BD2Decide

Big Data and models for personalized Head and Neck Cancer Decision support

TITLE	**The BD2Decide visualisation mock-ups**		
Deliverable No.	D5.3		
EDITOR	Vasilis Tountopoulos (ATC)		
Contributors	Thanassis Dalianis (ATC), Antonis Kosmas (ATC), Laura Lopez (UPM), Liss Hernández (UPM), Giuseppe Fico (UPM), Ana Ugena (UPM), Maria Teresa Arredondo (UPM)		
WorkPackage No.	WP5	**WorkPackage Title**	Visualizing the HNC Virtual Patient
Status[1]	FINAL	**Version No.**	1.0
Dissemination level	PU		
DOCUMENT ID	D5.3 The BD2Decide visualisation mock-ups		
FILE ID	BD2Decide D5.3		
Related documents	Annex I DoA version 27/07/2017 - BD2Decide D5.1, D5.2		

[1] Status values: TOC, DRAFT, FINAL

Distribution List

Organization	Name of recipients
AOP	T. Poli, E.M. Silini, E. Martinelli, G. Chiari, C. Caminiti, G. Maglietta, D. Lanfranco
VUMC	R. H. Brakenhoff, H. Berkhof, P. va de Ven, Th. Klausch
UDUS	K. Scheckenbach
INT	L. Licitra, E. Montini, G. Calareso, G. Gatta, A. Trama
UPM	G. Fico, L. Hernández, L. López, A. Ugena, M.T. Arredondo
POLIMI	L. Mainardi, V. Corino
Fraunhofer	S. Wesarg
ATC	V. Tountopoulos, Thanassis Dalianis, Antonis Kosmas
MAASTRO	F. Hoebers
AII	A. Algom, R. Shefi
MME	S. Copelli, F. Mercalli
UNIPR	S. Rossi, M. Silva
European Commission	Project Officer: and all concerned E.C. appointed personnel and external experts

Revision History

Revision no.	Date of Issue	Author(s)	Brief Description of Change
0.1	03.05.2018	V. Tountopoulos, A. Dalianis	Table of Contents
0.2	21.05.2018	V. Tountopoulos	Draft version of Sections 2 and 4
0.3	24.05.2018	A. Dalianis, A. Kosmas	First draft of the CDSS use
0.4	29.05.2018	G. Fico, L. Hernández, L. Lopez, M.T. Arredondo, A. Ugena	First draft of the VAT use
0.5	07.06.2018	V. Tountopoulos, A. Dalianis	Second draft of the document and editing for internal review
0.6	11.06.2018	V. Tountopoulos	Prefinal version for review
0.7	14.06.2018	G. Fico	Overall review of the deliverable
1.0	15.06.2018	V. Tountopoulos	Final version

Addressees of this document

This document is addressed to the BD2Decide Consortium and informs on Deliverable D5.3, which is a demonstrator. The main part of the deliverable is a set of functional mock-ups of the two main tools of the BD2Decide platform, namely the Clinical Decision Support System (CDSS), which supports clinicians in making decisions across the head and neck cancer (HNC) treatment process, and the Visual Analytics Tool (VAT), which enable researchers involved in HNC to analyse big datasets of clinical and population data. This document acts as a supportive material and links the respective deliverable D5.3 with the previous project deliverables, towards the implementation of a user centric design approach.

To this end, the main result of this deliverable is a comprehensive set of guidelines for accessing the functions offered to the target stakeholders through the current implementation of the CDSS and VAT tools. The respective mock-ups comprise the result of the fruitful cooperation between the clinical partners, as the early adopters of the perceived software, and the technical partners, who materialise their knowledge on the relevant fields into actionable software solutions.

The deliverable is an official result of the project and it will be delivered to the European Commission through this document.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abbreviations and definitions

BDI	Big Data Infrastructure
CDSS	Clinical Decision Support System
DPEE	Digital Patient Exploration Environment
e-CRF	Electronic Clinical Record Format
HN	Head and Neck
HNC	Head and Neck Cancer
HTTP	Hypertext Transfer Protocol
ICT	Information and Communication Technologies
IPDA	Interactive Patient Decision Aid
LSU	Local Storage Unit
PDS	Patient Documentation System
QoL	Quality of Life
REST	Representational State Transfer
SOA	Service Oriented Architecture
TBCE	Tumor Board Collaboration Environment
UI	User Interface(s)
URL	Uniform Resource Locator
VAT	Visual Analytics Tool

Abstract

The active involvement of end users and stakeholders in the development of software products aims to maximise the technology acceptance rate and ensures the sustainability of the product in the commercial line of the respective market. The use of ICT in the health domain introduces a radically new approach for the treatment decision-making process, since it brings the power of technology to the service of the clinical experts and other health professionals. As data is enormously increasing, the exploitation of mechanisms to harness this big data and present it to the health experts becomes a critical challenge, which requires these stakeholders to be centrally positioned in the development of new tools and services.

This concept has been adopted in the BD2Decide project, resulting in a continuous engagement of clinicians and researchers in the development of the two principal visualisation tools, namely the Clinical Decision Support System and the Visual Analytics Tool. This deliverable summarises the outcome of a long period effort to come up with functional and working prototypes of these tools that accomplish the expected tasks for the target stakeholders. More specifically, the deliverable presents the way that the Clinical Decision Support System supports the BD2Decide workflow for managing the collection, processing, storage and visualization of the patients' data and their assessment. It provides details on the usage scenarios of this tool from health professionals, either being physicians in a Head and Neck Cancer (HNC) case or undertaking a specific role in the HNC treatment process, like a surgeon, a medical oncologist, pathologist, a radiologist or a radiation oncologist. Such scenarios include the management of the information collected and/or processed / produced across the HNC treatment process, the combination of clinical, imaging, genomic and radiomic variables into aggregated visualisations for making assessments on the patient status and disease evolution, compared to similar cases and the support for assistance in making informed decisions on the selection of the appropriate treatment for the HNC patients, based on the expected life expectancy rate, which is developed using statistical prediction models or the integration of big data from cohort data and public sources.

Furthermore, the deliverable presents the Visual Analytics Tool, which supports the exploitation, representation and visualization of information retrieved from large-scale and heterogeneous data sources in the field of HNC. In detail, the tool supports clinicians in performing research activities through a set of intuitive graphs, charts and images, supporting exploratory and collaborative search, built on top of statistical modelling and big data analytics techniques developed in WP3, WP4 and WP6.

1 ABOUT THIS DOCUMENT

1.1 Introduction and scope

This deliverable aims to present the results of the user-oriented development approach for the two main visualisation tools of the BD2Decide platform, namely, the Clinical Decision Support System and the Visual Analytics Tool. These results include the implementation of functional and working prototypes for the user interfaces of these tools.

The respective work is driven by the design of user interaction illustrations, which has started in the BD2Decide Deliverable D2.2 [1], and the development of mock-up screens, which have been shaped in the BD2Decide Deliverable D5.2 [2]. Following the results of these two deliverables, this deliverable builds on top of user experience and evaluation workshops, in order to deliver working versions of the user interaction tools, which assist the health professionals in making treatment decisions in Head and Neck Cancer (HNC) patient cases and the researchers in executing their clinical investigation tasks. The results of this deliverable will comprise the principal vehicle in evaluating the effectiveness of the BD2Decide platform for a more personalised prognosis of the treatment outcome in HNC cases.

1.2 Top 15 Request

Clinicians have created the following list of 15 request to examine variable correlations and find patterns. The purpose of these clinician-driven requests is to evaluate if any of the combinations can change the prognostic curves of Stage III and IV of HNC patients.

1. Stage + Overall Survival
2. Stage + Radiomics
3. Stage + Other clinical factors (i.e. comorbidities, up to 10 variables selected by the investigator from the CRFs fields)
4. Stage + Other clinical factors (i.e. comorbidities, up to 10 variables selected by the Investigator from the CRFs fields) + genomic
5. Stage + Other clinical factors (i.e. comorbidities, up to 10 variables) + genomic + radiomic
6. Stage + Other clinical factors (i.e. comorbidities, up to 10 variables) + radiomic
7. Stage + Genomic
8. Stage + Genomic + Radiomics
9. Genomic + Radiomics
10. Stage + Treatment + Genomic
11. Stage + Treatment + Radiomic
12. Stage + Treatment + Genomic + Radiomic

13. Stage + Pathological +Genomic
14. Stage + Pathological + Imaging +Genomic
15. Stage + Pathological + Imaging +Radiomics + Genomic

1.3 Structure of the deliverable

Following the objectives set for this deliverable, the document is structured as follows:

- Section 2 makes an overview of the BD2Decide environment and the position of the visualisation tools in the high level BD2Decide architecture. In this section, we recap the architecture of the Clinical Decision Support System (CDSS) and the Visual Analytics Tool (VAT), describing their functionalities in the context of the BD2Decide system.
- Section 3 describes the development of a generic data management and acquisition platform for the collection of multisource, multiscale and multivariate data for Head and Neck Cancer treatment purposes.
- Sections 4 and 5 are dedicated to the presentation of the CDSS and VAT, respectively. These sections start with an overview of the expected scenarios of use and detail on the screens that have been developed to accomplish these scenarios and how the target end users could work with these tools. Furthermore, the sections briefly present the technologies used during the implementation phase of the tools, as well as the connection with other BD2Decide or external components.
- Finally, Section 6 provides concluding remarks for the work presented in this deliverable and how this work relates to future activities in WP5 and in the project as a whole.

2 OVERVIEW OF THE BD2DECIDE VISUALISATION TOOLS

The BD2Decide platform brings together a number of tools and services that serve the needs of various players in the e-health sector and specifically the HNC clinical research field. This section makes an overview of the platform environment and updates on the architecture design of the visualisation tools that the BD2Decide project delivers for the end users in the respective scientific and business fields. The remaining part of this section advances the descriptions already provided in the BD2Decide Deliverable D5.2 [2].

2.1 High level overview of the BD2Decide platform environment

The BD2Decide platform integrates various components, which are brought together to facilitate functionalities for the clinicians, the physicians, and other doctor specialties both on the level of exercising their knowledge for treatment purposes and experiencing with tools for enhanced research on the clinical field. To this end, the platform provides these stakeholders with the means to improve decision making along the three phases of the HNC treatment process, namely diagnosis, treatment, and post-treatment / follow-up periods.

An overview of the BD2Decide platform is presented in Figure 1: BD2Decide refers to multi-stakeholder groups from the health domain, offering a number of tools and services that add value to the multisource, multiscale and multivariate patient and population data and enhance the daily practices of these stakeholders. More specifically, the BD2Decide platform offers:

- The Clinical Decision Support System (CDSS), which refers to clinicians for supporting their decision making for the most appropriate treatment in HNC cases.
- The Imaging Analysis Tools, which refer to radiologists and radiation oncologists, including tools for the analysis and segmentation of CT and MRI images and the extraction of radiomic features.
- The Genomics Tools, which refer to bio analysts and support them in the extraction and analysis of genomic features from tissue samples.
- The Visual Analytics Tool, which refers to clinical researchers and supports them in clinical research tasks, which require the integration of patient and population data.
- The Statistical Analysis Tools, which refer to statisticians in the field of HNC and comprise a set of libraries with algorithmic approaches for the statistical analysis of the survival rate for HNC cohorts.
- The Big Data Infrastructure (BDI), which consolidates big data technologies for managing and processing the multisource, multiscale and multivariate patient and population data. This infrastructure supports data scientists in the HNC field to extract patterns with respect to the trends observed in big data structures for HNC datasets.

- The Interactive Patient co-Decision Aid (IPDA) tool, which refers to patients (and their physicians) and provides them with an explanatory environment for understanding the available treatments and their impact of the patient's quality of life.

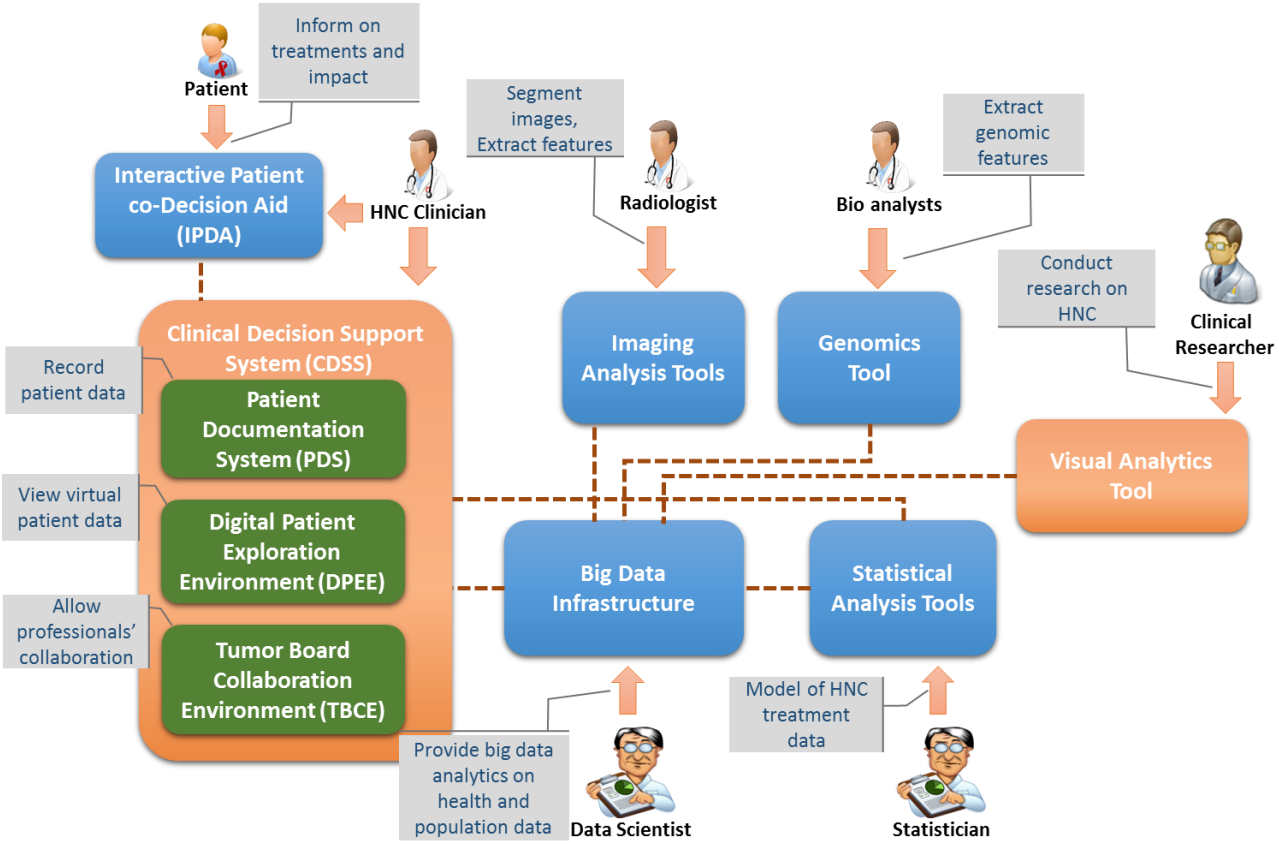

Figure 1: Overview of the BD2Decide platform environment.

BDI is the core component of the BD2Decide platform, which offers a data driven analysis of the patients' records. CDSS and the Visual Analytics Tool are the main visualisation tools, which build on top of the BDI and support the end users in accomplishing their tasks. In the following two sections, we will summarise the design and the functionalities for each of them.

2.2 Overview of the CDSS and VAT architecture

The three-layer architecture of CDSS and VAT tools is presented in Figure 2. The *Application Layer* is composed of the web application tool user interfaces. This layer is fed with the data available in the *Data Layer*, thanks to the usage of the components included in the *Services Layer*.

Both the CDSS and VAT benefits from the *Imaging & Radiomics Models* using the *Imaging & Radiomics* dataset, which are the data resulting from these services. *Imaging* data is used to observe the medical images at patient level in the CDSS together with the segmentation features. The images can be also useful for visualizing and processing in the VAT. On the contrary, *Radiomics* are useful in both tools but these data have to be processed before by the *Statistical Models* and/or *Big Data Analytics* available also in the *Services Layer*.

Similarly, *Genomic Models* produce the *Genomics* data available in the *Data Layer* in the BDI, and these *Genomic* data is accessible both in CDSS and VAT for patient data visualization and representation. Some *Genomic* data is only exploited in CDSS and VAT after the execution of the *Statistical Models* and the *Big Data Analytics* at *Service Layer*.

In summary, CDSS and VAT have direct communication with the *Statistical Models* and *Big Data Analytics*, whereas the *Imaging & Radiomics Models* and *Genomic Models* are intermediate services needed to get the corresponding data.

On top of all services, the *Data Management System* and *Knowledge Management System*, are the responsible elements to coordinate all the data transfer and usage through the different services and applications. Details will be reported in the BD2Decide Deliverable D6.4, due M36 (December 2018).

In order to use the services and to exploit the data detailed in the *Data Layer*, CDSS and VAT consume the dedicated RESTfull APIs implemented in the project (https://bd2decide.unipr.it/doc/).

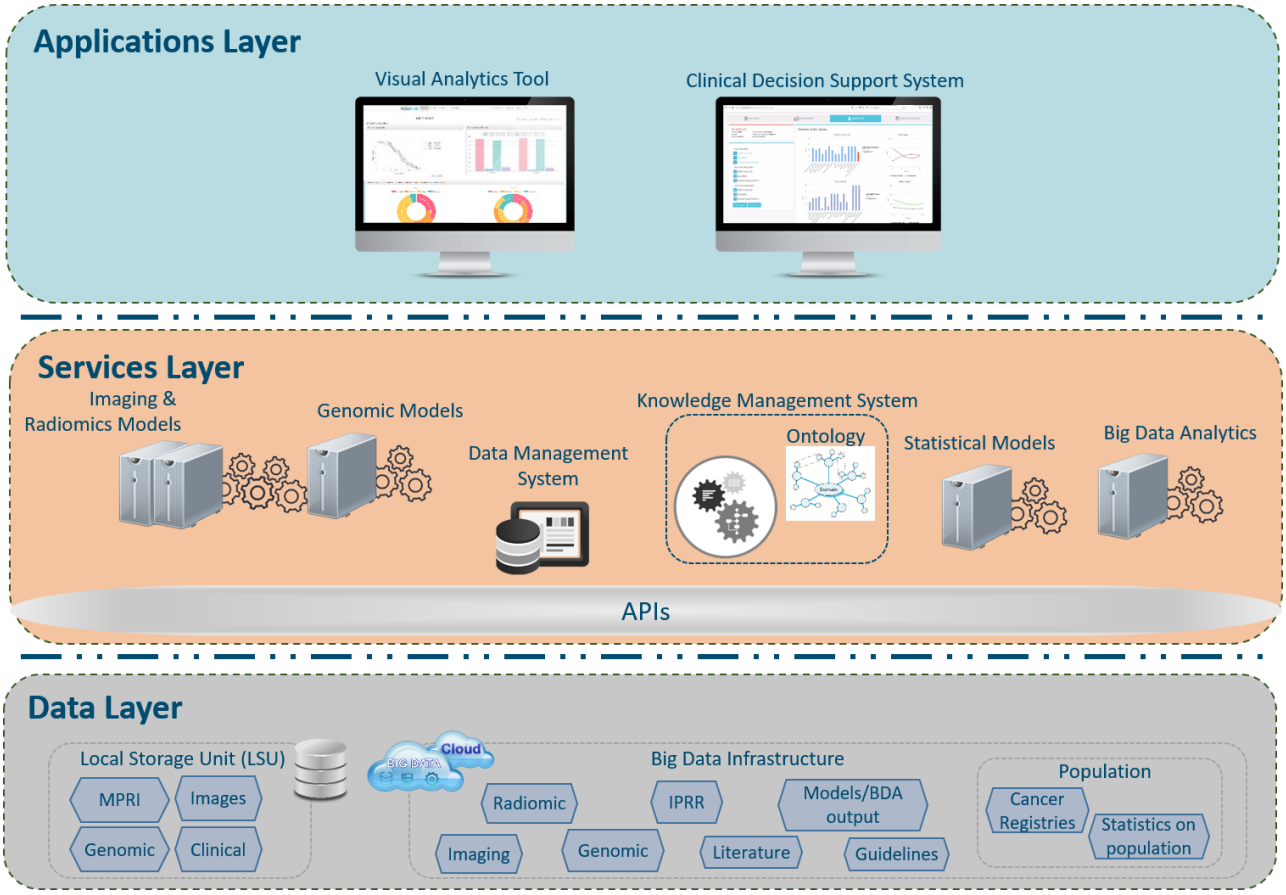

Figure 2: The architecture of the CDSS and VAT tools.

Apart from the data outputs generated for the services above mentioned, in the *Data Layer*, there are other types of data that are exploited by *VAT*:

- IPRR (Integrated Patients' Records Repository): it contains all patient clinical data (initially available in each hospital LSU) and inserted in BDI via OpenClinica user interface. The link with

the patient data stored in the BDI and the patient local ID is stored only locally in each hospital LSU, concretely in the Master Patients Records Index (MPRI).

- Literature: it holds PubMed articles information related to Head and Neck Cancer.
- Guidelines: reference clinical guidelines with recommendations on treatment decisions (e.g. ESMO).
- Population: head and neck cancer registries and statistics on population around Europe.

CDSS is also fed by the IPRR, guidelines and population datasets.

2.3 Overview of the CDSS functionalities

CDSS is one of the main visualization tools of the BD2Decide platform, as shown in Figure 1. The tool implements the BD2Decide workflow for managing the collection, processing, storage and visualization of the patients' data and their assessment. Around *CDSS*, we build various components and tools, which offer additional functionalities and enable the integration of individual services and components, such as the prediction analysis and prognosis of the life expectancy rate for a patient.

CDSS targets health professionals, being physicians in an HNC case or undertaking a specific role in the HNC treatment process, like a surgeon, a medical oncologist, pathologist, a radiologist or a radiation oncologist. This group of users aims to:

- Manage the information of the HNC treatment process, by browsing anonymised data, structured in the form of e-CRF, for the clinical status of their patients;
- Combine various data variables into aggregated visualisations for making assessments on the patient status and disease evolution, compared to similar cases;
- Receive assistance for making informed decisions on the selection of the appropriate treatment for their patients, by requesting the patient life expectancy rate, according to different statistical prediction models and based on the integration of big data from public sources in these models.

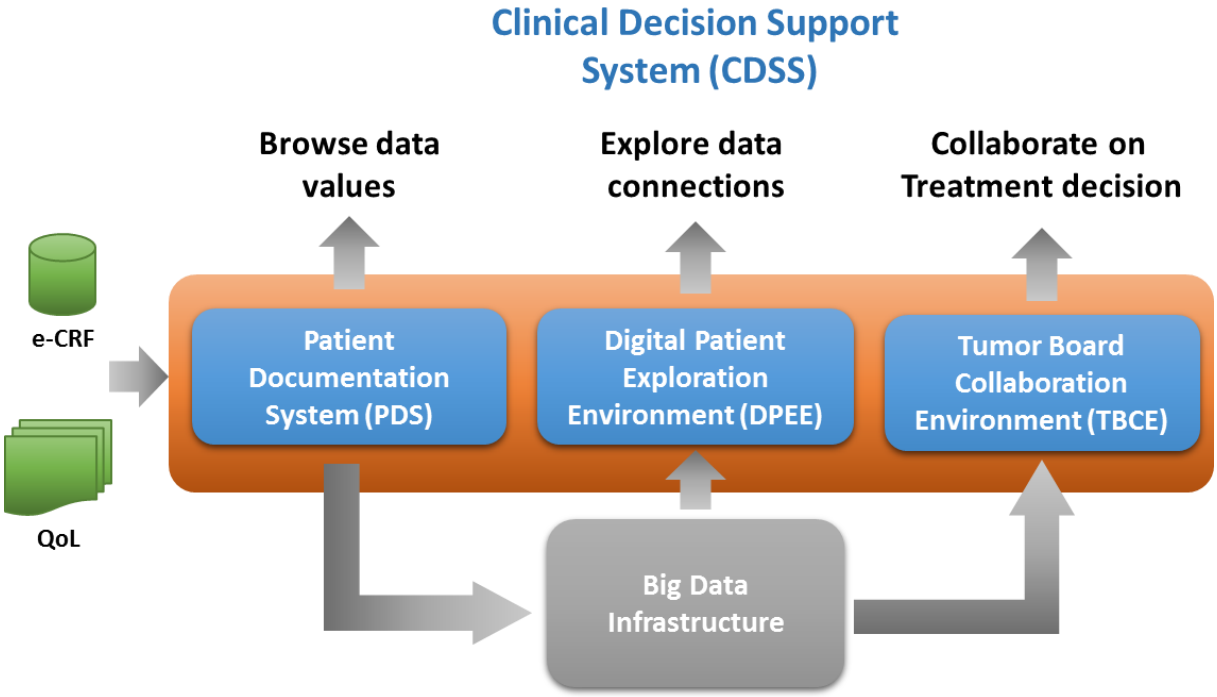

Figure 3: The conceptual design of CDSS.

