be sent around.

Considering the priorities, three levels can be defined:

Level 1	CRF + gene expression profiles
Level 2	CRF + gene expression profiles + imaging data (PET/CT, US, XRay, MRI)
Level 3	CRF + gene expression profiles + imaging data + (digital) pathology

In principle, the data should be gathered electronically. However, there should be accounted for that the infrastructure of the participating partners might not yet support this (thus an alternative path should be possible).

Each trial will have its own central review to ensure quality.

Standard arm data is to be shared/access by privileged parties (Pharma, participating sites), but there will be no access to the experimental arms before the trial is published. Afterwards that data may also be considered for sharing. The results from the experimental arms are only available to the trial sponsor.

For the trials there will be different setups in different countries, but the repository could be managed and maintained by BrEAST. Especially in the UK or Germany there may be local repositories as well. The BrEAST data centre is the default data centre. They do the digitization of the CRFs. The access could be widened to the data storage.

Data will be maintained for at least 5 years and there will be around 5 x 300 = 1500 patients from the first five trials. But when the results are promising new trials may be initiated. The life of the repository should be significantly beyond the duration of the trials.

The data collected from the standard arms could be used to plan new research and to refine the results through follow up trials. This data should be stored and managed by the data sharing platform and the sharing of the data in a secure and privacy-preserving way among the members of the NeoBIG consortia and other authorized parties (potentially at a cost) should be supported.

The experimental arms of each trial should only be available to the parties participating in that trial, when they are authorized to access the data.

Currently the only communication requirement is that all centres have e-mail and fax access. eCRFs could be an interesting option, but costs are considered very high. There is a push from the pharma companies to introduce eCRFs. But the current priority noeBIG is the gene expression and the clinical data.

- All CRFs are currently paper based (faxed)
- The process is as follows: Fax Server -> Image Server -> (manual) data entry
- eCRF solutions are currently explored (Oracle deemed expensive, an academic solution – hosted in the US considered as an alternative, building their own solution or customizing an OS solution currently rejected for the lack of time/resources)

To provide a platform that enables data sharing and collaboration between cancer research centres, NeoBIG requires a robust, secure IT solution that is compliant with a wide set of regulations and laws in the context of security, safety and privacy protection. The platform needs to be able to store, manage, and share the various types of data that will be generated by NeoBIG trials, as discussed further in Section 2.1.

Security is an important aspect of the NeoBIG data sharing infrastructure. NeoBIG deals with personal data obtained from patients, whose privacy needs to be protected (both from an ethical and a legal perspective).Secondly, future prospective clinical trials with targeted therapies will require a system capable of dynamically setting up collaborations of organizations around specific data sets. Data shared within such a group needs to be well protected. Therefore, the NeoBIG data sharing platform needs to assure secure data sharing, such as authentication of users (secure logon), authorization (access control), encryption (to guarantee confidentiality), trust establishment, and Virtual Organization Management. Additionally, the interactions with the NeoBIG data sharing platform need to be fully audited to enable traceability.

Strong requirements on the data sharing platform are production-level reliability and availability and full maintenance. The data sharing platform will be used and needs to be available long beyond the end of the clinical trials, as the data is highly valuable for further research.

Additionally, data interoperability and adherence to widely accepted international standards are important requirements which will enable the collaboration between BIG and other cancer organizations world-wide. In that context, well-known standards (HL7, DICOM, MIAME, MAGE, etc.) and terminologies (SNOMED, LOINC, etc.) are relevant, but also new standards emerging with the development and adoption by the US research community of relevant NeoBIG tools.

As collaboration with the US cancer research community is desired and the US market is important for the pharma organizations participating in the NeoBIG trials, additional requirements need to be extracted from regulatory frameworks (such as FDA 21 CFR part 11) to which compliance needs to be assured.

The figure below identifies all the aspects that we have considered as relevant in the requirements collection phase. We will briefly discuss them further.

2.1 Data access

The main purpose of the NeoBIG platform is to provide seamless access to the clinical trial data stored in the system, to be shared by the sites that have participated to the trials but also by other research organizations for meta-analysis, hypotheses building, etc. There are several aspects relevant in this respect.