In order to implement these principal functions, *CDSS* is structured into three main components, as shown in Figure 3 and explained in the following:

- The *Patient Documentation System* (PDS): PDS is responsible for maintaining the records of the patients' data. This tool implements the functionalities for importing patients' data into the Integrated Patients' Records Repository (IPRR) and retrieving them to visualise either the complete electronic Clinical Record Format (e-CRF) as raw information, through the PDS User Interface (UI), or feed the other tools of the BD2Decide platform.
- The *Digital Patient Exploration Environment* (DPEE): DPEE provides an enhanced view of the patients' profile, as it is defined in the BD2Decide project. The tool implements the digital patient concept to provide focused analysis and presentation of the patient's information and assist clinicians in a better interpretation of the patient's HNC status, through connections to the results of the prediction models, the patient's prognostic factors and the incorporation of the radiomic and genomic features into the patient life expectancy rate. The DPEE is essentially a presentation layer tool, which visualises the patient personalised risk scoring, the comparison of the patient's prognosis to similar cases (based on the output of different prediction models) and the assessment of qualitative factors, such as the collective responses and relevant scores from the Quality of Life questionnaires.
- The *Tumor Board Collaboration Environment* (TBCE): TBCE comprises an integral part of the treatment decision-making process and allows clinicians to collaborate with each other in the context of a virtual tumor board.

2.4 Overview of the Visual Analytics Tool functionalities

The Visual Analytics Tool, hereinafter named VAT, is a tool for the exploitation, representation and visualization of information retrieved from large-scale and heterogeneous data sources.

It is a web application tool that supports clinicians to carry out research activities, providing guidance through the interaction workflow implemented in the tool. This tool provides the clinicians, as the target end users of the tool, with the analysis outcomes together with graphs, charts and images in an intuitive way. VAT also allows the users to work in a collaborative manner through the 'Project' functionality. This feature works as a virtual container to manage the exploratory search projects via the execution of the different tasks available in the tool.

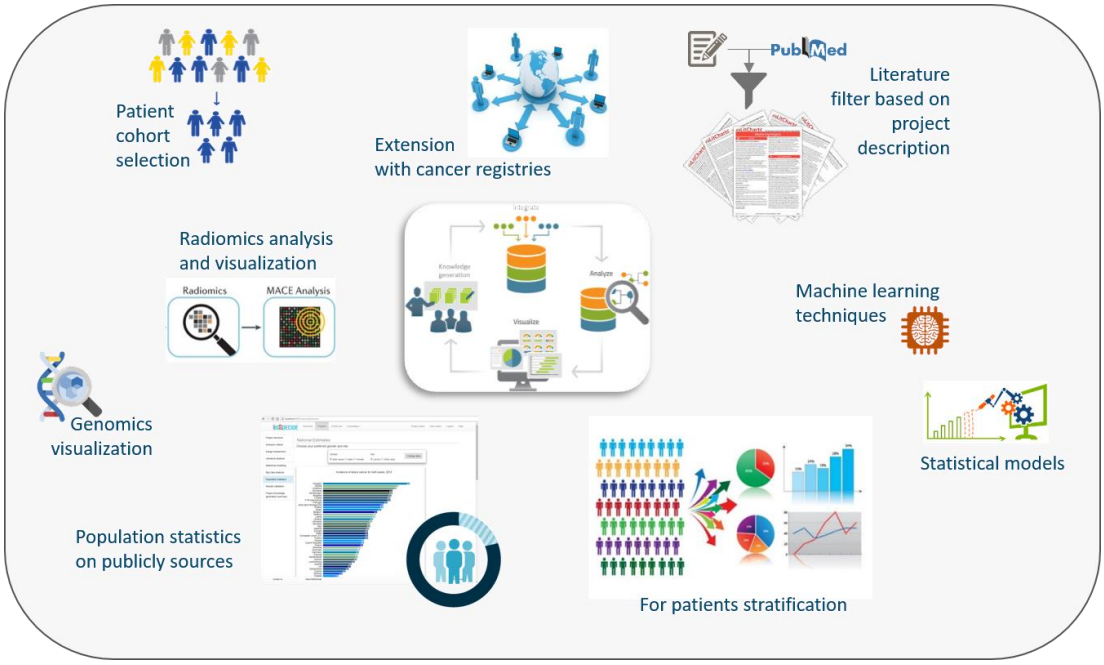

Figure 4: The *VAT* functionalities.

The VAT includes several functionalities (Figure 4) that allow clinicians to query and analyse the BD2Decide data. These functionalities are:

- To select a patient cohort on which to execute the research by filtering through the different fields defined in the e-CRF.
- To enrich the analysis with external cancer and epidemiological registries publicly available.
- To be up-to-dated with the literature in head and neck cancer research, personalized according to the research purpose.
- To execute machine learning techniques and statistical models taking into account the selected cohorts and the type of research performed.
- To visualize large amounts of quantitative features (radiomics and genomics) in an understandable way for the users.

3 DEVELOPMENT OF A GENERIC DATA MANAGEMENT AND ACQUISITION PLATFORM

3.1 Rationale for this section

The scope of this section is to make an overview of the work in Task T5.1 with respect to the development of a generic data management and acquisition platform, which drives the implementation of the Patient Documentation System. The latter is part of the CDSS tool, as it was introduced in Section 2.3.

The role of the generic data management and acquisition platform is to facilitate the collection of multiscale and multivariable patients' data from the different sources, within a hospital environment. More specifically, as stated in T5.1, the BD2Decide project deals with sensitive patient information, such as clinical health records, raw and segmented diagnostic images, radiomic features and genetic Binary Alignment Map (BAM) files and their genomic signatures. This information is coming from hospital information systems (HIS), laboratory information systems (LIS), and picture archiving and communication systems (PACS).

Currently, the collection of such data is mostly manually accomplished. Each clinical center has appointed a dedicated Network Attached Storage (NAS) equipment within their hospital network environment, in which all the necessary BD2Decide related information has been manually collected. Through this NAS, the appointed clinical partners can make use of the BD2Decide tools and process this information in in an anonymised or pseudonymised way, before uploading the respective data on the BD2Decide platform environment. This process is subject to the National and EU regulations and the internal security and privacy policies of each center.

Following the first review report, the project needs to provide evidence on a more automated process for the collection of the patients' data from the different information systems, being operating in the clinical centers. Thus, the scope of this section is to introduce the roadmap for the development of such a data management and acquisition platform.

3.2 Existing data integration alternatives

Collecting data from multiple sources, such the ones that can be found in an e-health environment, can be a complex task. It involves tying into existing information systems and storage tools and integration with applications and products that might include modern software tools as well as legacy systems, which are deployed inside a hospital or clinic environment. The collection of big e-health data requires the implementation of the appropriate integration pattern [4] that can be used in order to allow the communication of the various systems with the BD2Decide data acquisition platform. Such patterns range from the simple File Transfer approach and the Shared Database scheme to the Remote Procedure Invocation and the Messaging methods.

Each of the above patterns has advantages and disadvantages, which have to be taken into account in order to come up with the most suitable solution to be adopted in a data integration problem. A hybrid

approach that uses a combination of these integration styles would be able to handle the diversity of sources and overcome limitations that these sources might exhibit.

3.2.1 File Transfer

This approach relies on having a system (producer) to produce files that contain all the necessary information to be consumed by another system (consumer), which will read this file. This process implies that data has to be represented in a format that both systems can understand. In many cases, this might not be entirely possible, so it is the responsibility of the integrator to ensure that extracted files from the producer can be transformed to the format that the consumer expects.

Moreover, the nature of the data will dictate the time intervals of file generation, e.g. hospitalisation data might need to be produced on a daily basis. Once this is defined, the file transfer method between the two systems has to be also defined. They can be either applications that share the same storage or remote applications that will have to communicate the files via the Internet. The latter case reveals one of the possible drawbacks of the File Transfer approach, which is security. Files have to be sent in a secure way, which will prevent any possible loss of confidentiality or data integrity.

Another disadvantage is scalability. Data files usually need to contain big volumes of information, e.g. BAM files. This could make transferring take much longer to complete and therefore jeopardise the integration and co-operation with the other application(s). Finally, adaptability is another aspect to take into account, since emerging user requests for modifications or changing business needs can easily end up to a broken file transfer process, i.e. custom-made scripts enabling data transformation to the common transfer format need to be updated every time something changes.

Mitigation methods for the above problems can of course be planned and enable the establishment of a File Transfer process that can support the integration of applications that could otherwise not communicate to each other. This is in fact one of the big advantages of File Transfer, communicating applications need to have no information about the internals of each other and are entirely decoupled from each other. As long as the produced files follow the same format, the application can make internal changes without notifying or affecting the other applications.

3.2.2 Shared Database

A Shared Database gives the opportunity to all applications that have access to it to be always consistent and have immediate access to data updates. The data scheme and relationships are known to all the applications making in it quicker to develop functionalities on top of that scheme without the need to know how other applications interact with it. However, this flexibility can also be a problem for applications sharing a common data storage.

Coming up with a suitable design for the shared database that covers the needs of all applications is quite difficult. Furthermore, initial lack of required data structures will need database schema updates. Changes to the schema might take too long to implement as well, due to the impact they possibly have to other applications. Moreover, it might be even impossible to predict what functionality could be affected by a data schema update until this update is in place. Another problem can arise from an increasingly high rate of updates. Multiple applications reading and writing to the same database

increase the possibility for locks and deadlocks, thus making the database a big bottleneck of the system and eventually a single point of failure.

Working on a database copy and then applying the changes to the production instance reduces some of the risks but introduces other issues, like e.g. the database synchronization. Nevertheless, a Shared Database can be the solution for some cases where e.g., there is a limited number of applications that share the database and there is little or no probability of big changes to the data schema. In such cases, the Shared Database approach can be a fast to develop integration solution, which provides a reliable communication between applications.

3.2.3 Remote Procedure Invocation

The Remote Procedure Invocation applies the principle of encapsulation to integrating applications. The applications communicate directly with each other in order to exchange information via a predefined mechanism for common formatted messages. This mechanism usually relies on a Service Oriented Architecture (SOA), in which the applications provide a set of services to be used by other applications. This allows the applications to act as black boxes to other applications. They only need to define the expected input and output parameters of the exposed services and they keep the responsibility for maintaining their internal data integrity without any other application being affected. If one application wants to change the data of another application, it does so, by making the appropriate call to the other application.

The architectural approach that is mainly used in modern SOA is Representational State Transfer (REST). REST is an architectural style, and, while it employs standards, it does not constitute a standard itself. The main characteristics of REST web services are described below:

•*Client-Server Architecture:* The resources are located in a server and a client can request them through a REST web service;

•*Stateless:* No session state is maintained at the server; each request from client to server contains all of the information necessary to understand the request.

•*Cache:* The requested resources can be cached and therefore the client gets faster results;

•*Uniform interface:* REST operations use the Hypertext Transfer Protocol (HTTP) operations GET, PUT, POST and DELETE; thus, if a client can deal with hypermedia, it can start dealing with REST.

•*Named resources:* Every resource is accessed using a clear name, its Uniform Resource Locator (URL).

Using RESTful services to collect data from various sources can be a relatively simple and reliable integration solution, which is currently supported by frameworks in various programming languages, testing tools and design guidelines. It can offer scalability and does not require any heavyweight framework or toolchain to get started. On the other hand, breaking down complex functionalities coming from legacy systems in single HTTP messages could potentially require quite an amount of effort and redesign of the involved procedures.

3.2.4 Messaging Events and Messaging Channel

In order to facilitate real time communication of low level information from various (mainly Internet of Things) sources, one could employ the messaging integration pattern. In the e-health domain, a number of legacy medical devices may comprise such an information source, which in turn produce messaging events. The underlying communication paradigm is completely different from conventional, service-based architectures as they permit high-speed, low latency and high throughput, avoiding time-consuming communication handshakes.

A typical implementation of the messaging approach is via a messaging channel, which operates on TCP/IP enabled networks and is mainly supported by the messaging bus, an intermediary that supports multiple communication interfaces, each one defined by a protocol and a port. Messaging relies on transferring of data through a common channel in an asynchronous manner. Messages are sent from an application without knowing if the other application is up and ready at the same time. In the same way, an application that was not connected to the messaging system for whatever reason, might receive a message, which was sent some time ago right after it successfully reconnects to the messaging system. Asynchronous messaging strengthens the decoupling of the applications and allows integrators to choose between broadcasting messages to multiple receivers, routing a message to one of many receivers, or other topologies.

A messaging event is the object that is published to the messaging bus (channel, queue, peer-to-peer, or data group) and is passed on to subscribed consumers as required. An event may be a simple byte array or contain more complex structures, such as complex data structures, dictionaries, Google Protocol Buffers, JSON, or JMS events. The underlying messaging model is a publish/subscribe infrastructure [5], an asynchronous messaging model in which the sender (publisher) of a message and the consumer (subscriber) of a message are decoupled. These two objects simply share a common channel/topic. The publisher published data to the channel, which exists within the messaging bus. As messages arrive on a channel, the server automatically sends them onto all consumers subscribed to the channel.

A Messaging system usually involves:

- *Message Channel/Topic:* the way that two applications can message each other;
- *Message BUS:* performs the processing steps to send a message to the proper receiver(s);
- *Message Translator:* converts the data into proper formats;
- *Message Endpoints:* the way to connect to a message channel;
- *Message Body:* the actual pieces of information exchanged among applications;

As message events are handled by the message bus via topics, a set of preliminary topics is created to separate data from different sources into logical channels rather than using filtering. In addition, separated data can utilize DataStreams for a single target, which facilitates the communication between different platform modules but also reduces the communication overhead as modules only receive data relevant for their proper operation.

3.3 Standardised efforts in integrating health data

The BD2Decide platform integrates tools and services for the analysis and processing of multisource, multiscale and multivariate patient and population data. This data is multisource, because a number of information sources is involved in the collection phase, ranging from user input at the time of first visit up to the analysis of images and tissue samples. This data is multiscale, because it refers to the different levels of the human body construction and behaviour, such as pathology data, biomarkers, and alcohol consumption. Then, this data is multivariate, because the BD2Decide environment exhibits different variables, through which the data can be considered and processed. Such variables include the refresh rate of data values along the HNC treatment process and the type of population data that should be integrated in the prediction models to give a better prognosis of the patient life expectancy rate.

In order to collect such data from existing systems, a standards-based approach must be employed, which considers the existing practices across the enterprise health systems for managing patient records worldwide. The Integrating the Healthcare Enterprise (IHE) group has worked towards this direction. Thus, it has introduced a set of profiles, in order to guide the standards-based access to enterprise health systems. These profiles are technically realised through technical frameworks, which can relate to one or more domains in the health sector. For the BD2Decide case and the implementation of the data acquisition and management platform, we focus on the IHE IT Infrastructure profiles, and especially the Cross-Enterprise Document Sharing (XDS) Integration Profile[2], which is now in Release 4.0, since 2016. This release consists of three different volumes describing the structure of the integration profile, the transactions that can be established in this profile and the respective metadata. This integration profile is compliant with the underlying HL7 standards[3] for Electronic Health Records (EHR), such as the Clinical Document Architecture (CDA), and the DICOM standards[4] from the Medical Imaging and Technology Alliance.

3.4 Current approach in the development of the platform

Due to the current restrictions applied in the participating clinical organisations, access to the internal enterprise health systems is not allowed to the technical partners of the BD2Decide project. To this end, up to now, the implementation of the IHE profiles for accessing the patient data of the BD2Decide HNC cases, both the retrospective and prospective ones, has not been automated. Instead, a manual approach has been followed.

More specifically, the project has adopted and customised an e-CRF based software for managing clinical data about the HNC patients (see Section 4.4.1). Both single and batch import processes of patient data are available. The data is collected from the representatives of the clinical centres in their

[2] https://wiki.ihe.net/index.php/Cross-Enterprise_Document_Sharing

[3] http://www.hl7.org/

[4] https://www.dicomstandard.org/

internal NAS systems and the provided software guides them on how to upload the anonymised data into the BD2Decide platform.

Regarding imaging data, again the available (and most relevant for the project) images are selected in the form of CT, MRI or DWI-MRI scans. The representatives of the clinical centres use the local NAS to store these images and their analysis, using the BD2Decide imaging analysis tools of Figure 1. Once the images have been analysed, the BDI offers the appropriate mechanisms, like the ones described in Section 3.2.3, so that the images and their analysis are uploaded on the BD2Decide platform.

As for the tissue samples, the project has formalised the procedure for transferring the samples to INT for analysis. The role of the BD2Decide data management and acquisition platform starts once this analysis has been finalised. Both the resulting genetic BAM files and their genomic analysis are maintained in the local NAS of each centre, while, as it happens for the images, the BDI offers the appropriate mechanisms (see Section 3.2.3), so that these results are uploaded on the BD2Decide platform.

As part of the forthcoming BD2Decide deliverable D5.5, we will provide further details on how the data management and acquisition platform can be automated and be fully interoperable with the existing enterprise health systems that are currently operating in the hospitals and the clinical centres. This is a challenging approach for the ongoing research in the field of big data analytics for the health domain and it has been pointed out by several relevant organisations and initiatives, like the Big Data Value Association (BDVA) [6].

3.5 Data Management Tool

This tool allows the hospital data managers or administrators to control the data collection process and perform the data quality check. This tool allows data managers to work with the data available in the BDI (Figure 5) without the need for learning any technical language.

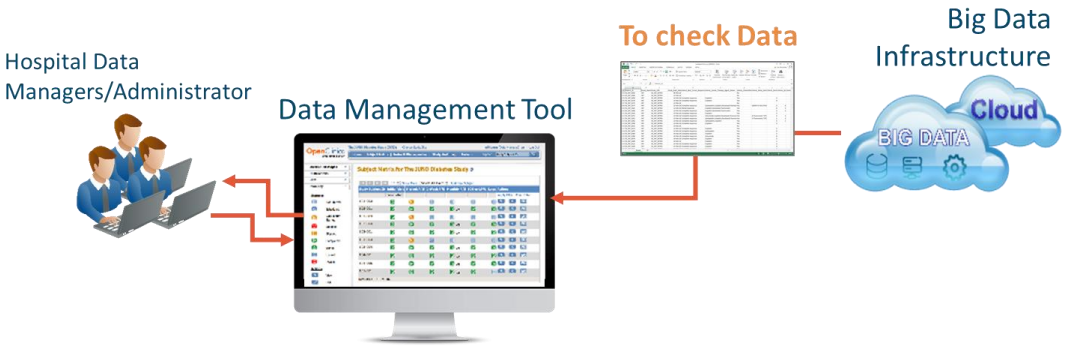

Figure 5: Data management tool scheme.

Among the features already defined, the users can download the complete BD2Decide patients' database. Users have also available a summary table that reports the status of the data insertion process in the BD2Decide system, as shown in Figure 6. The list of the summary data collection has been defined with the corresponding partners. Additional features will be included following further

interactions with the clinicians, to provide a complete functionality to ensure that data collection complies with expected data quality.

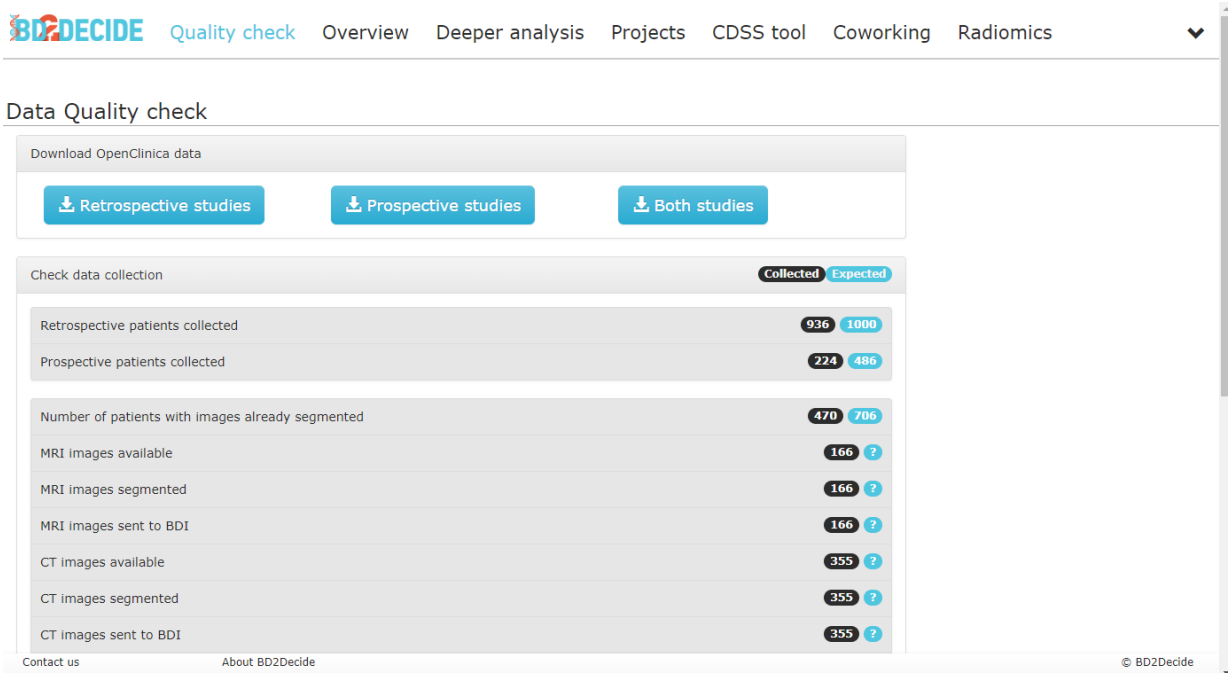

Figure 6: Data quality check. OpenClinica dataset download and patient collection data summary.

A specific functionality to allow some data managers to report the status of data collection or quality will be included. This will serve mainly the genomic and imaging data collection confirmation.

4 FUNCTIONAL VISUALISATION MOCKUPS OF THE CLINICAL DECISION SUPPORT SYSTEM

As mentioned in Section 2.3, CDSS intends to support clinicians and other doctor specialties in managing the collection, processing, storage and visualization of the HNC patients' data and their assessment. In order to do so, this tool integrates various components and services, which offer advanced functionalities for the processing of such data towards making prediction analysis and prognosis of the life expectancy rate for an HNC patient. In this section, we analyse the CDSS functional mock-ups through the expected scenarios of use for this tool.

4.1 Overview of scenarios of use

In order to demonstrate the use of CDSS, we have defined a set of scenarios, in which the health professionals, as the target CDSS end users, interact with the tool, through the screens that we have already implemented. These scenarios aim to highlight the use of CDSS in different phases of the HNC treatment process.

It must be noted that the definition of scenarios has been used as part of the user-oriented design in CDSS to come up with the final set of functional screens. During this process, we have presented to the evaluation team a set of mock-up screens that implement the expected BD2Decide functions. This process has resulted to a better understanding of the clinicians' needs and their expectations from CDSS. More details on this process and the results of the various evaluation iterations will be provided in the context of the BD2Decide Deliverable D5.7 "Usability evaluation results", which is due M36 (December 2018).