- Types of data:

Of course, the types of data that need to be stored are a source of requirements with respect to amount of storage needed, repositories that need to be built, standards used for storing and exchanging the data, etc.

The NeoBIG trials will include the following data:

- Clinical Report Forms.
- Genomic data: gene expression profile and SNP data.
- DICOM images that can be stored as flat files or in a PACS system.
- Patient (data) tracking.
- Lab data, both raw data and reports.
- Various metadata describing the trials, the arms of the trials, the patient context, the samples, the image files, etc. This data needs to be curated and a relevant aspect is to decide who will carry out the curation for each type of data.

- How to find data:

Once the data is in the system, it becomes relevant to decide how the users are going to search and retrieve the data they need. These aspects need to be decided based on the research workflow, by looking at the type of questions users may need to answer based on the data. Additionally, one needs to take into account the legal, ethical and privacy requirements and the vision of the owners of the system and of the data with respect to what should be allowed to do with the data.

The authorized user could be allowed to access the data per trial, per trial arm, by querying the actual data content, or by querying the metadata, such as the eligibility (inclusion/exclusion) criteria for a trial, the demographics, the description of the data content, or the description of the trial.

- When should the data be accessible:

For analysis the data is accessed after the completion of the trial, however there are use cases for accessing the data during the running of the trial, such as to verify how many patients are enrolled so far, etc.

- How should the data be accessed:

One needs to decide how to offer access to data to the user of the system. The user may be allowed to access directly relevant parts of the data, to retrieve raw data files, to retrieve encrypted data, or to access the data via analysis services provided on top of the data sharing platform. The chosen solution will cover only one or some of these options.

- Who will access the data:

Potential users of the data sharing system may be the participating research institutes, other academia and research institutes, interested pharma organizations, patients, general public. It is expected that the NeoBIG system will be open for the participating institutes and pharma organizations, and access will also be provided to other organizations from academia and industry interested to carry out research on the data.

- Where will the data reside:

The choice was whether to preserve the data decentralized, at all participating institutes, or centralized at one or several centres. It was chosen for the second alternative, where one or a few centres will be in charge of the data. Most likely, the data will reside at the BrEAST data centre.

2.2 Data collection

An important step in building this data sharing system is to carry out the data collection. In that respect there are several aspects to be considered:

- The data can be gathered offline, and filled into the system at a later time at the centre managing the system, or data can be collected online by allowing the users to upload it directly into the system. In the NeoBIG case the first option will be initially chosen.
- It is also relevant to consider what expertise and IT infrastructure is expected from the contributing partners. What kind of network we expect, do we expect IT expertise and dedicated IT personnel, etc. We have chosen to assume minimum infrastructure and expertise, which supports the choice at the previous point for an offline data gathering.
- The collection granularity refers to the volume of data that will be provided at one time, to be loaded into the system. We could consider collection per patient endpoint, per test, per batch of tests, and per trial. All these are possible, but collection per patient endpoint and per patient test seems to be more likely.

2.3 Certification

To be used in a clinical research setting and in collaboration with pharma organizations the system needs to comply with various legal and privacy regulations. An example of such regulatory framework to which compliance needs to be assured is FDA 21 CFR part 11. We will opt for self-certification of compliance with those frameworks as obtaining a formal certification by an external party incurs unreasonably high costs.

Additionally, pharma organizations require finalized products and are not interested to work with research prototypes.

2.4 Privacy and security

As previously mentioned data privacy and security are essential for such a system. Aspects considered are:

- The grain of access control:
    - All or nothing
    - Per trial/ trial arm
    - Per patient
    - Per test
    - Per part of a test
    - Parts that will become public

All these alternatives are still considered relevant and the choice of which will be supported will be made later in the design of the solution.

- One needs to choose between anonymization of the data and the pseudonymization of the data (with trusted third party). The location where the anonymization/pseudonymization will take place also needs to be decided.
- Audit trail for all transactions concerning the data needs to be provided as well.

2.5 Process

Several other requirements concern the process, such as staging requirements for a fast initial deployment, deciding for the available budget for development of the system taking into account the necessary investments in hardware and software, the end of life for the components and further reinvestments for maintaining the system operational, the budget for maintenance and the budget available for running the system (e.g. is help desk required). These aspects are highly relevant but will be decided at a later time, closer to the actual design of the system.