4.1.1 Exploring patient's data

The first scenario relates to the exploration of the HNC patient data on the individual level. The aim of this scenario of use is to allow the clinicians to have direct access to the low-level patient data at any time. More specifically, in this scenario, the health professionals are authenticated into CDSS to be able to browse both the raw and analysed information about their HNC patients. Through this scenario, the clinicians may access mainly to three blocks of data, namely the e-CRF data, the Quality of Life (QoL) data and the image data.

For the e-CRF section, the clinicians can have access to data with respect to the following patient information:

- Demographic data, like the gender, the age, the ethnicity, etc.;
- Clinical data, like the time of first diagnosis, the HB level, the PLT level, the value of lymphocytes, etc.
- Risk factor data, such as smoking and alcohol habits and the overall hygiene view of the patient.
- Data with respect to the clinical T- and N-characteristics, which describe the morphology of the tumor and the nodes.

- Pathology data, like HPV, etc.
- Treatment data after following chemotherapy or radiotherapy.
- Information on potential surgery treatment followed.
- Details about the follow-up period, such as reoccurrence, etc.
- Any toxicity data, which describes the impact of radiotherapy or chemotherapy on the patients' clinical status.

For the QoL data, the clinicians can explore the patient's responses to three standardised QoL questionnaires, namely:

- The EQ-5D-5L questionnaire developed by the EuroQol Group[5] for the measurement of health outcome;
- The QLQ-C30 from EORTC[6], which is a questionnaire developed to assess the quality of life of cancer patients;
- The QLQ - H&N35 from EORTC, which is a questionnaire developed to assess the symptoms or problems arisen from a treatment followed by cancer patients.

For each of these questionnaires, this scenario of the CDSS use envisions that the clinicians will be able to browse the individual responses to the questionnaires and for the different periods across the HNC treatment process. The tool should also allow them to visualise the individual scores per questionnaire and compare them to the average QoL score of the cohort.

The third block of data refers to the patient images. In this case, the clinicians should be able to browse the diagnostic images (CT, MRI and DWI-MRI scans) that have been analysed for the selected patient. Such images have been analysed using the BD2Decide Imaging Analysis Tools. Thus, CDSS should be able to present analysed segments to the clinicians, especially those that are the most relevant for the HNC case of this patient.

4.1.2 Assessing the patient survival rate

The second scenario of the CDSS use refers to the ability of the health professionals to exploit the collected and/or analysed patient data and evaluate the patient's individual survival probability. This is a risk stratification scenario that, currently, the health professionals try to address using the past research or bibliographic sources and not on a patient level. With BD2Decide, this scenario allows for a more personalized risk stratification, which is enabled through both standard statistical and big data analytics methods.

More specifically, CDSS is used to assess the individual patient's risk through different models. The tool should allow the clinicians to select among various predefined statistical models and evaluate the

[5] http://www.euroqol.org/eq-5d-products.html

[6] http://groups.eortc.be/qol/

evolution of the survival probability of the patient in time, compared to the one for the cohort of patients, belonging to similar cases. Apart from these statistical models, CDSS is configured to present additional survival probability curves, which are constructed through big data analytics techniques from the processing of the datasets available in the big data infrastructure. The way that these curves are built is explained in the BD2Decide Deliverable D6.2, which is due M29 (May 2018).

It must be noted that the survival probability curve, either being built from a statistical model or through big data analytics, is based on a set of prognostic variables. These factors may have different influential factor on the construction of the curve. By allowing the clinicians to compare the outcome of different curves, the CDSS gives them the ability to assess the impact of each prognostic variable on the selection of the appropriate treatment.

4.1.3 Organising a Tumor Board

An integral part of the HNC treatment process is the organization of a tumor board. This is another feature offered in CDSS, which aims to extend the previous scenarios of use to a more collaborative fashion. The aim of this scenario is to support the organization of an online tumor board and manage the recording of the decisions made for each patient. It is clarified that the execution of an online tumor board, in the mode of a teleconferencing system, is out of scope of this scenario of use.

In more details, in this scenario, CDSS is used to schedule a forthcoming tumor board and build the agenda of it. The latter consists of HNC patient cases for which a treatment decision is pending. Once a board is scheduled, CDSS can support the health professionals, who are participating in the board, in having access to the most relevant information about a board case. To this end, CDSS provides a chatting feature for enabling clinicians in sharing opinions and discussing about a patient case. Once a consensus is achieved, CDSS manages the sealing of the treatment decision for the specific patient.

4.2 Visualising the scenarios of use

In this section, we present a step-by-step guidance on how the health professionals can access CDSS and use the provided visualisations to accomplish the intended scenarios and tasks presented in the previous Section 4.1.

4.2.1 Entering the CDSS environment

CDSS is accessible through a Web browser. One can navigate to cdss.bd2decide.eu and fill the authentication form with the provided username and password, as shown in Figure 7, in order to log in to the tool.

In case of wrong credentials, the platform will inform the user accordingly (see Figure 8).

After the successful login, one will be redirected to the home page of the CDSS environment (see Figure 9), giving an overview of the data for the patients assigned to the user. The level of access rights is determined by the role assigned to the user.

In this screen, the user can also navigate to some graphics, summarising some basic statistical information about the cohort that the user has access.

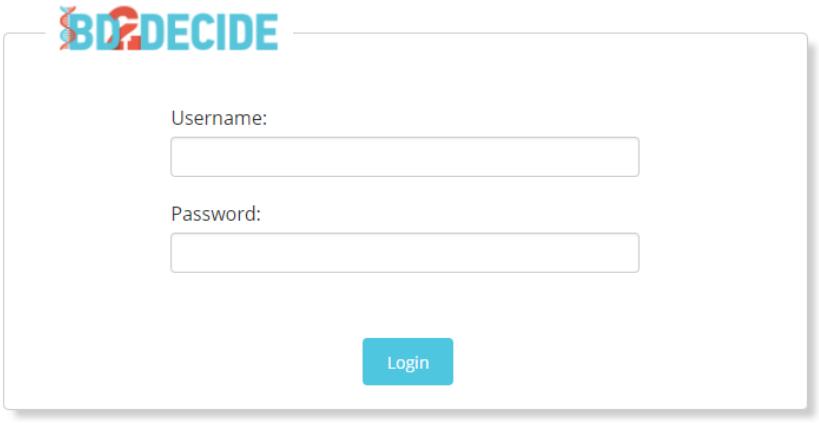

Figure 7: The CDSS login page.

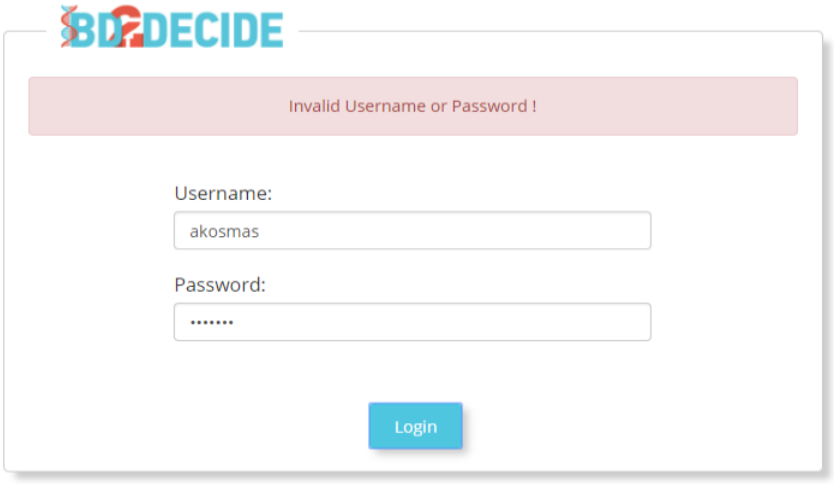

Figure 8: The CDSS response to invalid credentials during the login process.

Through the menu bar of the home page (see Figure 9), the user can have access to the individual clinical records of the patients and the collaborative tumor board functionalities.

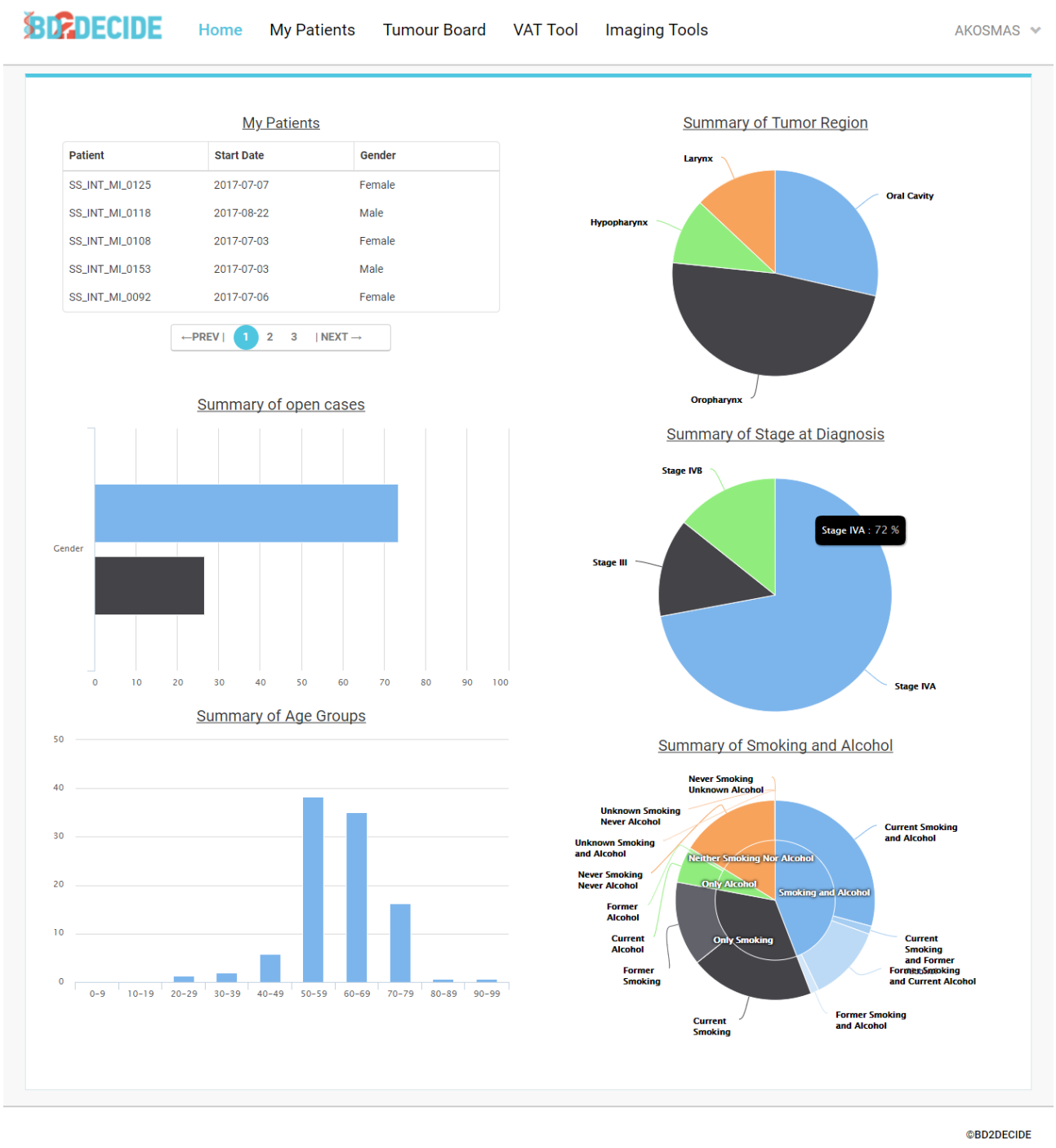

Figure 9: The CDSS home page.

4.2.2 Exploring patient's data

From the menu bar, one can select the "My Patients" menu item in order to navigate to the full list of HNC patients assigned to the user role. As shown in Figure 10, this page summarises some basic details about the patients, like the patient ID, the date of first entry and the gender, in a tabular form,

while it provides a button for each patient, so that the use can explore the clinical information and data registered for this patient.

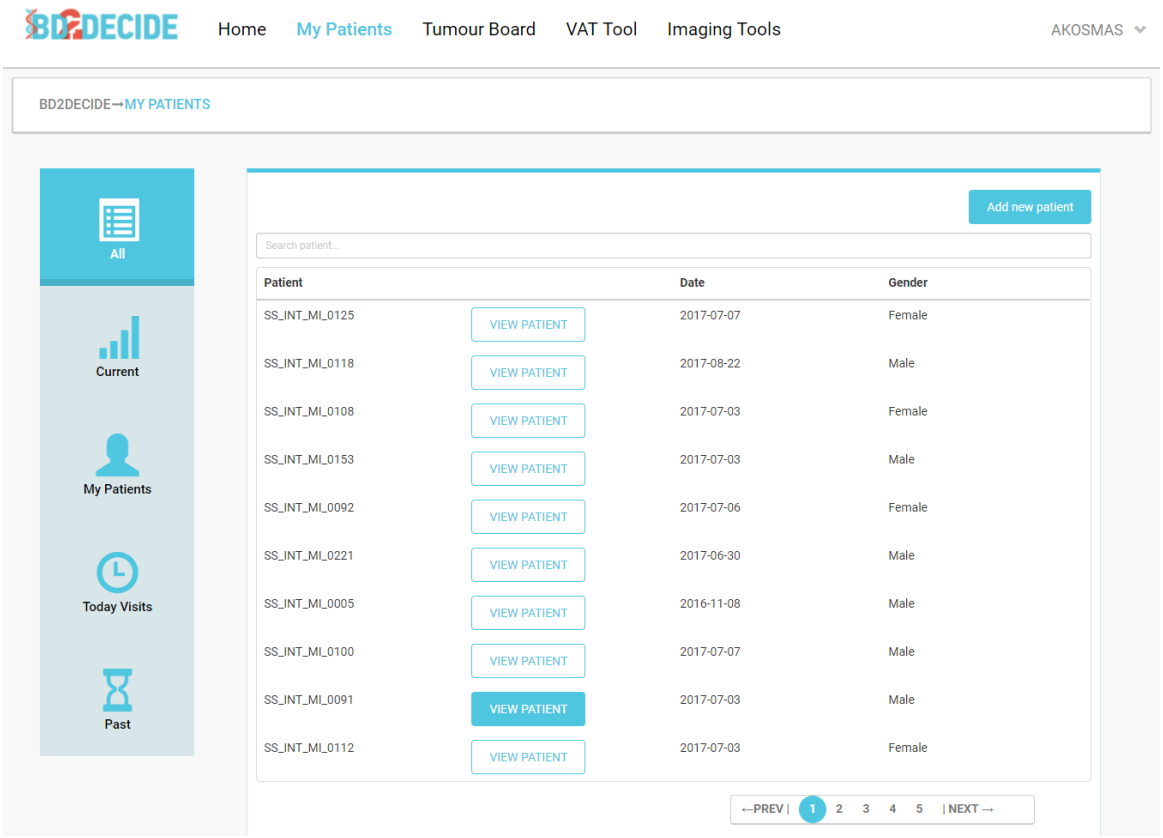

Figure 10: The CDSS patients' overview page.

It is noted that the user is always instructed on the current page through the breadcrumb like text box, just below the main menu bar.

Once the user clicks on the "View Patient" button, the screen of Figure 11 is loaded. The default view navigates to an overview page that displays the main clinical related information about the selected patient. We highlight that this view also represents the logical structure of the CDSS screens for the implementation of the first two scenarios of use, as they were summarised in Sections 4.1.1 and 4.1.2. Thus, just after the breadcrumb like text box, the user is offered with tab like choices to select the specific sections that one wants to access. For each section, the CDSS view is split into three parts:

i. The top left being marked with the red line comprises the patient quick label, listing the patient ID, the gender, the age, the smoking and alcohol consumption habits, the tumor region and the stage at diagnosis.

ii. The left panel allows the users to interact with the currently selected section. The interaction takes the form of either selecting an option to change the view of the section on the right or an action that needs to be taken by the user (i.e. upload an image)

iii. The right panel presents the results of the interaction of the user with the features offered in the left panel.

On top of that, the breadcrumb like text box offers the user the option to access the IPDA tool. More details are provided in Section 4.4.3.

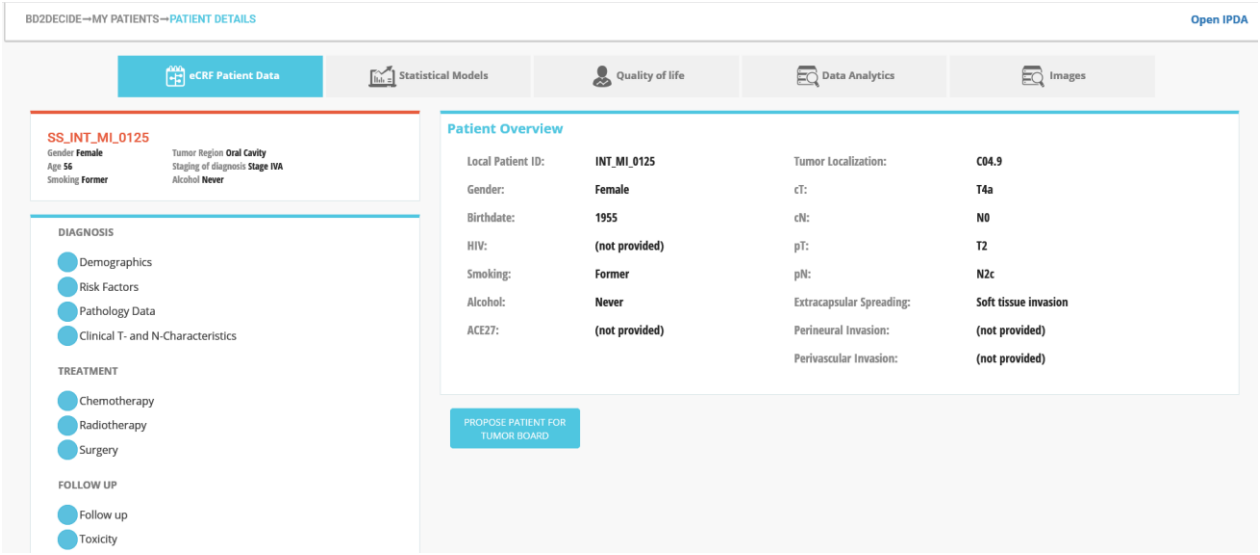

Figure 11: The individual patient overview page in CDSS.

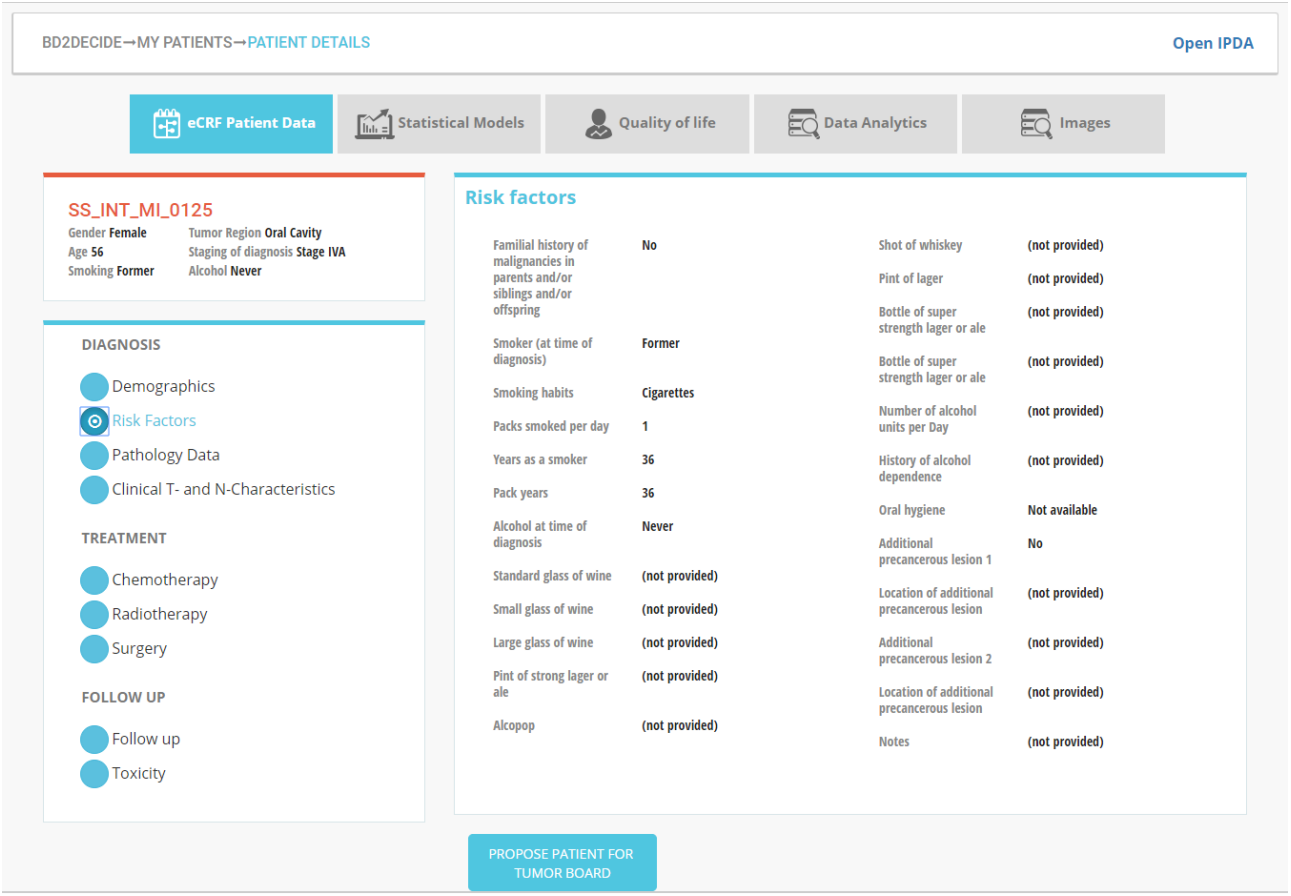

Figure 12: An example of the CDSS view for presenting the risk factor data collected for the selected patient.

By default, the patient overview page, displayed in Figure 11, is logically categorised under the "eCRF patient data" section. For all the pages under this section, the user, as the physician of the patient, has the option to propose this patient for a tumor board discussion, based on the scenario presented in Section 4.1.3. The relevant screens will be presented in Section 4.2.4.

One can use the left panel to further explore the patient's data, which are grouped according to the e-CRF structure. Thus, the user can select the desired category and view the collected data for the patient (through the process that is described in Section 4.4.1). For example, if the user selects the "Risk factors" category, then the screen in Figure 12 is presented. Any value that has not been inserted into the BD2Decide environment is marked with the label "(not provided)".

A second block of data relates to the patient QoL. This dataset is accessible by selecting the section "Quality of life". The default view of this section is presented in Figure 13. As shown there, the right panel lists the available questionnaires for the selected patient per period (at the time of diagnosis, and six, eighteen or twenty four months after treatment) in a tabular format. Below this table, the user can see the evolution of the score for the three questionnaires in time, both for the patient individual responses and the cohort as a whole.