2.6 Non-functional requirements

The desired performance of the final system also needs to be considered. One needs to evaluate the minimum performance expected for queries, data retrieval and data upload. The desired uptime, reliability and availability of the system also need to be estimated. The system is not considered to be time-critical, but of course it should be able to provide the desired level of service to the users. At this moment actual values have not been attached to the parameters considered.

The available maintenance represents an important element. The system needs to be available long after the last trial has finished and it is expected that new programmes following NeoBIG may also add their data to the same system. Additionally, the main goal of the system is to make data available post-trials to be used for research by interested parties. The amount of partners that will use and maintain the system and the sustainability aspects are also to be considered.

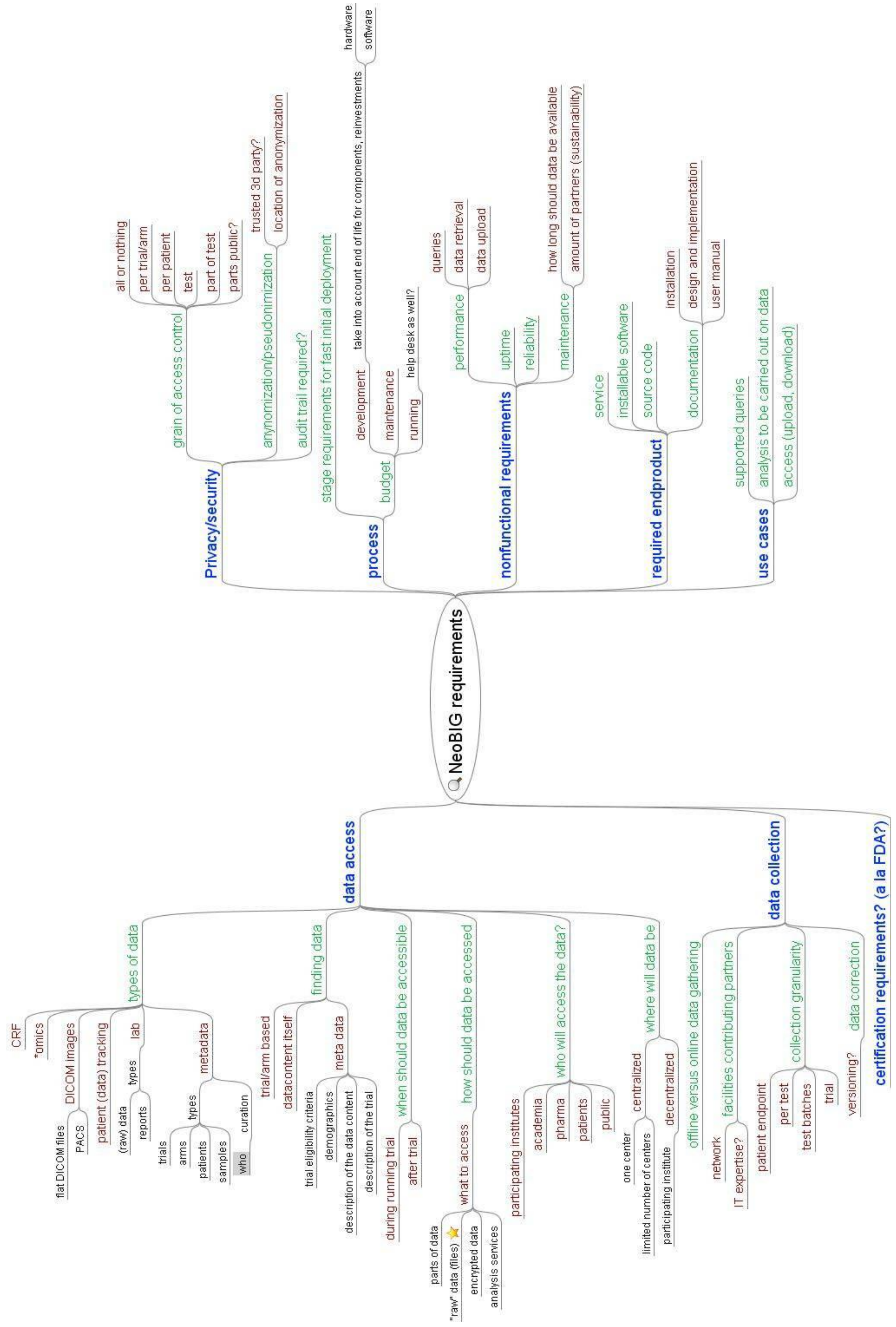
2.7 Required end-product

Before starting the development of the data-sharing platform it needs to be decided what the desired end-product is. That may be a commercially-provided service, an installable software or a source code. The current discussions suggest that the most viable alternatives are to provide a service operated by external parties in collaboration with BrEAST or to provide an installable software that will be maintained by BrEAST or by another party.

The level of documentation to be provided is of course not decided at this stage, but it is highly relevant and should be taken into consideration. Documentation could cover the installation of the system, the design and the implementation, and the use of the system by end-users (user manuals).

2.8 Use cases

To proceed with the design of the system, relevant use cases need to be elaborated together with the end users. These use cases should cover the typical queries that need to be supported by the system, the analysis that the user needs to carry out on the data, and the required access to the data both for upload and download. Several scenarios that need to be further refined into use cases are included in the next section of this document.


NeoBIG requirements
data access
types of data
CRF
*omics
DICOM images
flat DICOM files
PACS
patient (data) tracking
lab
types
(raw) data
reports
metadata
types
trials
arms
patients
samples
curation
who
finding data
trial/arm based
datacontent itself
meta data
trial eligibility criteria
demographics
description of the data content
description of the trial
when should data be accessible
during running trial
after trial
how should data be accessed
what to access
parts of data
"raw" data (files)
encrypted data