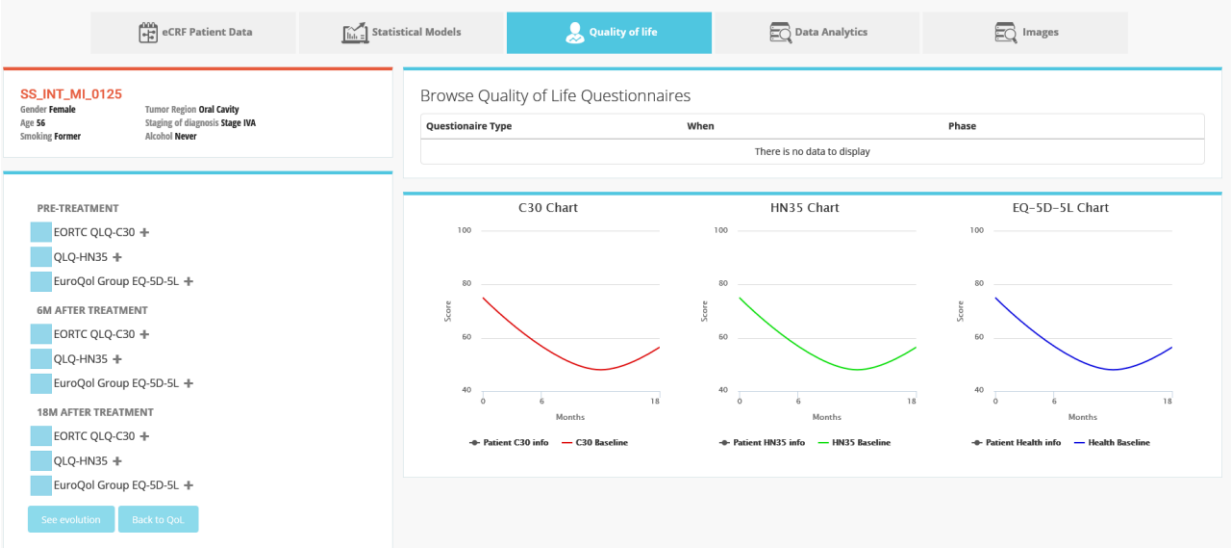

Figure 13: The default view of the QoL section in CDSS (no QoL data available).

In Figure 13, the selected patient has not completed any questionnaire, so the right panel is empty. In order to access the questionnaires, the user should use the left panel and click on the "+" symbol next each questionnaire. This navigates the user to the process, which is described in Section 4.4.2.

An example of a patient with completed QoL data is presented in Figure 14. In this case, the left panel on the default page of the "Quality of life" section does not include the "+" symbol (since the patient has already completed the questionnaires), but CDSS allows the user to select to view the different types of available responses and their evolution in time.

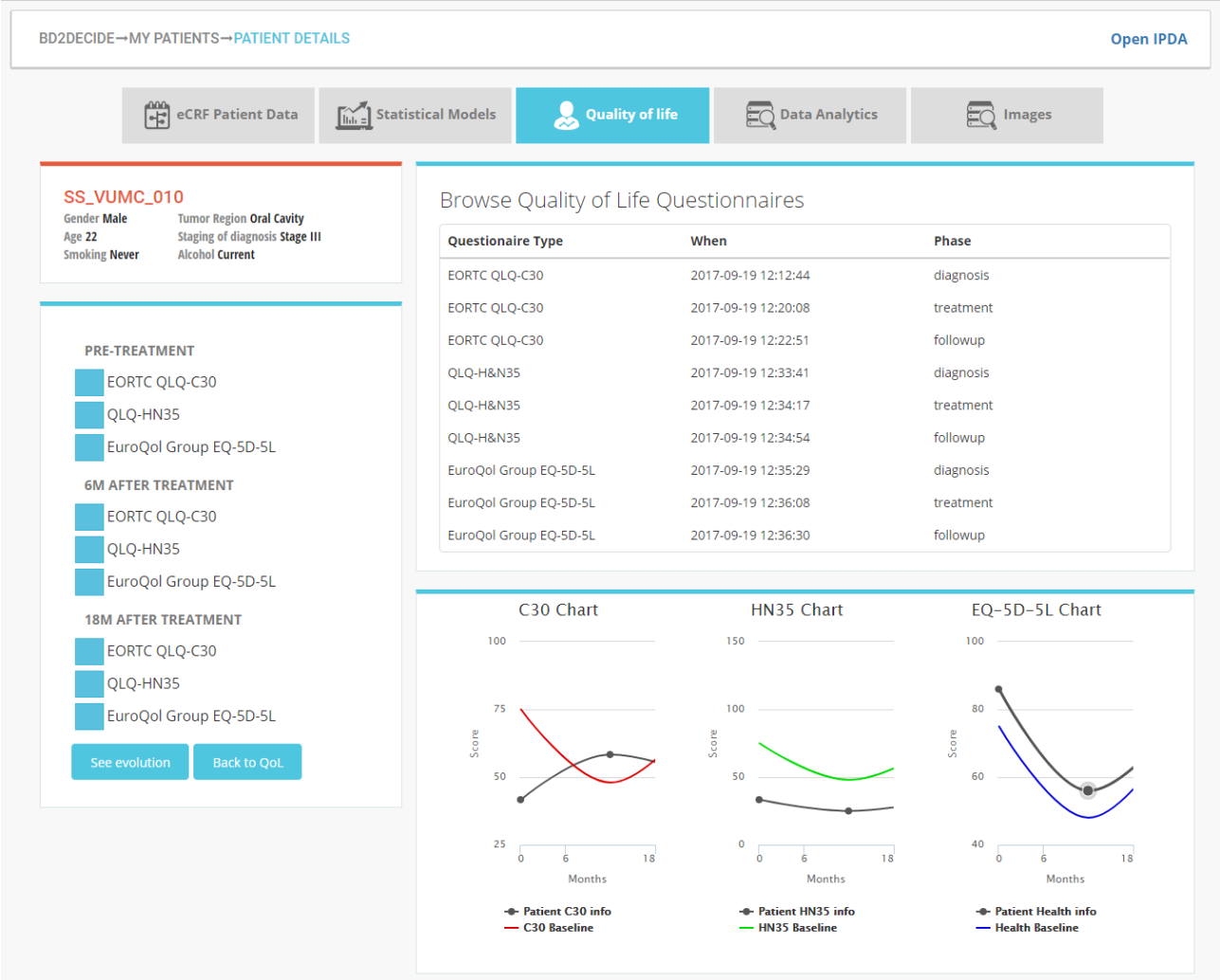

Figure 14: The default view of the QoL section in CDSS (with mocked QoL data).

By selecting one or more questionnaires, the user is navigated to the screen of Figure 15. Here, the user can browse through the individual responses of the patient in a certain questionnaire and their score.

The third block of data that belongs to this scenario relates to the patient images. In order to access the imaging data, one should select the “Images” section. By browsing this section, the user can see the list of images taken from this patient, either through the radiologists (CT, MRI and DWI-MRI scans) or by the physician through a mobile camera (see Figure 16). The right panel of this screen is, subsequently, split into diagnostic and doctor images. The latter, taken from the physician via a camera, can be uploaded on CDSS using the “image upload” button on the left panel of the screen.

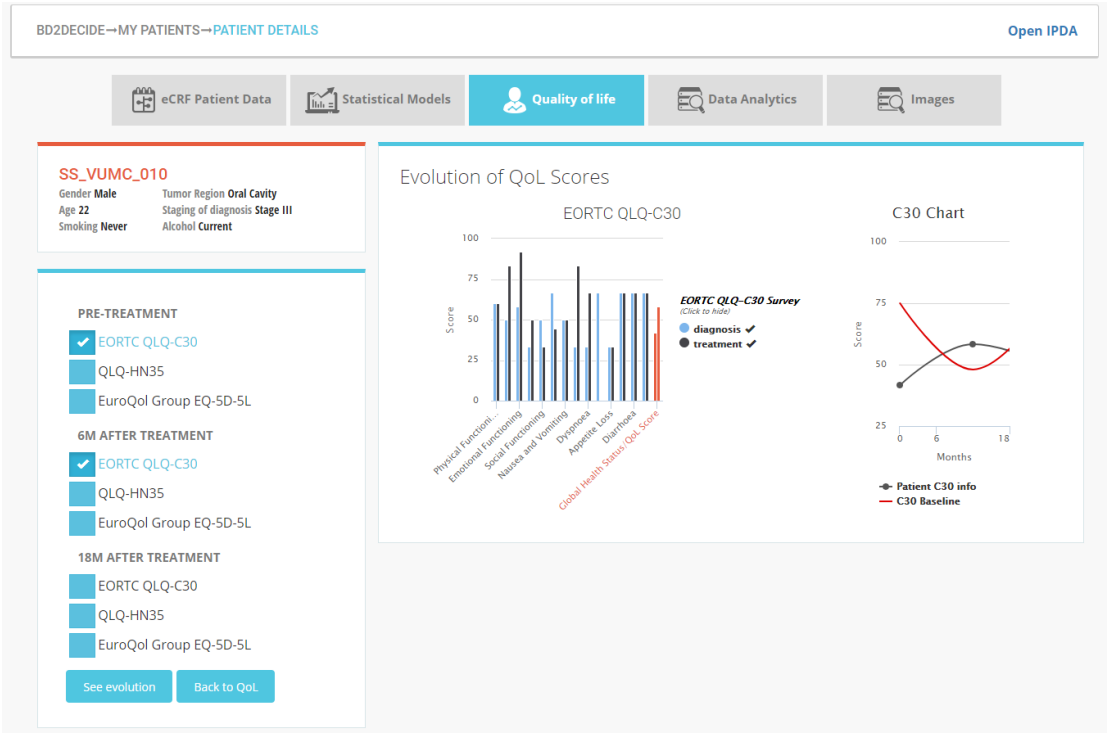

Figure 15: The detailed view for navigating into the quality of life questionnaires in CDSS.

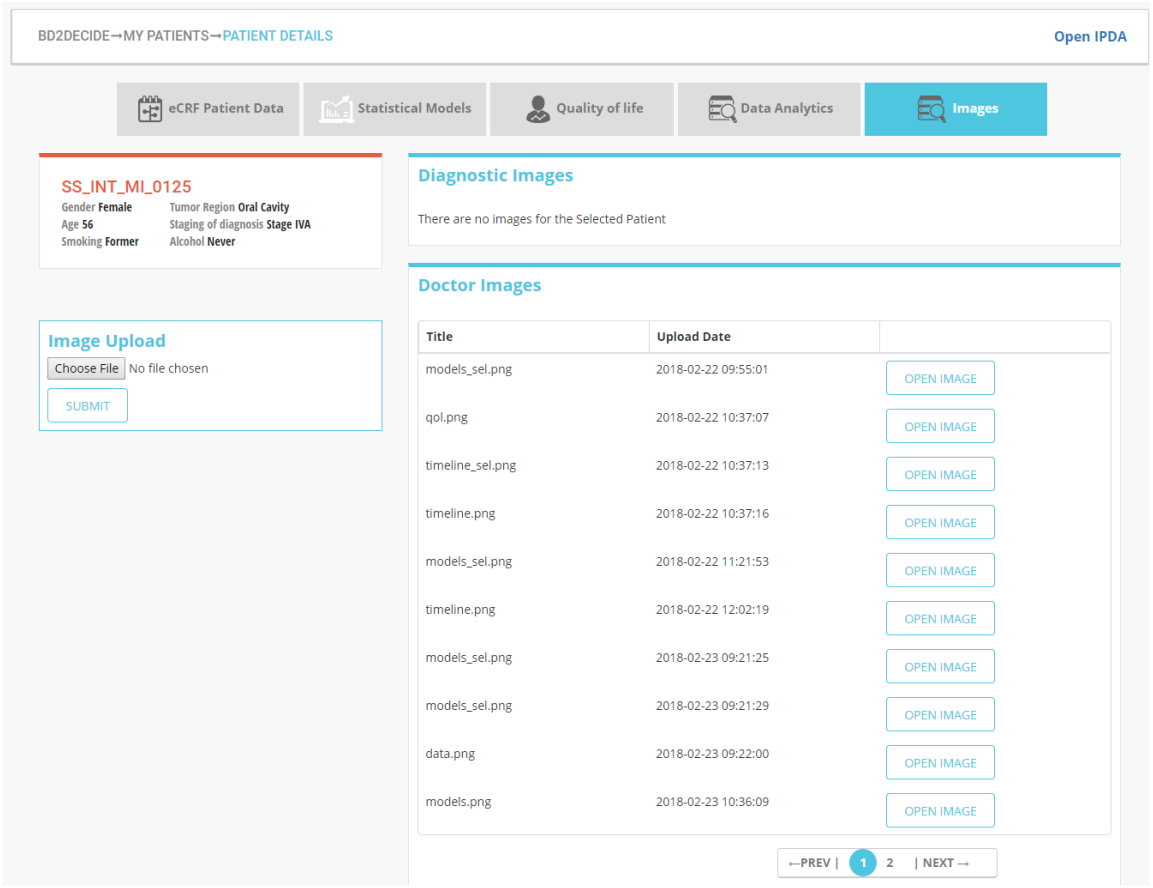

Figure 16: Browsing the imaging data in CDSS.

4.2.3 Assessing the patient survival rate

The scenario of Section 4.1.2 refers to the assessment of individual patients' cases with respect to similar cases observed in the past. In this scenario, the user navigates through different survival probability curves, which are produced either from statistical models or the analysis of past patient and population data, using big data analytics techniques.

One can access the environment for the risk stratification by using the section tabs under the "My Patients" menu item. Two options are available. The first one refers to the assessment of the patient risk using the statistical models and is selected by clicking the "Statistical Models" section. The second option refers to the risk stratification using big data analytics techniques and is selected by clicking the "Data Analytics" section. It is noted that, as we explained for the other sections in the "My Patients" menu item in Section 4.2.2, the screens for this scenario are split into three parts, with the top left part comprising the patient quick label, the left panel being the controlling part to allow user interactions and the right panel presenting the results of these interactions.

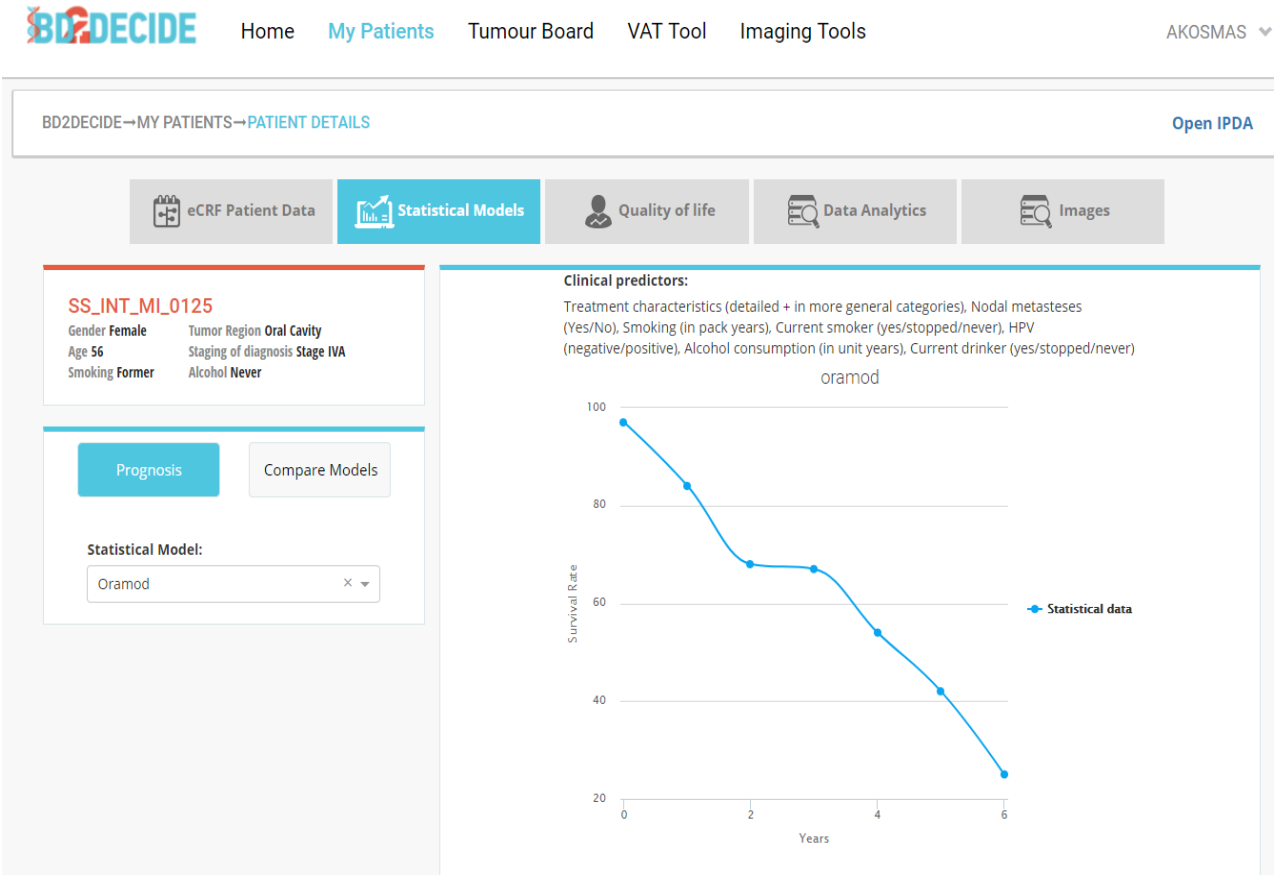

Figure 17: Accessing the patients' risk stratification results through statistical models in CDSS.

When the user selects the "Statistical Models" section, CDSS loads the screen of Figure 17. In this figure, the left panel allows the user to switch between two different features. The first feature refers to the presentation of the prognosis for the patient individual survival probability, by allowing the user to select a statistical model from a predefined list. This is illustrated in the right panel of the

screen in Figure 17, in which the user also receives the set of clinical predictors that affect the development of the curve.

Currently, the results presented in this figure are based on mocked data. In the next version of this illustration, the user will be able to receive the high, medium and risk probabilities for the individual patient's prognosis.

In case that one wants to compare the results from two different models, then the user clicks on the "Compare Models" button on the left panel and select the models for comparison. An example is presented in Figure 18.

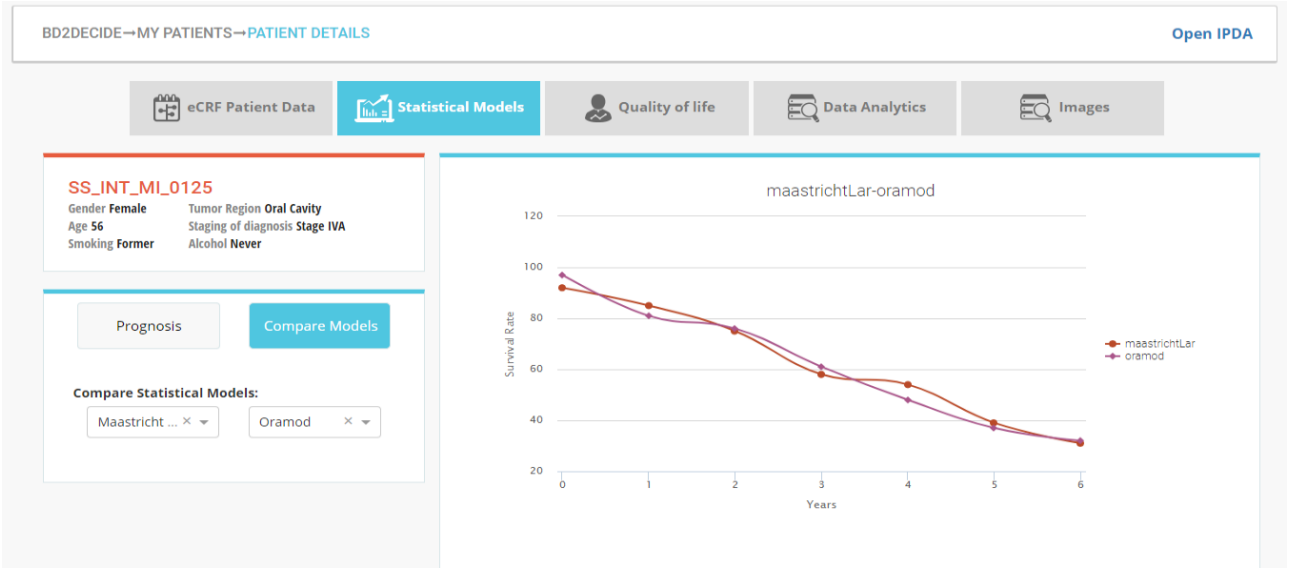

Figure 18: Comparing the survival probabilities for two different statistical models in CDSS.

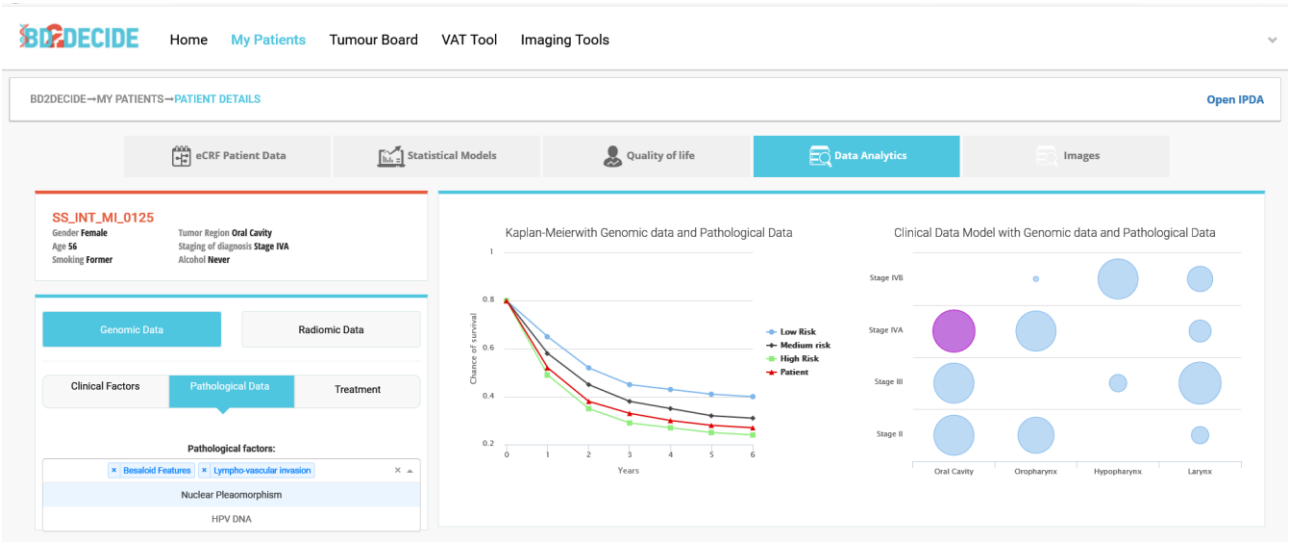

Figure 19: Accessing the patients' risk stratification results through big data analytics in CDSS.

Moving to the "Data Analytics" section, the user is recommended the risk stratification for the specific patient, based on the big data analytics techniques. As shown in Figure 19, the risk

stratification is illustrated in two different views on the right panel. The first one is similar to the presentation of results using the survival probability curve, as it happens for the statistical models in Figure 17. In this view, the user is presented with the high, medium and low risk curves of the cohort of patients belonging to similar cases and the individual prognosis for the patient. The second chart classifies the patient in the whole patient's cohort through different variables. In the example of Figure 19, we are using the tumor region and stage for this classification, while the size of the bubble declares the size of the group. It is noted that, in Figure 19, we are using mocked data.

On the left panel, CDSS provides a predefined set of categories of clinical data variables, like genomic and radiomic features, clinical factors, pathology data and treatment data. When these variables are selected, it means that they are considered in the development of the survival probability curve, which is presented on the right panel, using the big data analytics techniques. Both radiomic and genomic features are single buttons, while the other three categories are expandable to indicate the specific variables, which are considered in the curves for the category. In the example of Figure 19, the genomic features and specific pathology data (the baseloid features and lympho-vascular invasion) are the dominant predictors for the recommended survival probability curve of this patient. Of course, the user has the ability to select additional categories and variables from the left panel and check the impact on the curves presented on the right panel.

4.2.4 Organising a Tumor Board

Following the description of the scenario that we described in Section 4.1.3, in order for the clinicians to organize a Tumor Board, they first need to propose patients for this board. As we explained in Figure 11, this is enabled by selecting a patient and then clicking the button "Propose Patient for Tumor Board" in the "eCRF Patient Data" section of the "My Patients" menu item. In case that the selected patient has already been proposed for a tumor board, instead of the button, a message informs the user that the patient has already been proposed for it or the decision has already been taken, as shown in Figure 20.

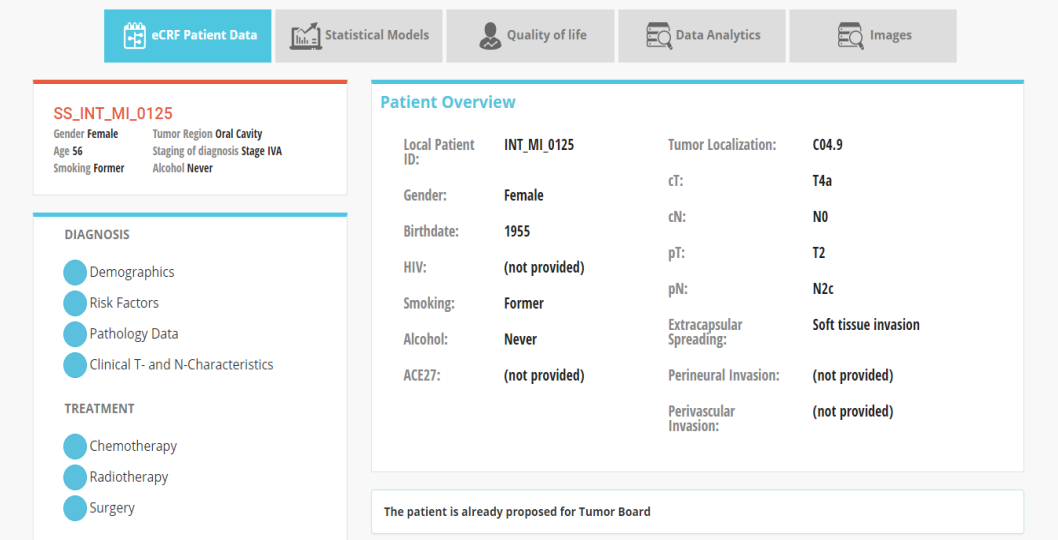

Figure 20: Informative message in CDSS that the patient has been selected for a tumor board.

Figure 21 presents the home page of the TBCE, which is enabled by selecting the "Tumor Board" item in the menu bar. This is comprised of three sections, namely:

- The "Schedule Boards" section, which is used to organise a tumor board, based on the patients that have been selected for it.
- The "Planned Boards" section, which lists the boards that are already planned and through which the user can launch one of them.
- The "Past Boards" section, which lists the boards being executed in the past, along with the patients, for whom a decision has been made in these boards.

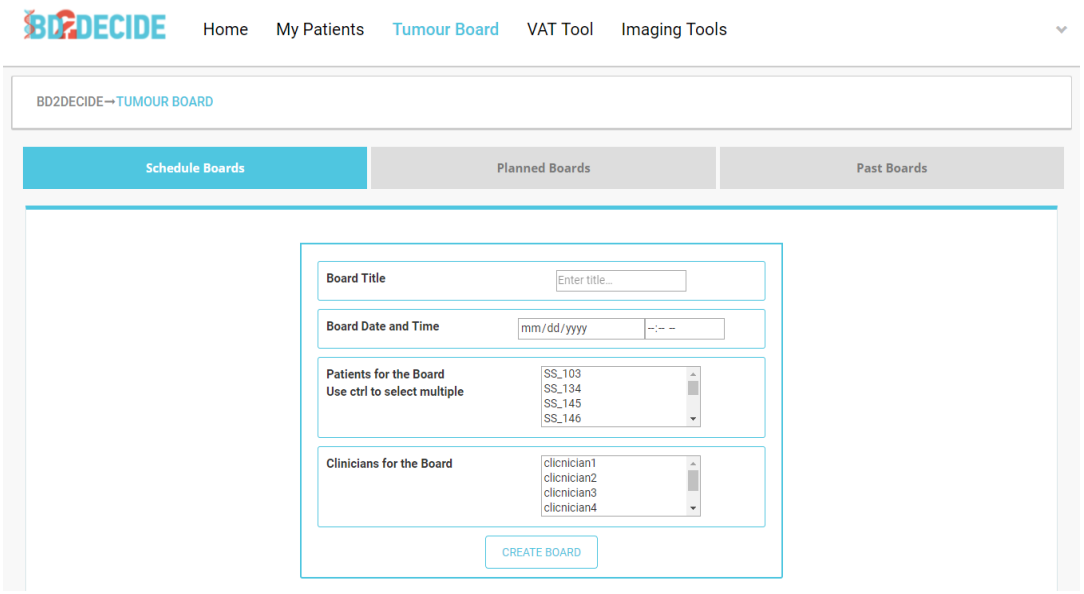

Figure 21: The home page of the TBCE in CDSS for the organisation of a new tumor board.

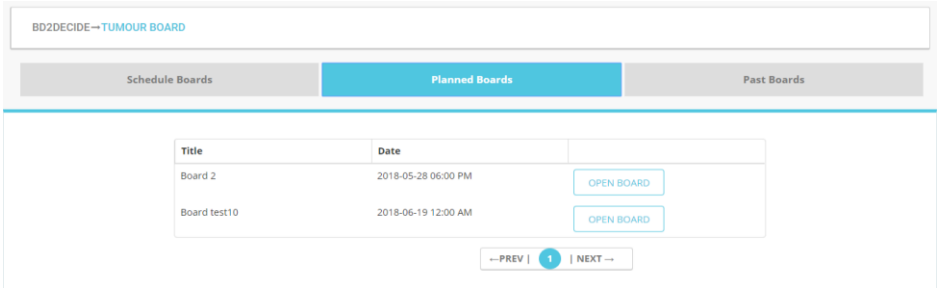

Figure 22: Selecting a board to participate in CDSS.

In Figure 21, we present the fields that the user has to fill in so that a tumor board is scheduled. As shown there, the user is prompted to provide the title of the tumor board, the proposed date and time for organising it, the list of patients that have been proposed for a tumor board and the list of candidate participants (clinicians and other health professionals) in this board.

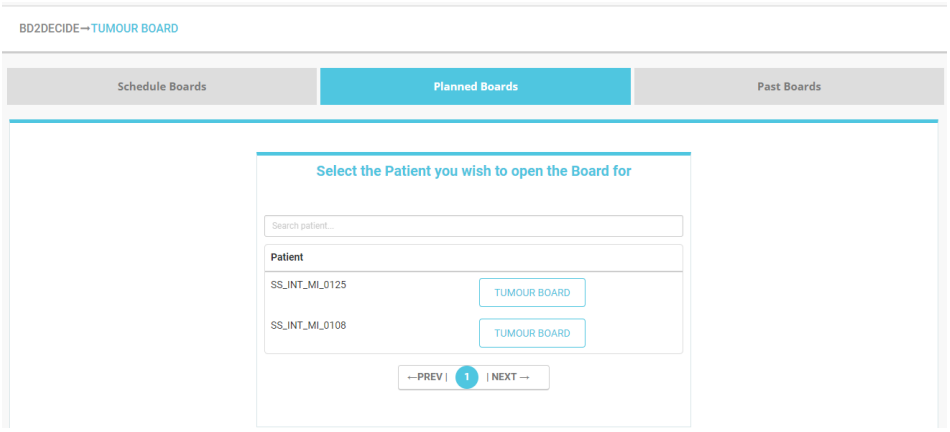

Figure 23: Selecting a patient to start the discussion in the CDSS tumor board.

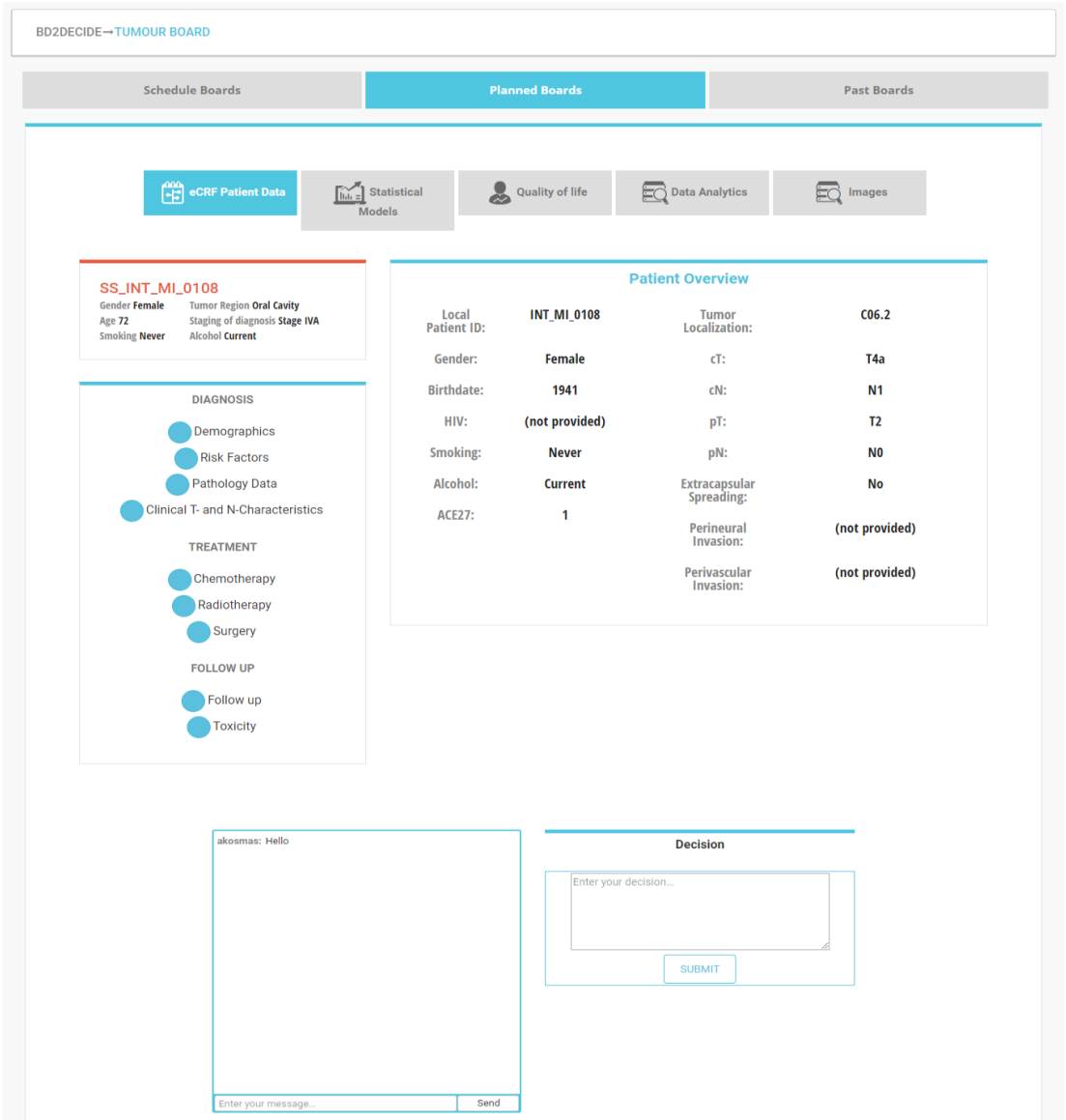

Figure 24: Discussing a patient case in the tumor board in CDSS.

Once the user clicks on the "Create Board" button, the TBCE turns to the second section with the list of planned boards, as shown in Figure 22. Then, the user is prompted to select the patient case that will be discussed in the context of this board (see Figure 23). This loads the screen in Figure 24.

Figure 24 resembles to the "My patients" pages, as it consists of the different sections containing the information about the patient, but in a limited version. Along to it, this page loads a real time chat functionality that allows the participating clinicians to remotely discuss about the patient case. The outcome of the discussion is recorded as the tumor board decision, which is done through clicking the "Submit" button.

In order for the user to browse a decision of a tumor board occurred in the past, one can click on the "Past Boards" section and then select the patient. This directs to a page containing the board decision for the selected patient, as shown in see Figure 25.

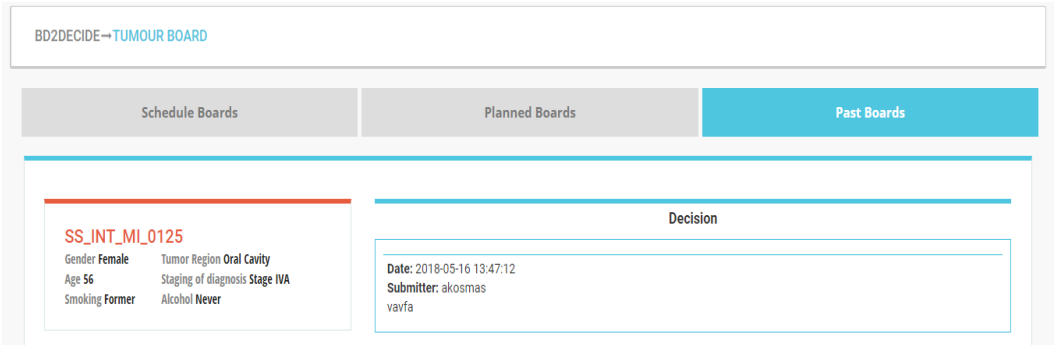

Figure 25: Browsing the decisions of past tumor boards in CDSS.

4.3 Implementation technologies

This section describes the tools and technologies that we used in the development of the BD2Decide CDSS components.

The core of the CDSS front end has been developed in React. React js[7] is a javascript library for building user interfaces, introduced for the public in 2015. React js can be paired with other javascript libraries like Redux[8], a fact that contributed in gaining a lot of popularity and becoming one for the first trends for web development. The version used is 15.6.2.

In order to implement the backend services of the tools, Java[9] and Laravel[10] were used.

Java is an open source, object oriented programming language, designed to run in all platforms and with as few implementation dependencies as possible. Version used for this project is 1.7.0_80.

[7] https://reactjs.org/

[8] https://redux.js.org

[9] https://go.java/index.html

[10] https://laravel.com/

Laravel[11] is an open source php web framework that allows the creation of web applications following the Model-View-Controller architectural pattern. The version used is 5.2.

Tomcat[12] is an open source java servlet container, which provides an HTTP server environment for java web applications and services to run. The version used is tomcat 7.

Apache[13] is a free open source web server that can run in multiple platforms and can support many features like url rewriting, certificate authentication etc which makes it a valuable tool for all web applications. Usually it is used in front of an application server in order to apply forwarding or restrictions rules to the access of web resources.

4.4 Integration with other tools

4.4.1 Adding eCRF data into the BD2Decide

The part of the PDS component, which is responsible for the collection of the e-CRF-based patient data has been developed in the OpenClinica[14] tool. This is used in order to allow the clinicians to enter the patient's data into the system. OpenClinica is a clinical data management tool, open source and web based that is widely used from clinicians all over the world. The version used is 3.10.1.

In order to store the patient's data the PostgreSQL[15] database is used. PostgreSQL is a reliable and secure database that offers a wide range of features and is widely supported by many tools and programming languages. The version used is 8.4.22.

Access to OpenClinica

In order to access OpenClinica, one should follow these steps:

- Open an Internet browser;
- Type http://openclinica.bd2decide.eu/;
- Enter the user name and password that one have been given and click on 'Login.

[11] https://laravel.com/

[12] http://tomcat.apache.org/

[13] https://httpd.apache.org/

[14] https://www.openclinica.com/

[15] https://www.postgresql.org/

Figure 26: OpenClinica login page.

After successful login, the user will be redirected to OpenClinica home page for the study assigned to the user.

After successful login, the OpenClinica home page appears, like the screen in Figure 27. This page shows an overview of the selected study, while it provides the menu for accessing, among others, the Subject Matrix, which gives an overview of all patients (subjects as they are called in OpenClinica) in the study.

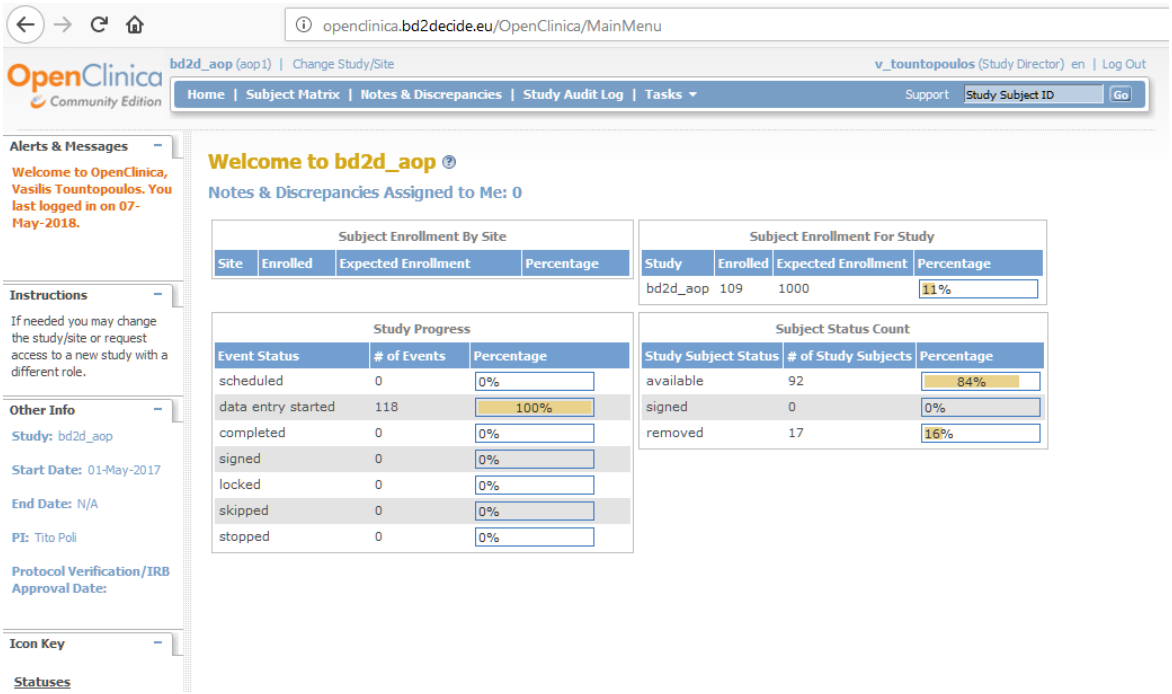

Figure 27: OpenClinica home page for a clinical user.

Manage Patient Records

On the Subject Matrix area of the study, the user can enter a new patient by clicking "Add New Subject" (green box in Figure 28).

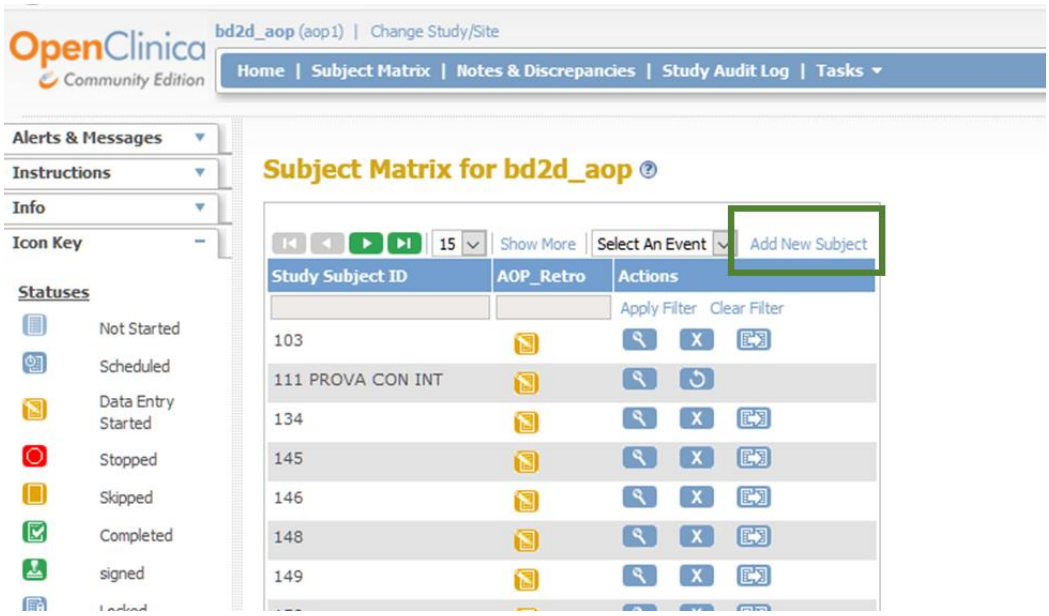

Figure 28: The OpenClinica Subject matrix providing the list of patients inserted in the study, assigned to the user.

Then, the following window appears, in which the user is prompted to provide a 'Study Subject ID' (this is not the BD2Decide patient ID) and the year of birth, select the sex and the study from a list, and confirm the patient enrollment in the study and the start date of inserting data (automatically set to the current date).

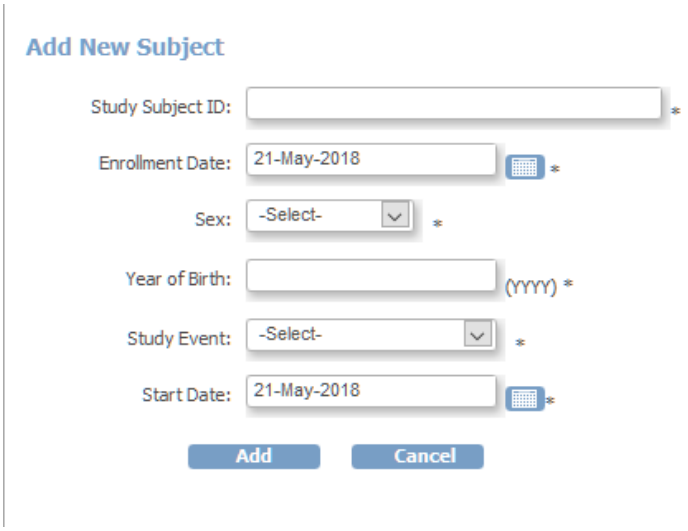

Figure 29: Adding new patient in OpenClinica

Then, the user should click the "Add" button. The patient overview table appears (see Figure 30). In this screen, the "Actions" column in the middle refer to actions (view, edit, remove) for managing the label of the patient, while the right end buttons refer to actions with respect to patient data (continue entering data, view, print, remove, etc.).

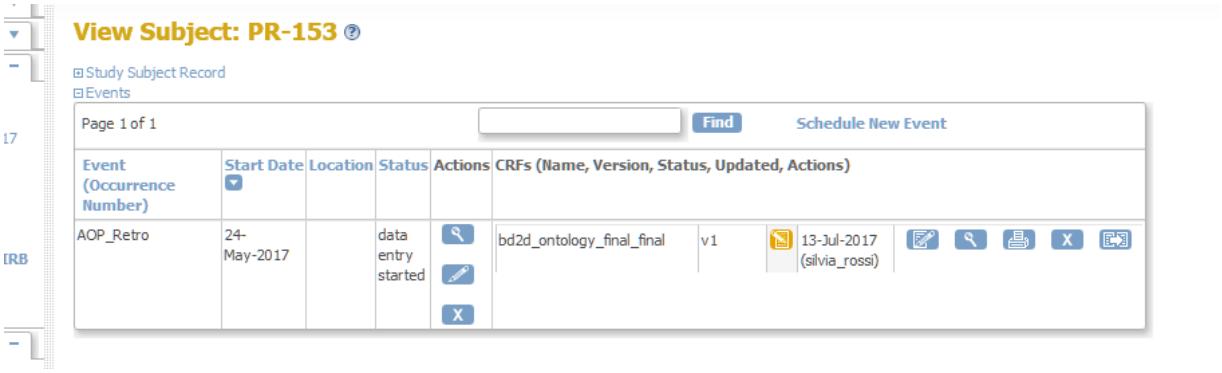

Figure 30: Patient overview table in OpenClinica

By clicking the "view" button in the "Actions" column, the user gets the following screen (Figure 31). In this screen, the "study subject ID" (orange box) is the ID the user provided when entering the patient, while the "study subject OID" (green box) is a unique identifier, which is automatically created in OpenClinica and corresponds to the **BD2Decide patient ID** that should be used for all BD2Decide analysis.