analysis services
who will access the data?
participating institutes
academia
pharma
patients
public
where will data be
centralized
one center
limited number of centers
decentralized
participating institute
data collection
offline versus online data gathering
facilities contributing partners
network
IT expertise?
collection granularity
patient endpoint
per test
test batches
trial
data correction
versioning?
certification requirements? (a la FDA?)
Privacy/security
grain of access control
all or nothing
per trial/arm
per patient
test
part of test
parts public?
anynomization/pseudonimization
trusted 3d party?
location of anonymization
audit trail required?
process
stage requirements for fast initial deployment
budget
development
take into account end of life for components, reinvestments
hardware
software
maintenance
running
help desk as well?
nonfunctional requirements
performance
queries
data retrieval
data upload
uptime
reliability
maintenance
how long should data be available
amount of partners (sustainability)
required endproduct
service
installable software
source code
documentation
installation
design and implementation
user manual
use cases
supported queries
analysis to be carried out on data
access (upload, download)

3 Scenario’s

This chapter describes various scenarios in which users interact with the data platform.

3.1 Clinical researcher accesses data of a clinical trial for basic research

In this scenario, the users - a clinical researcher from institute X - accesses both arms of clinical trial A to assess standard hypothesis 1 “Benefit of at least 1 surrogate endpoint in the experimental arm versus the control arm in neoadjuvant setting”, and to refine hypothesis 2 “Molecular markers profile trigger this benefit” to determine the molecular markers profile.

1. The user logs into the data platform and authenticates himself.
2. The user sees an overview of all clinical trials to which he has access.
3. The user selects clinical trial A.
4. The user sees the arms of clinical trial A to which he has access. In this case, both arms are available.
5. The user sees an overview of the types of data that are available for download (such as case report forms, microarrays, ...)
6. The user chooses to download the case report forms (as csv[1] file)
7. The user analyses the case report forms on his local computer infrastructure to confirm hypothesis 1: there is at least 1 surrogate endpoint better in the experimental arm versus the control arm in neoadjuvant setting.
8. As hypothesis 1 is confirmed, the user chooses to download the microarrays of the clinical trials.
9. The user performs a statistical analysis on the microarrays on his local computer infrastructure and defines the molecular marker profile, forming hypothesis 2.