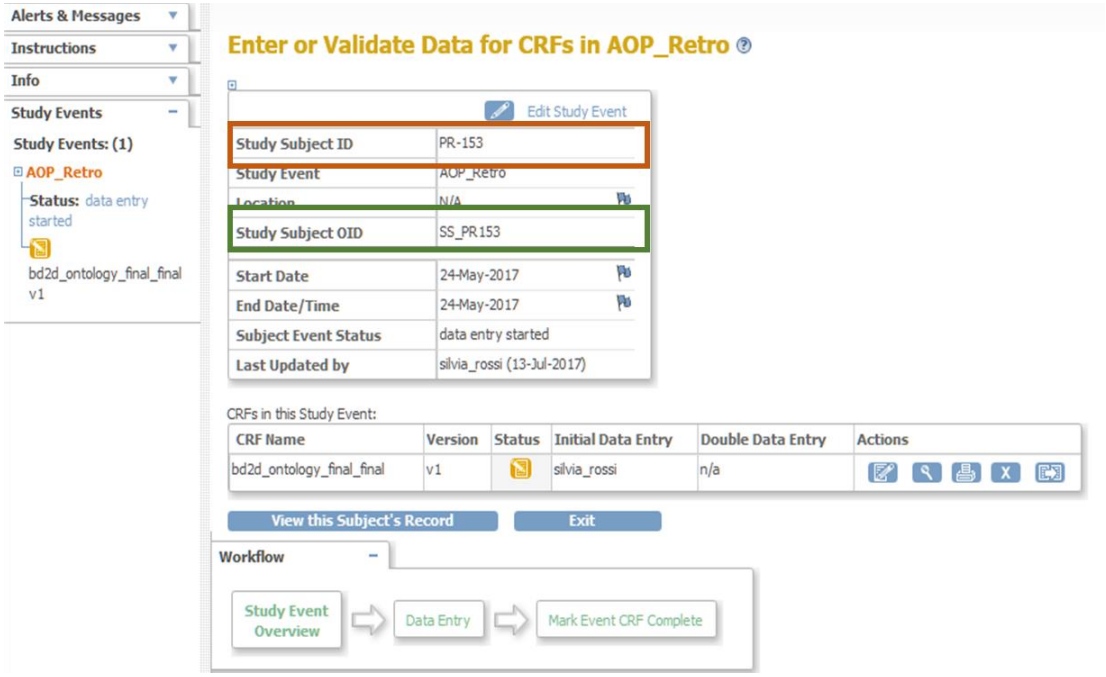

Figure 31: Patient 'view' option

From the Subject Matrix page and the patient overview table (see Figure 30), the user can access the e-CRF patient data space by clicking the "Continue Entering Data" button. This presents the entry form of Figure 32, which is split into the different sections of the e-CRF. The navigation to the different sections can be done either with the left and right arrows or directly by selecting the relevant section from the dropdown list. Above this view, the user can see the username of the patient (which is the study subject ID, named 'Username') and the BD2Decide patient ID (which is the automatically created study subject ID, named 'Patient ID').

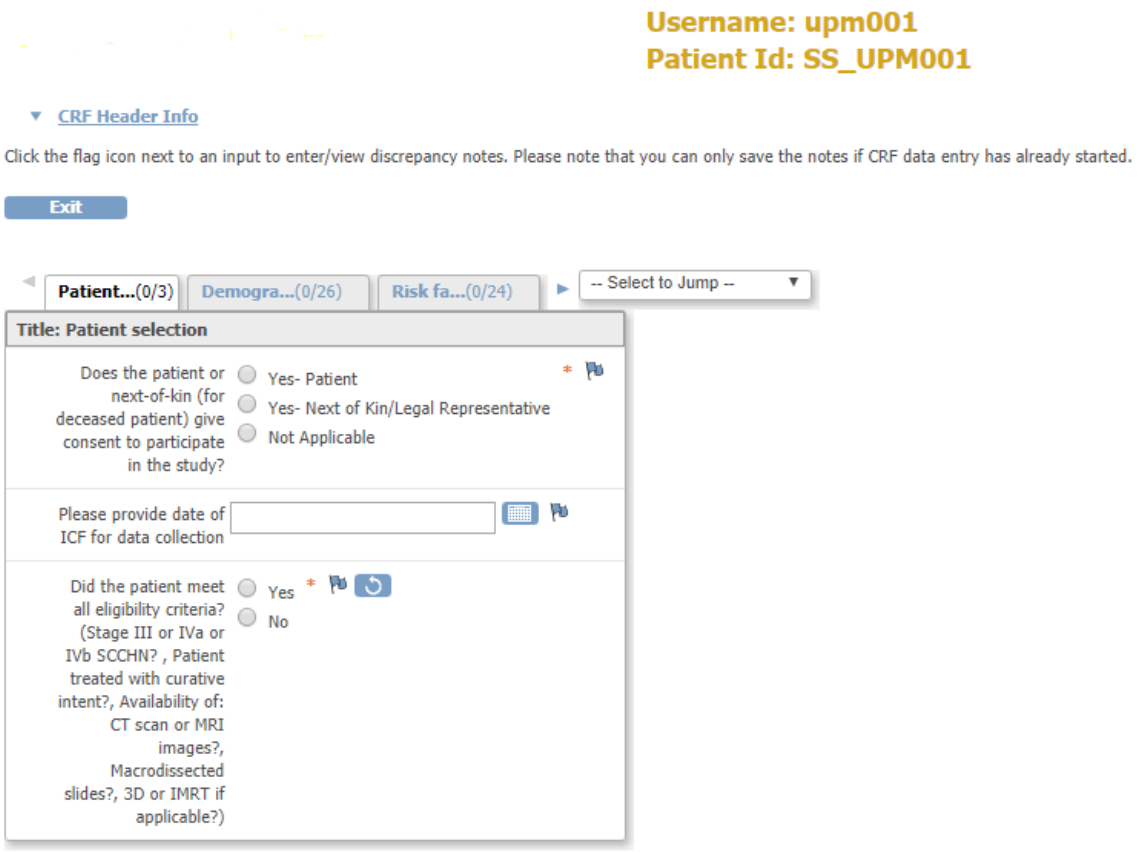

Figure 32: Patient data entry.

In each section:

- The mandatory fields are highlighted with a red asterisk (*)
- In order to place the value for one field, the user has to make a single or multiple selection from predefined answers (in most cases), either from a dropdown or radio button list.
- The values for some fields are automatically calculated, based on the responses given in previous fields in the respective section.
- We offer guidance on the proper format of the field values, by providing notes or links to external pages.
- In a couple of cases, a visual explanation of the candidate responses is provided, so that the user provides a self-assessment of the best possible value. This is for example implemented in oral hygiene field of the risk factors section of the eCRF, as shown in Figure 33.

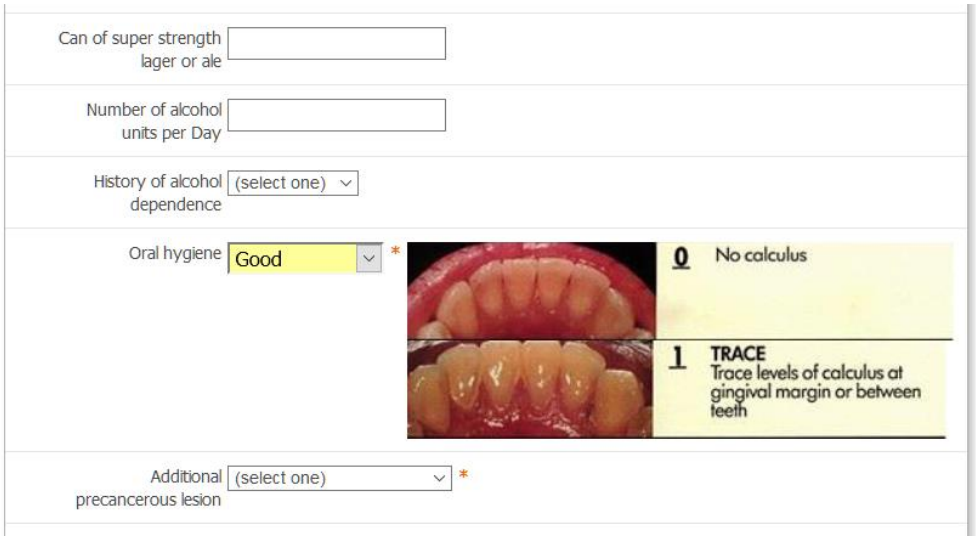

Figure 33: Example of help support in OpenClinica

Switching between studies

In top left corner, one can see the name of the study and the option to "Change Study/Site" (green box on Figure 34). The user should click on this option and, then, choose the preferred study and press the button "Change Study" (see Figure 35). The user is prompted to confirm this action (see Figure 35)

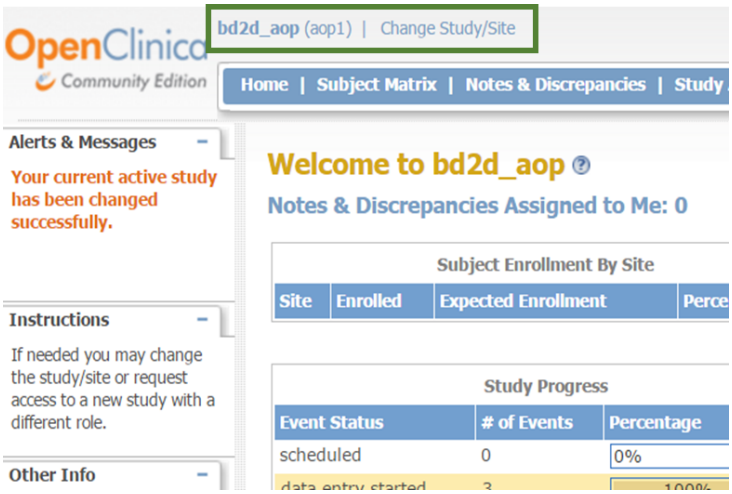

Figure 34: Switching between studies

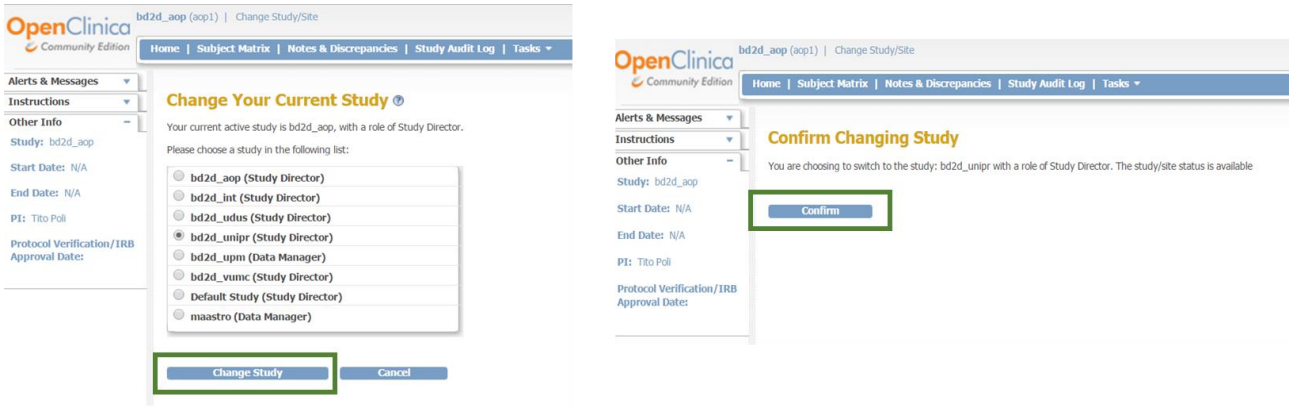

Figure 35. Changing study

Now, the user has moved to the home page of the selected study (see Figure 36).

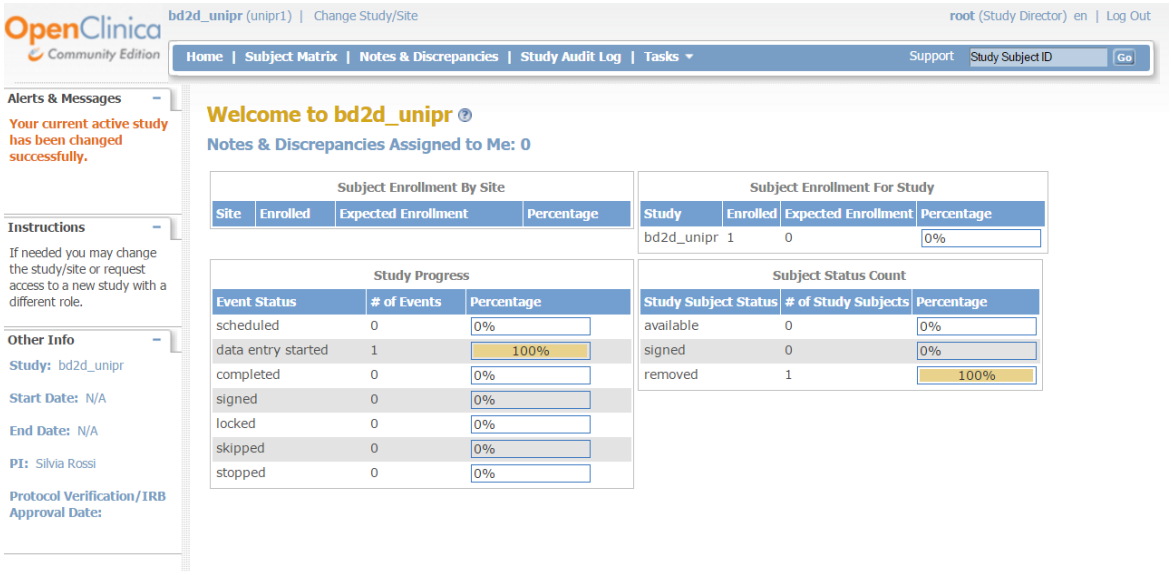

Figure 36: Home page of the newly selected study.

4.4.2 Adding QoL data into the BD2Decide environment

Within the BD2Decide project, the health professionals in the HNC domain use three instruments to assess the quality of life of their patients. As it was introduced in Deliverable D2.2 and explained further in D5.2, these instruments are namely the following questionnaires promoted by the international standards:

- The EQ-5D-5L questionnaire developed by the EuroQol Group[16] for the measurement of health outcome;
- The QLQ-C30 from EORTC[17], which is a questionnaire developed to assess the quality of life of cancer patients;

[16] http://www.euroqol.org/eq-5d-products.html

[17] http://groups.eortc.be/qol/

- The QLQ - H&N35 from EORTC, which is a questionnaire developed to assess the symptoms or problems arisen from a treatment followed by cancer patients.

As a complementary part of the CDSS environment, the BD2Decide project has followed the guidelines of the respective bodies and has implemented an online version of these questionnaires to facilitate the collection of data from the patients. The development of the questionnaires has considered the use of personal mobile devices (and especially tablet devices of 10 inches' screen, as reported in the guidelines) through Web browsers. The technology behind this implementation is the open source software LimeSurvey[18], which has been extended with custom implementation for the development of specific parts of these questionnaires, such as the visual analogue scale used in the EQ-5D-5L product (see Figure 37).

Figure 37: Quality of Life assessment – screenshot from the implementation of the visual analogue scale in the EQ-5D-5L product, with Limesurvey.

It must be noted that, since the three questionnaire instruments refer to the patients of the clinical centres in three different countries, they have been developed in four different languages (based on the official translation provided by the responsible bodies, where this was available).

4.4.3 Interface with the Interactive Patient co-Decision Aid tool

In order to access the Interactive Patient co-Decision Aid (IPDA) tool from CDSS, one can click on "My Patients" menu item, select a patient to open the Patient Overview page and, then, click on the IPDA link on the top right, as shown in Figure 38.

[18] https://www.limesurvey.org/

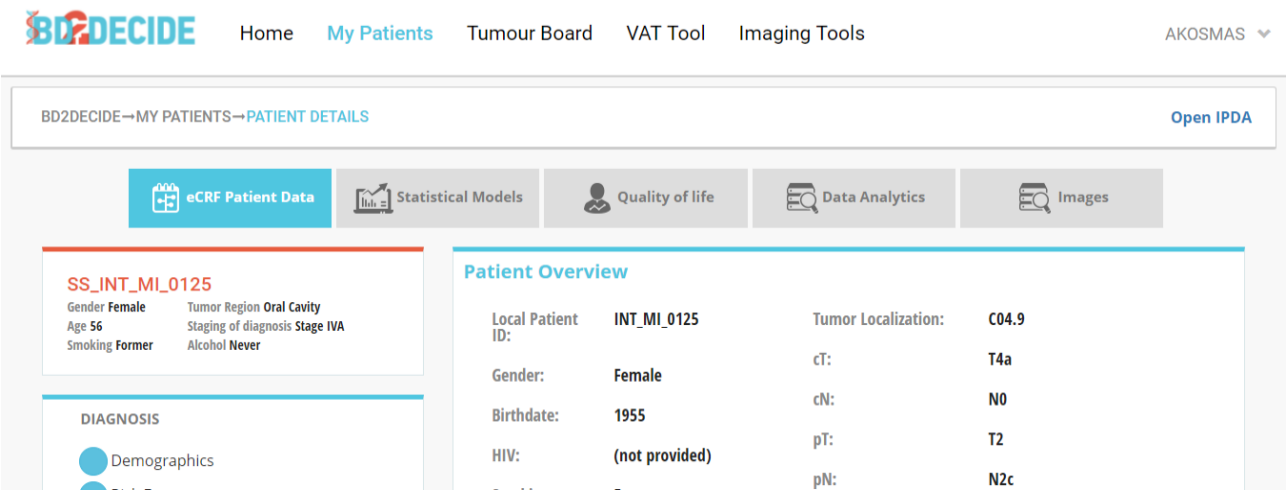

Figure 38: Interaction of CDSS with the IPDA tool.

The IPDA tool will open in a new tab with the required information without the need for further action from the user. This information relates to the BD2Decide patient ID, the year of birth and the tumor stage. The use of IPDA has been reported in Deliverable D5.4 [3].

4.4.4 Connection to the Big Data Infrastructure

All the information that is collected and generated for a specific patient, including the e-CRF-based clinical data, the genomics and the radiomic features are stored and processed in the BDI. In order for the PDS to upload the clinical data from OpenClinica to the BDI, a specific java web service has been developed. This service is described on the following Table 1.

Table 1: The specifications of the service for loading clinical data to the BDI.

Service Endpoint	
[GET] /api/openclinicaServicesAPI/openclinica/getopenclinicadata?study={study_id}	
Description of the Service	
This web service allows the retrieval of all the clinician data for a given study	
Input	
Parameter and Value	Brief Description
study_id	The id of the study as it is provided by OpenClinica e.g. S_BD2D_INT
Outputs	
A list of the study data in JSON format	
Sample Response	

[

{

"personID":null,

"studySubjectLabel":"INT_MI_0226",

"studySubjectKey":{"name":"http://ontology.lst.tfo.upm.es/BD2D/clinical#Patient_ID","value":"SS_INT_MI_0226"},

"studyEventData":{"startDate":"14-Nov-2017","studyOID":"SE_INT_PROSP"},

"gender":"m",

"itemData":[

{"name":"http://ontology.lst.tfo.upm.es/BD2D/risk#Additional_precancerous_lesion_2","value":"http://ontology.lst.tfo.upm.es/BD2D/static/add_prec_lesion2@@1"},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/risk#Additional_precancerous_lesion_1","value":"http://ontology.lst.tfo.upm.es/BD2D/static/Additional_precancerous_lesion@@1"},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/clinical#Age_at_Diagnosis","value":"62"},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/risk#Alcohol_at_Time_of_Diagnosis","value":"http://ontology.lst.tfo.upm.es/BD2D/static/Alcohol@@3"},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/risk#Alcopop","value":""},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/ctn#Anatomical_Tumor_Location","value":"http://ontology.lst.tfo.upm.es/BD2D/static/Anatomical_Tumor_Location@@45"},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/clinical#ASA","value":"http://ontology.lst.tfo.upm.es/BD2D/static/ASA@@6"},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/tox#Aspiration_Toxicity_12","value":""},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/tox#Aspiration_Toxicity_24","value":""},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/tox#Aspiration_Toxicity_6","value":"http://ontology.lst.tfo.upm.es/BD2D/static/aspiration_toxicity_6@@1"},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/tox#Asthenia_Toxicity_12","value":""},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/tox#Asthenia_Toxicity_24","value":""},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/tox#Asthenia_Toxicity_6","value":"http://ontology.lst.tfo.upm.es/BD2D/static/asthenia_toxicity_6@@1"},

{"name":"http://ontology.lst.tfo.upm.es/BD2D/patho#Basaloid_features","value":""},

.....

]

}

]

5 FUNCTIONAL VISUALISATION MOCKUPS OF THE VISUAL ANALYTICS TOOL

The VAT is a web application that aims to support the clinicians during research activities. This tool integrates data coming from heterogeneous data sources and also the analytics techniques needed for data exploitation and visualization. All the functionalities and interaction workflow are detailed in this chapter.

5.1 Overview of the scenarios of use

Following the user centred design methodology described in the book "About Face" by Cooper [7], we have created a set of goal-directed scenarios in the design phase of the VAT. These scenarios allows understand the real usage of the tool in a daily research context. Details will be provided in the BD2Decide Deliverable D5.7 "Usability evaluation results", which is due M36 (December 2018).

Taking into account the work that clinicians do, like searching, exploring and hypothesis testing, the VAT allows to implement two main scenarios.

5.1.1 Exploratory search

This scenario allow clinicians to first analyse the data before launching any exploratory specific project. Through this scenario, users can interrogate into the BD2Decide dataset in different ways:

- Having a first overview of the BD2Decide data, consisting on a descriptive analysis of the *top 15 Requests* (see Section 1.2). The users can interact with this module switching among the different data types and selecting the type of chart they prefer to use.
- Executing on-demand queries to all the BD2Decide dataset. This functionally offers the possibility to query the overall BDI. Users can select the data they want to analyse, selecting inputs, conditional variables and outcome of interest.
- Discovering prognostic factors through the adoption of machine learning techniques applied to a selected cohort of interest. Both supervised and unsupervised techniques will provide respectively survival and classification outcomes, clustering and factor analysis.

5.1.2 Exploratory search projects

Once the users have explored the data and have formulated some hypothesis, they can execute a search project, to carry out an analysis oriented to publications or outcome generation. Working in search projects, allows the users to work in a collaborative way.

During the creation of a project, the principal investigator (i.e. the user that creates the project) should specify a *project name, project type, project description* and *the research problems* that he/she wants to address. Once created, the following functionalities are available:

- Based on the type of the project, the VAT will facilitate the users some specific tasks to execute, guiding them to carry out their research activities.

- Users can select the *Inclusion criteria* to work with, pointing out which are the outcomes expected for the selected data. After this variable selection, patient cohort is represented grouped by the overall survival.
- The features available in the previous scenario are also accessible within the project.
- This scenario also allows run statistical models to the selected cohort, namely: Randon survival forest, Cox proportional hazards regression and Survival prediction.
- Similarly, the project also allows to select among different machine learning techniques, such as Supper Vector Machine Survival, Logistics, K-means or Principal Component Analysis. The possibility to select one technique or another will be based on the project outcomes and exploratory search questions.
- The VAT also allows to perform a literature analysis, based on the description of the project provided by the user.
- The user can access to similar cases from external studies, supporting the identification of relevant clinical factors, risk factors that may be underestimated and not described by patients during visits. The integration of the external population data enables VAT to execute the following scenarios: i) a quality of care evaluation in a set of European countries, according to defined quality criteria; ii) analysis of time trends of the incidence, survival and mortality rates for different European countries and HNC sites, combined also with the BD2Decide data; and iii) analysis on risk factors changes over time, examining the exposure to risk factors in different countries or correlating risk factors and context data with HNC incidence rates.

As an extra functionality, the VAT integrates a tool to extract radiomic features from both MRI and CT images. The results of this implementation are being analysed and compared with the results obtained in the radiomic feature extractor tools that were developed in BD2Decide, to identify the added value that this feature provides in the VAT.

Once the project is finished, the users will get a **Project summary** with the report of the overall research carried out.

On top of this, a knowledge generation workflow will be integrated into the tool, extracting the validated results and new evidence in the field, this functionality will be reported in the BD2Decide Deliverable D6.4, which is due M36 (December 2018).

Figure 39 represents the functionalities above described related with the services mentioned in the architecture (Figure 2). In light blue there are the imaging and genomic models that feed the Statistical models and Big Data Analytics. These components (in dark blue) are integrated in the major part of the research functionalities implemented in the Overview, Variable analysis, Population statistics, Data analysis and Literature search scenarios presented in the next sections.

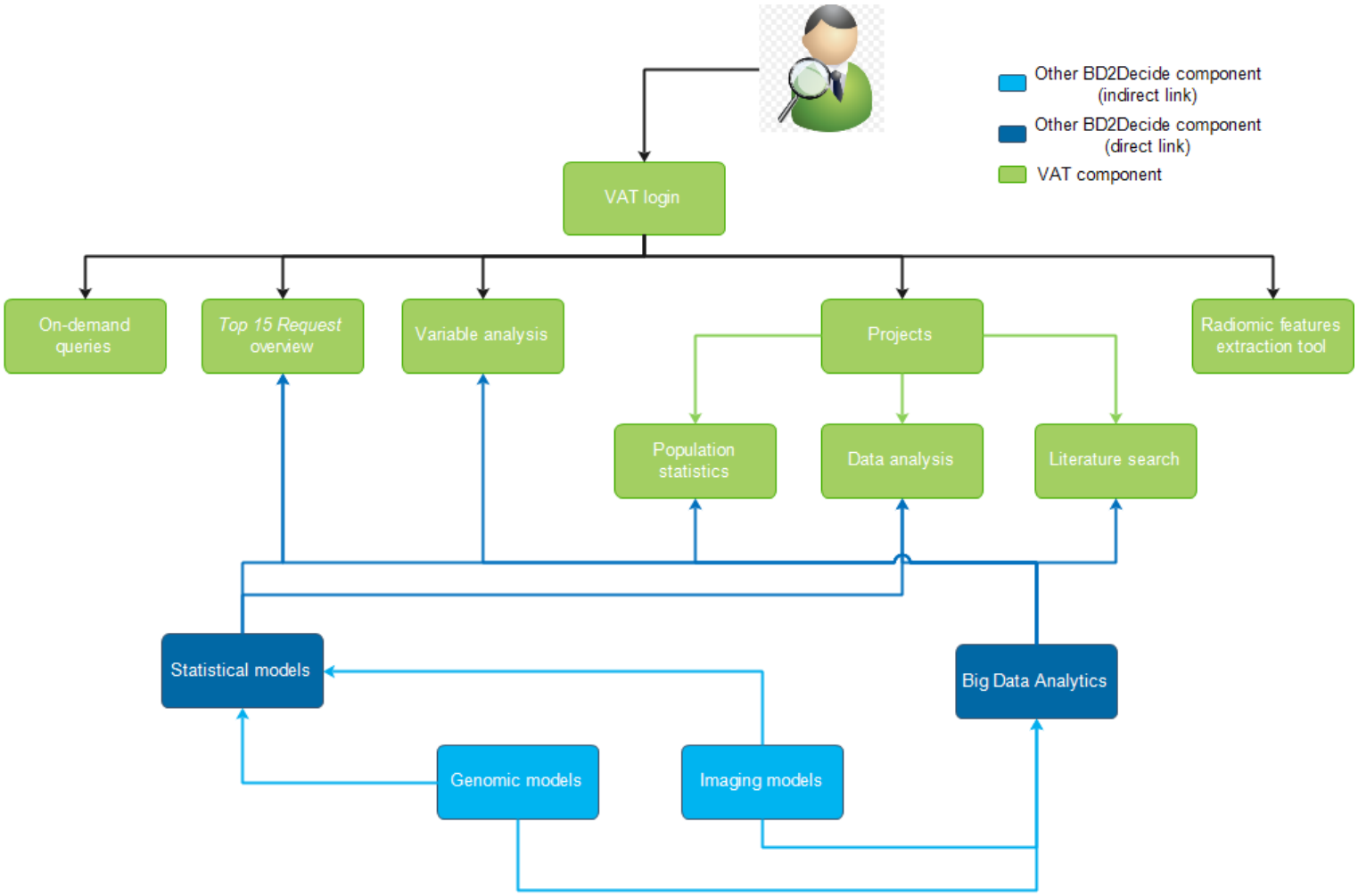

Figure 39: VAT relation diagram.

5.2 Visualising the scenarios of use

5.2.1 Exploratory search

Once the users log into the VAT, they are provided with an overview of the system (Figure 40, Figure 41, Figure 42, Figure 43, Figure 44, Figure 45 and Figure 46) that is a descriptive analysis of the BD2Decide dataset based on the *top 15 requests* defined by clinicians.

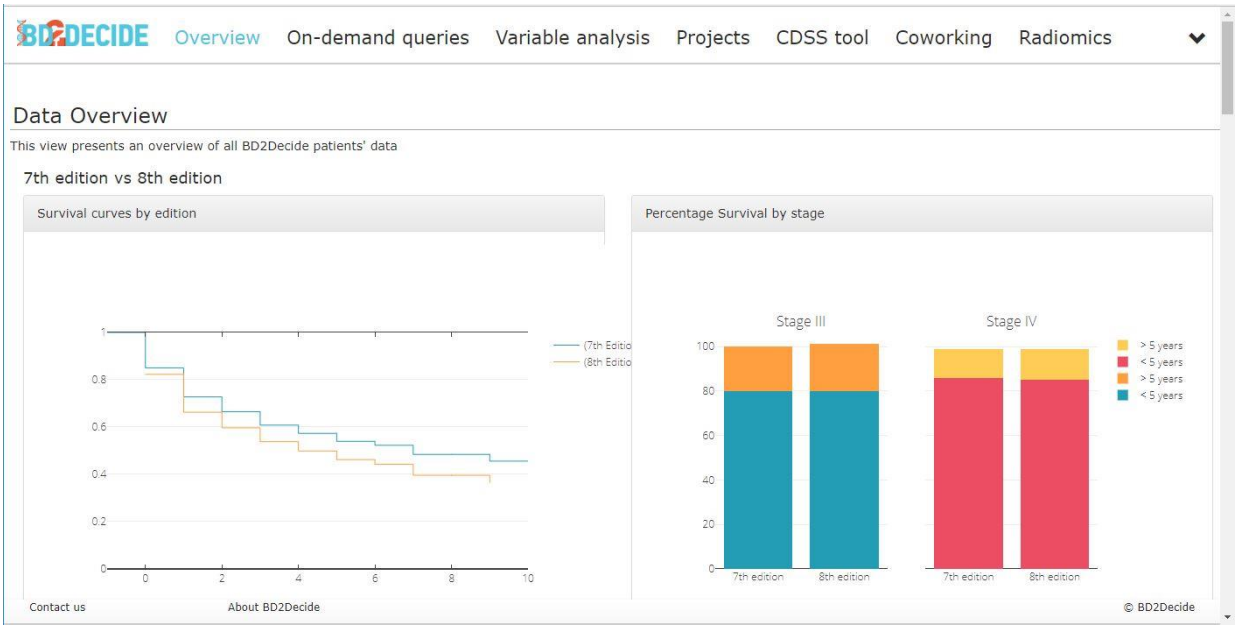

Figure 40: Stage comparison for the entire BD2Decide cohort presented in VAT overview.

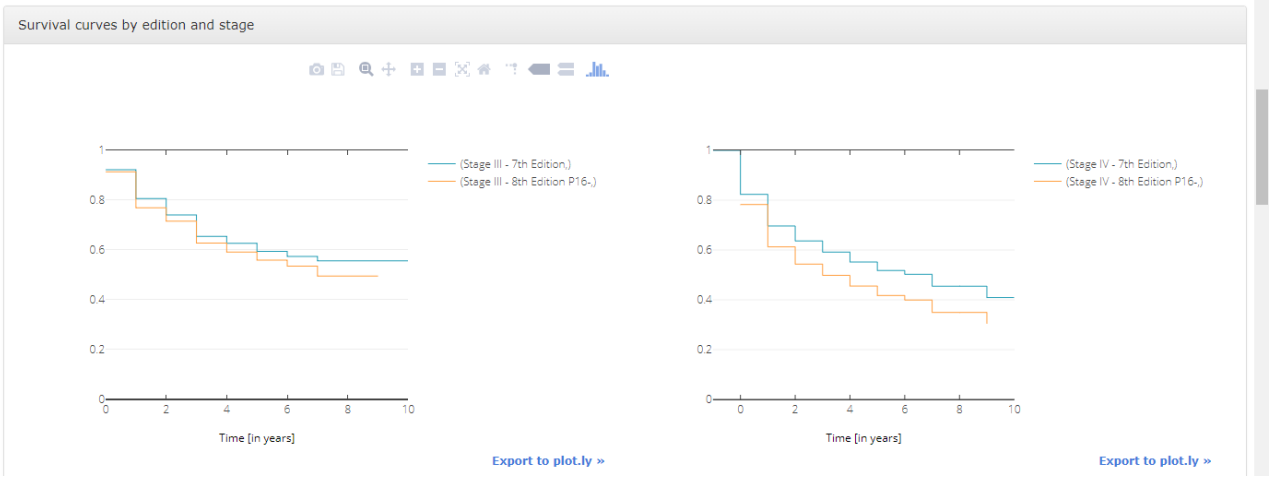

Figure 41: Survival curves by stage for the entire BD2Decide cohort, presented in VAT overview.

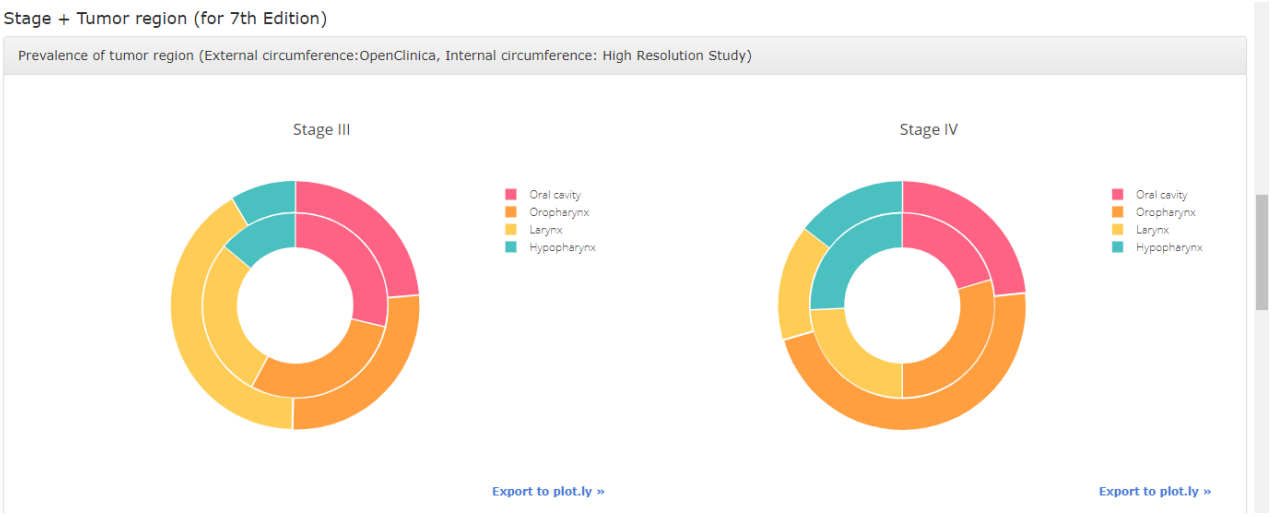

Figure 42: Doughnut chart with the comparison by stage and tumour region for the BD2Decide patients and the cancer registries cases, presented in VAT overview.

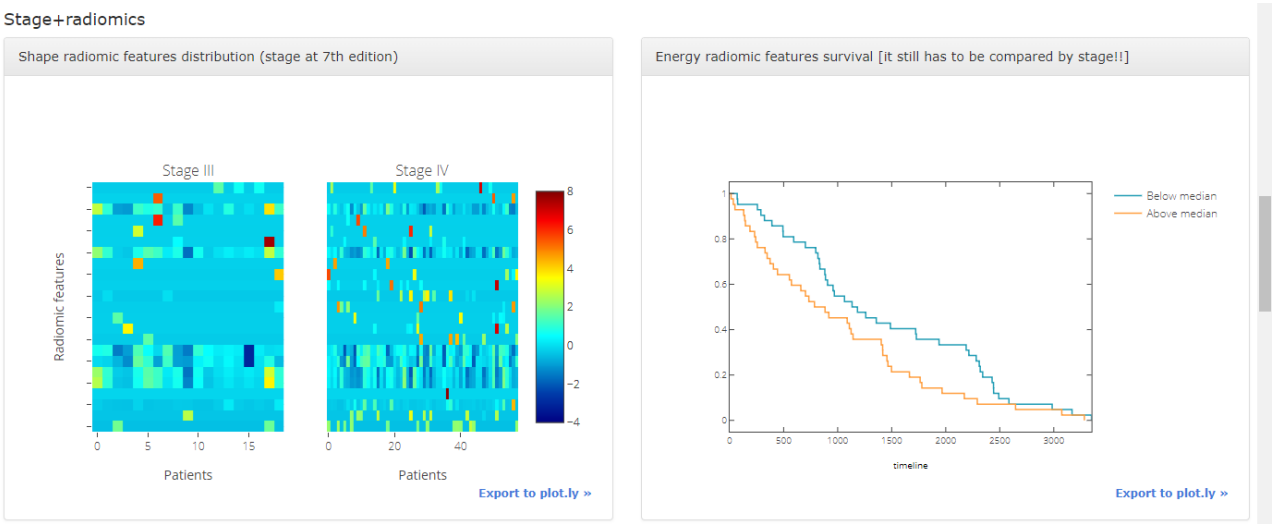

Figure 43: Radiomics distribution (left) and survival (right) representation in VAT overview.

This section is divided in each of the clinician request, and includes different types of representation for the same request in order to provide more information on the specified *request*. For instance, in Figure 43, the combination of radiomics and stage is represented in a heatmap to see the distribution of the features, and in a Kaplan-Meier curve, to see the survival curves of the patients that have these features values below or above the median.

Also the 'Stage + Other clinical factors requests' has the peculiarity to select among different clinical factors (up to ten, as requested by clinicians). Additionally, users can switch among different types of chart, as shown in Figure 44, Figure 45 and Figure 46.

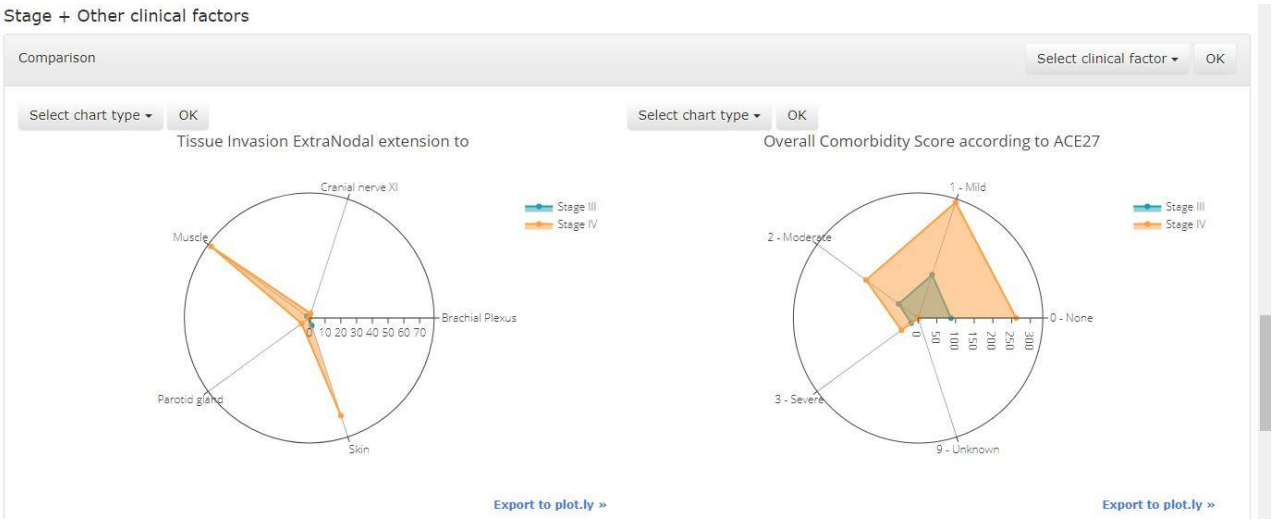

Figure 44: Tissue invasion extra nodal extension raking (left) and comparison among different comorbidity score (right) VAT overview. Radar chart representation, it is possible to change the type of chart selecting on 'Select chart type' button. It is also possible to switch among other clinical factors via selecting the 'Select clinical factor' button.

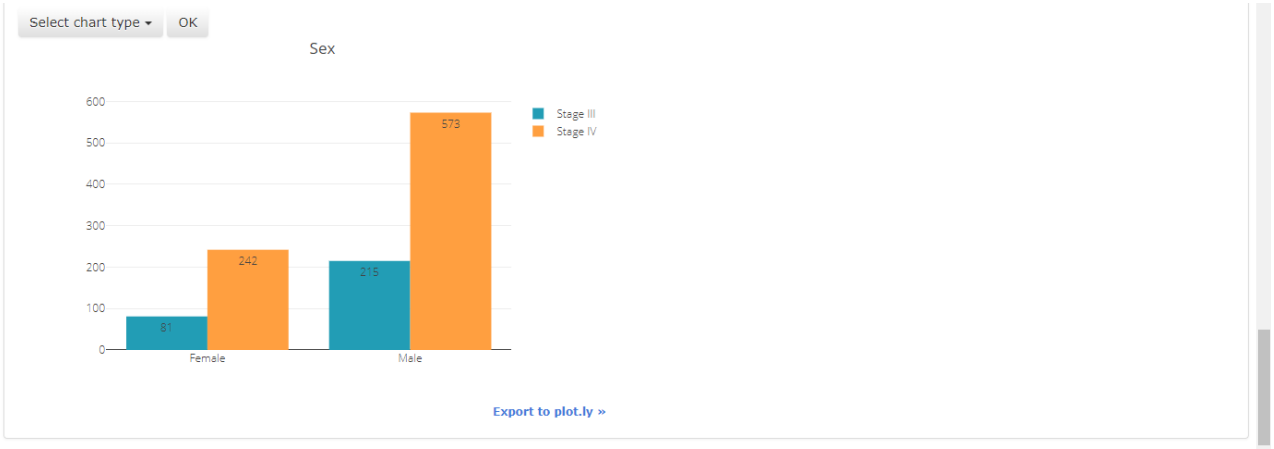

Figure 45: Sex comparison among stage III and IV, presented in VAT overview.

The user can then select the ***'On-demand queries'*** tab, next to the 'Overview', to query the BDI with customizable requests from clinicians without the need for any knowledge on SQL language. It is an ongoing activity, and the current mock-ups are presented in Figure 47, Figure 48 and Figure 49. This functionality will allow the clinicians to specify 'what they are looking for' and 'with which conditions (filters)'. In the summary table, they will be able to specify which are the fields of interest and customize the results based on this. As a result, the system will provide the outcome, which can be a table or a number along with some descriptive charts (Figure 48 and Figure 49).

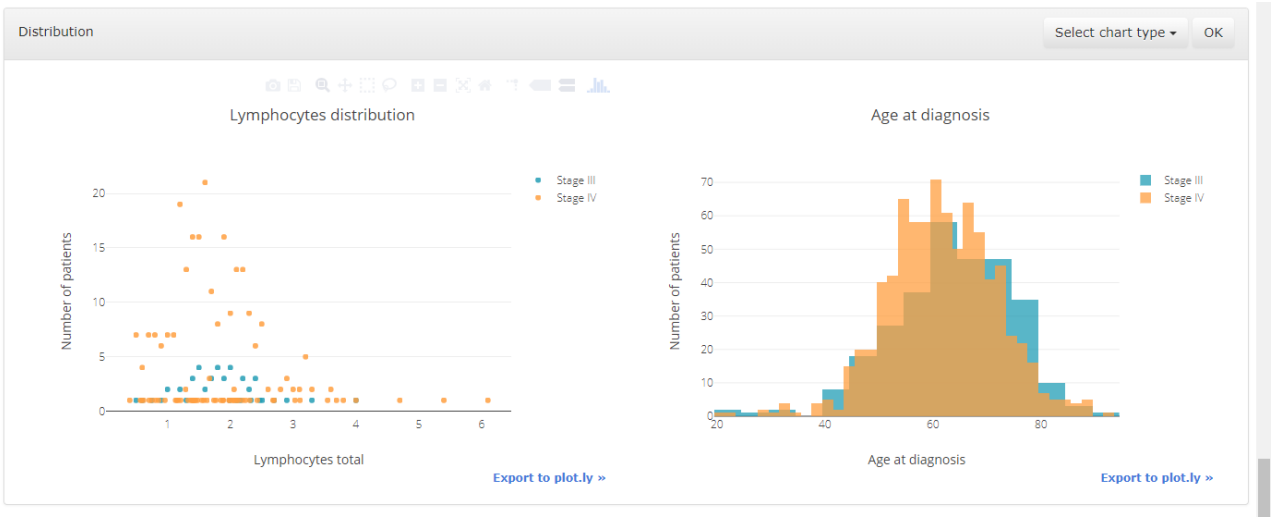

Figure 46: Lymphocytes distribution by Stage (left) and age at diagnosis distribution for all BD2Decide cohort, divided by Stage. Presentation in VAT overview.

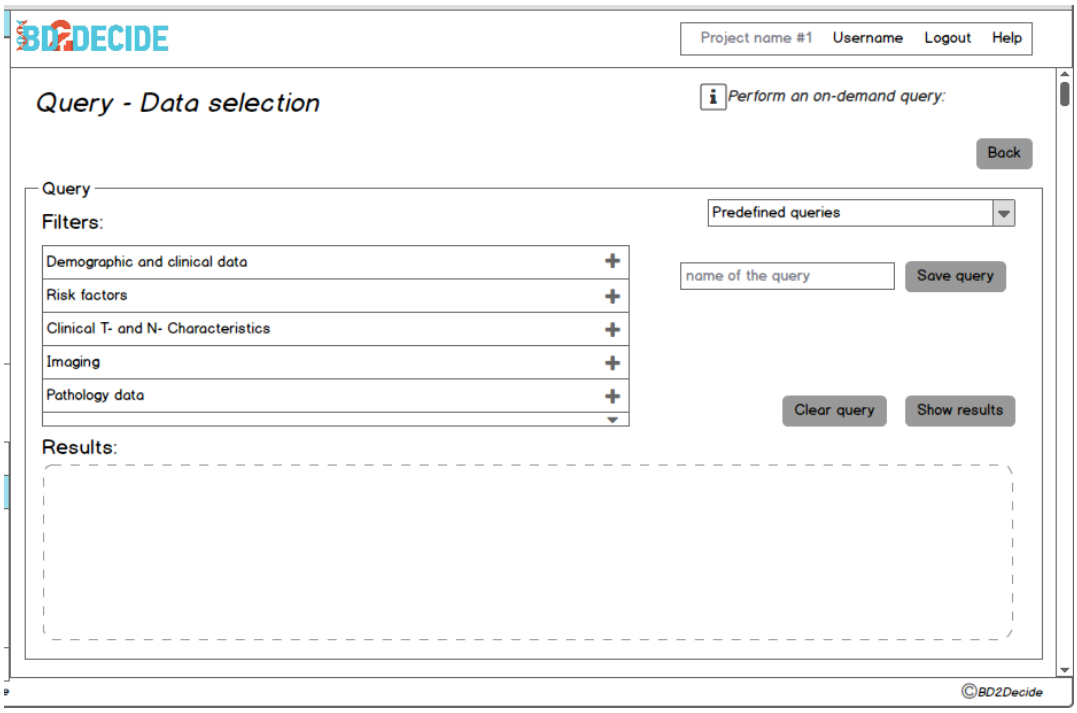

Figure 47: Data selection in the 'On-demand queries' to choose the filters to apply to the overall BD2Decide dataset. Filters are first categorized by e-CRF sections.