3.2 Pharma accesses data of a clinical trial control arm

A pharmaceutical company considers sponsoring a new trial. Upfront, the pharmaceutical company wants to assess the probability that enough patients can be enrolled into the trial.

1. The user logs into the data platform and authenticates himself.
2. The user sees an overview of all clinical trials to which he has access.
3. The user selects all clinical trials.

[1] Comma separated value

4. The user sees the arms of the clinical trials to which he has access. In this case, the user is only authorized to access the control arms (the standard of care arm).
5. The user sees an overview of the types of data that are available for download (such as case report forms, microarrays, ...)
6. The user chooses to download the case report forms (e.g. as csv file).
7. The users performs an analysis on the local computer infrastructure to see what percentage of the enrolled patients match the eligibility criteria of the envisioned trial.

3.3 Generation of a new hypothesis for a clinical trial

In this scenario, a clinical research reuses data from a clinical trial to generate a new hypothesis.

1. The user logs into the data platform and authenticates himself.
2. The user sees an overview of all clinical trials to which he has access.
3. The user selects clinical trial A.
4. The user sees the arms of clinical trial A to which he has access. In this case, the user is only authorized to access the control arm (the standard of care arm).
5. The user sees an overview of the types of data that are available for download (such as case report forms, microarrays, ...)
6. The user chooses to download the case report forms (as csv file) and the microarrays of the clinical trial.
7. The users performs an analysis on the local computer infrastructure and tries to find a molecular marker profile which can predict the outcome of a selected surrogate endpoint.

3.4 Clinical researcher uses data of multiple trials for meta analysis

In this scenario, a clinical researcher integrates the data of multiple trials to arrive to at a new hypothesis.

1. The user logs into the data platform and authenticates himself.
2. The user sees an overview of all clinical trials to which he has access.
3. The user selects clinical trial A and B
4. The user sees the arms of clinical trials A nd B to which he has access and downloads the control arms.
5. The user sees an overview of the types of data that are available for download (such as case report forms, microarrays, ...)
6. The user chooses to download the case report forms (as csv file) and the microarrays of the clinical trials.

7. The microarrays are different between the different trials. The user performs an analysis to link the different microarray designs into a new “virtual” microarray, allowing for a subsequent analysis based on the combination of all microarrays of the two control arms.

4 Conclusion

This document carries out an initial requirements collection and analysis for the data sharing platform meant to support the neoBIG research program of the Breast International Group. We have also briefly described several scenarios concerning the expected uses of the platform.

Part of this investigation, we have also looked at ways in which this new platform could benefit of the tools and services developed in the ACGT project, but also in the much larger caBIG project. We have concluded that there is a lot of ACGT expertise that could be used for the neoBIG data sharing platform, especially with respect to data storage, management and sharing, and with respect to privacy and security. On the other hand, due to the very strict requirements for a production-level system, with available documentation and user support, commercial deployment and long term maintenance, we have concluded together with the BIG that current ACGT prototype tools and services cannot be directly used for the neoBIG project. The same was preliminarily concluded about available caBIG tools and services. caBIG should still be evaluated as several standards emerging from that community should be taken into account in the development of the platform to enable interoperability and facilitate the collaboration between BIG and the US research community.

BIG will further make use of this investigation to refine their requirements concerning to the data sharing platform for their new research programme, to make the necessary choices and to set up a future initiative focusing on the design and development of the platform.

On the ACGT side, this investigation has allowed us to confirm (and sometimes infirm) our choices and research ideas. Fortunately, we can conclude that although due to the prototype status of our solutions and to the fact that we are not able to provide commercial service level agreements, long term maintenance and some of the more focused requirements such as certification, our solutions cannot be directly used in the neoBIG scenarios, much of our expertise and ACGT work in privacy, security and data access have proven highly relevant for the neoBIG data sharing platform.

5 References

[1] Phuong Dinh MD, Breast International Group, “The NeoBIG program; Research program to accelerate drug & biomarker development in early breast cancer 2009 - 2014”, Pre-IMPAKT Training Course , May 6th 2009, Brussels, http://www.esmo.org/fileadmin/media/presentations/1324/opening/Dinh.ppt

[2] Breast International Group, http://www.breastinternationalgroup.org/