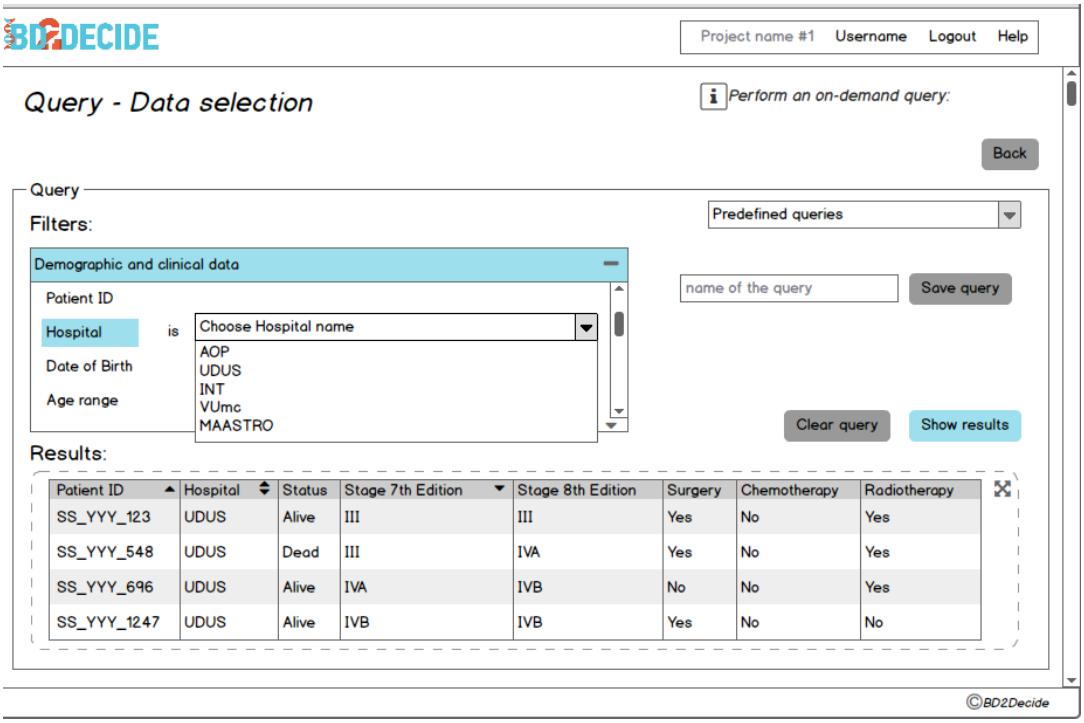

Figure 48: On-demand queries table result. After selecting the e-CRF section (in this example 'Demographic and clinical data'), the variable to filter is selected (in this example the filter is applied by 'Hospital').

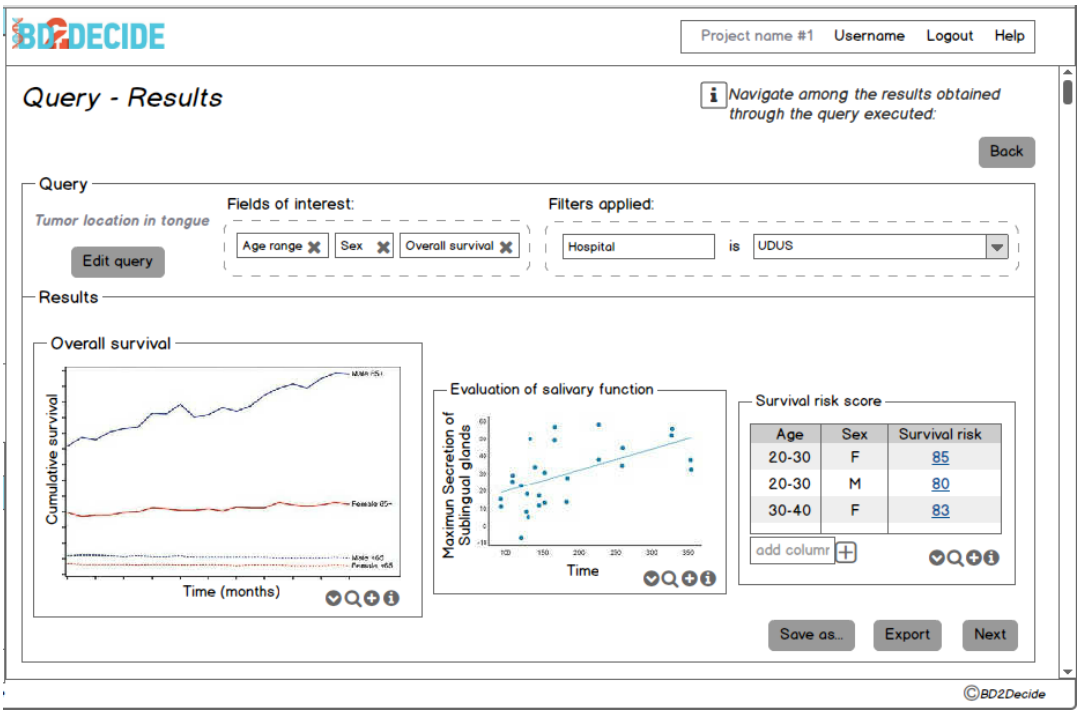

Figure 49: Results in the 'On-demand queries' accompanied with graphs as survival curves or summary tables.

If the users wants to execute more specific data analytics, to discover data correlations, the tab 'Variable analysis' is the one that fulfil these conditions. This functionality will integrate machine-learning techniques to allow clinicians to:

- study the prediction factors of the BD2Decide dataset,
- evaluate the clinical variables relations and
- understand the meaning and relevance of the patient variables.

Figure 50 shows the univariate analysis already implemented by All-In-Image, where it is possible to evaluate, in terms of survival, the significance of the e-CRF variables.

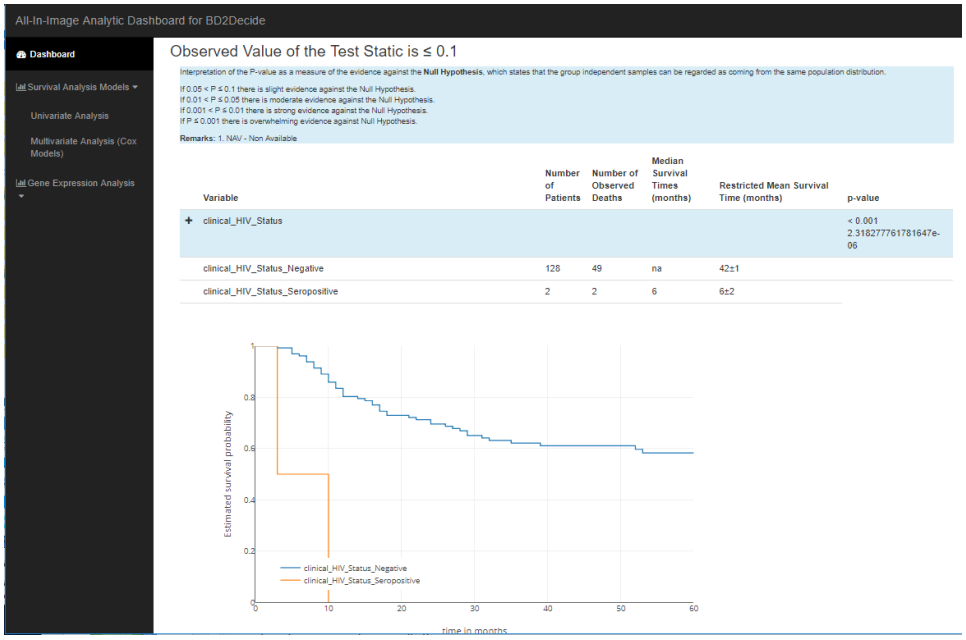

Figure 50: All-In-Image Univariate Analysis example to be integrated in the VAT.

5.2.2 Exploratory search projects

If the users, after analysing for instance the *Overview* and performing a *Variable analysis*, have a specific research question to address, they can open an existing project or create a new one. The project should contain at the beginning (see Figure 51 and Figure 52):

- Project name;
- Project type (Population based study, Case control study or Systematic review among others);
- Project description;
- Research problems.

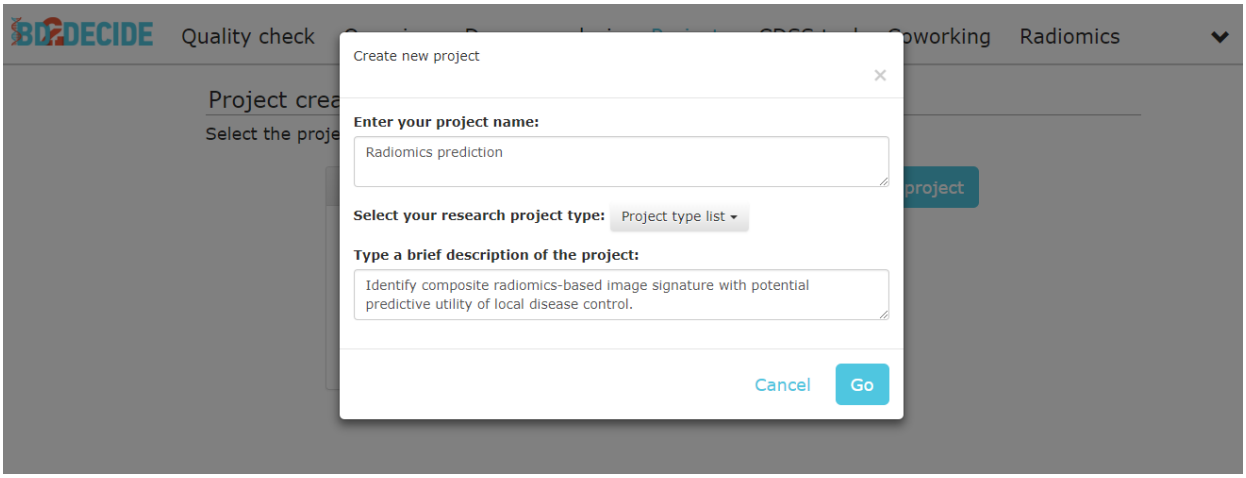

Figure 51: Project creation in VAT. Project name, description, type of project and project description insertion.

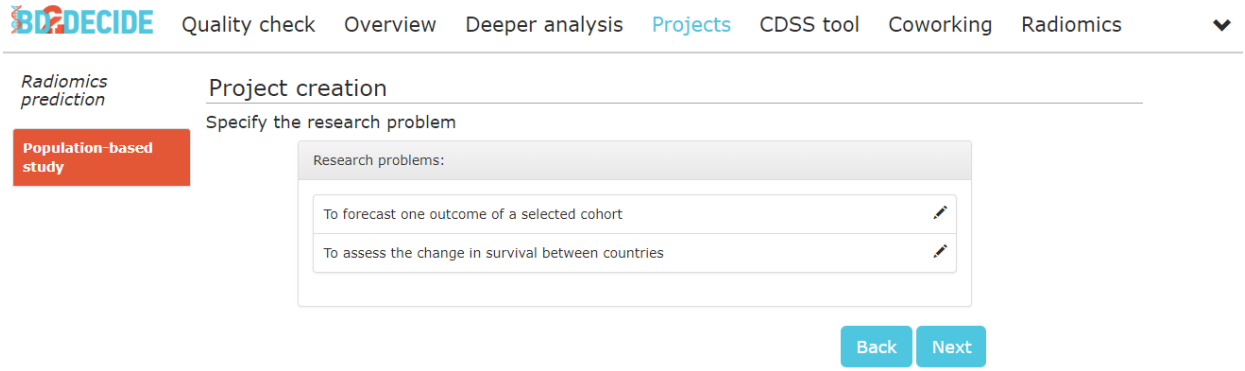

Figure 52: Specifying the research problem(s) in VAT.

With the above information, the tool suggests a research plan (see Figure 53), based on the project type.

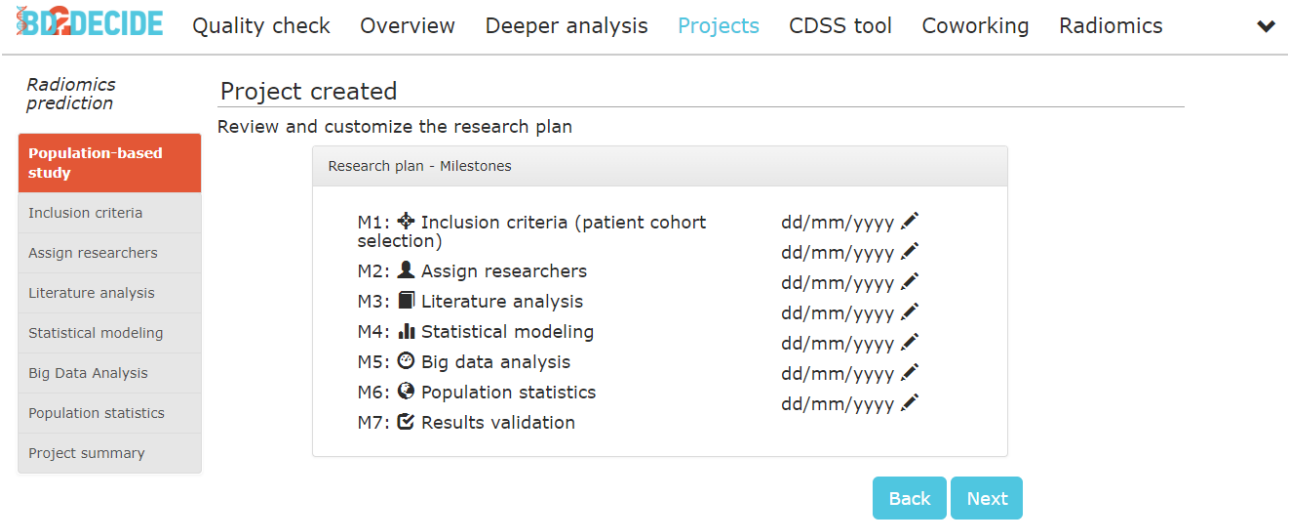

Figure 53: Project plan visualisation proposed through the VAT tool.

At the beginning, the users can describe the cohort, on which they want to execute the analysis (see Figure 54). Also, they can include the variable that they want to measure, depending on which other variable (exposure of interest), and to conclude by indicating the Outcome of interest. The selection of these variables allows the system to provide the adequate analysis and data representation.

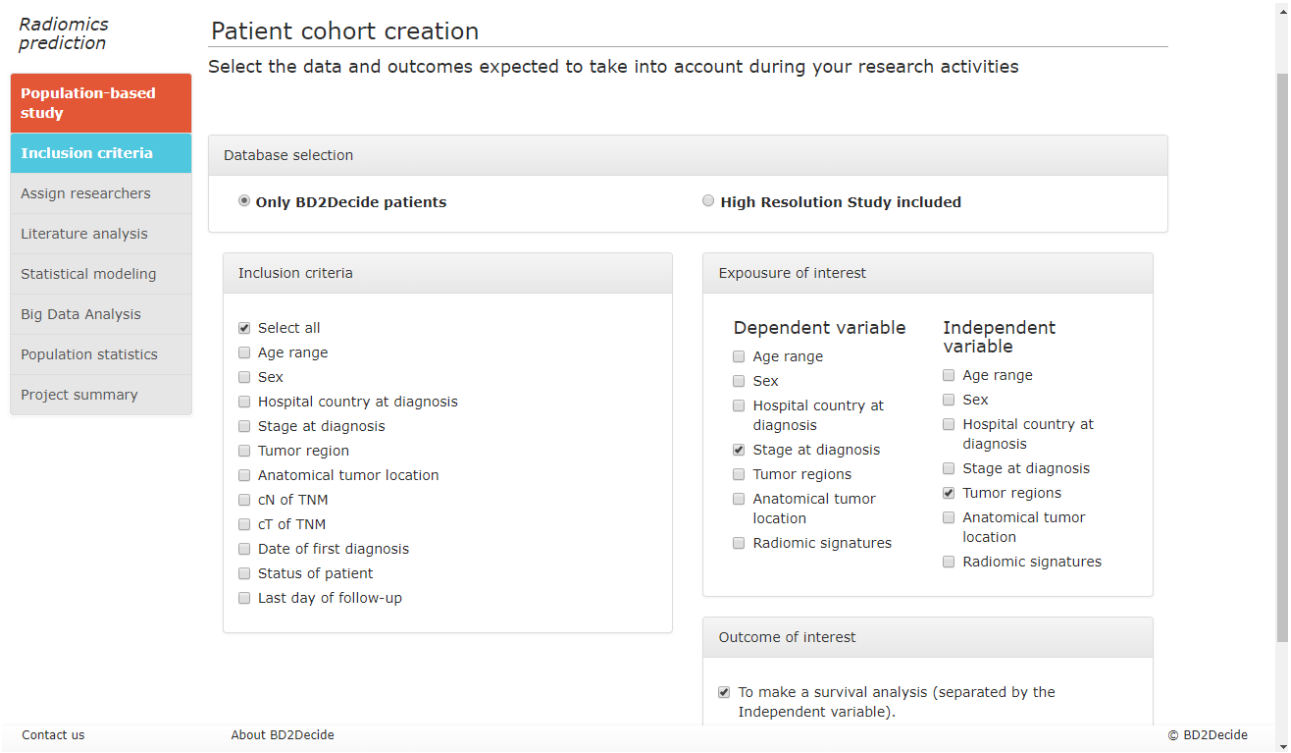

Figure 54: Inclusion criteria for the creation of patients' cohort in VAT.

Now that the project has been created and the cohort and outcomes have been identified, a brief summary of the cohort is provided (see Figure 55).

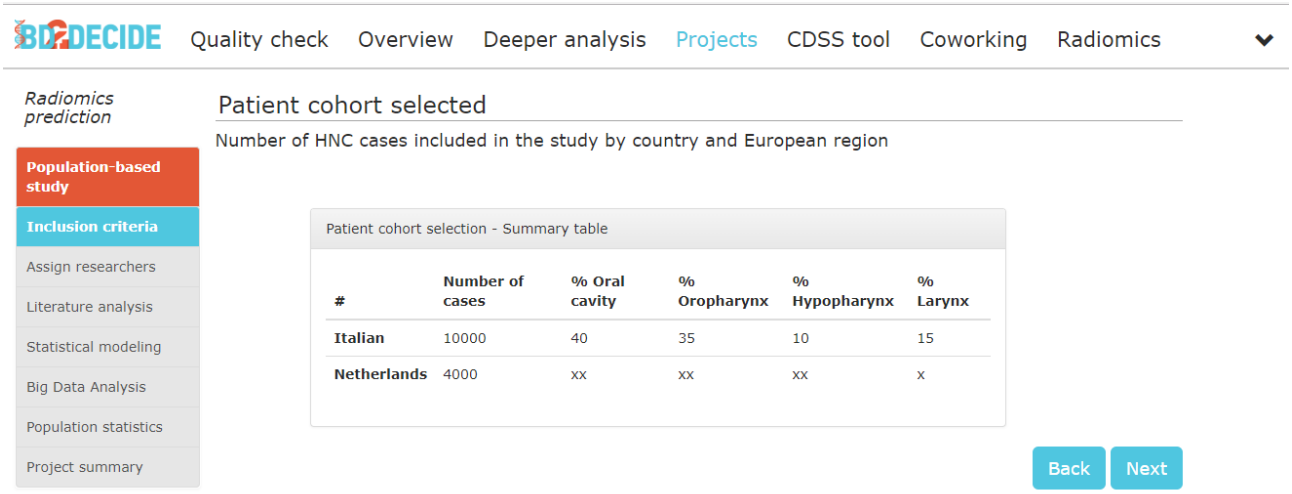

Figure 55: Summary of the cohort data, based on the inclusion criteria.

The user is able to assign more researchers to the project, to work in collaboration, as shown in Figure 56.

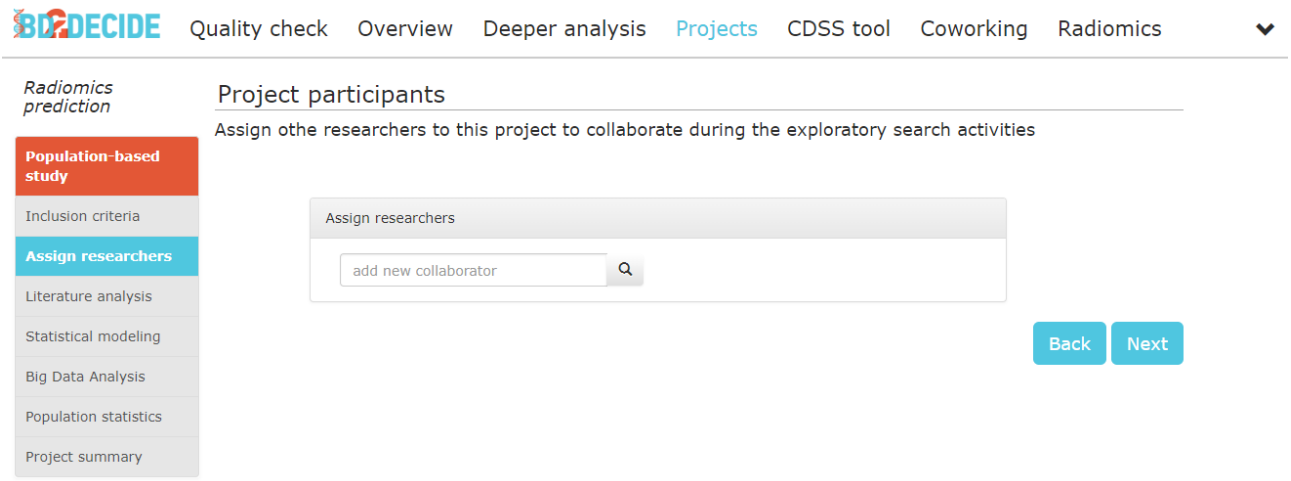

Figure 56: Assign researchers to the project through the VAT tool.

Then, the system guides the users through the different activities that have been scheduled in the project creation. For each one of them, the tool uses the input already provided by the users, which is:

1. literature analysis

The literature analysis extracts the relevant keywords of the research from the project description provided previously by the user. These keywords will be matched with the analysis of PubMed articles to suggest the users a list of recommended articles.

The VAT retrieves the HNC literature from PubMed[19] and analyses them with a Natural Language Processing Toolkit named NLTK[20], by extracting the relevant words from the articles abstract.

The list of relevant words is employed to recommend the appropriate articles to the users, based on the description of the research they want to carry out. To minimize the users work, this process is transparent to them.

VAT provides a list of the recommended articles specifying why they are so relevant (based on the match of the relevant words extracted), the year of publication and the number of authors among others fields (Figure 57).

If the users prefers to have a graphical overview of the articles recommended, by clicking on 'View descriptive charts' button in Figure 57, they will have a descriptive analysis of the suggested literature, as it is presented in Figure 58. This is an interactive representation of the literature list. Those charts work as filters through the selection of the specific section of the chart. For instance, by

[19] https://www.ncbi.nlm.nih.gov/pubmed/

[20] https://www.nltk.org/

clicking on the type of publication 'In process' (light blue in Figure 58), the system will provide the recommended lectures, filtered by the type of publication 'In process'.

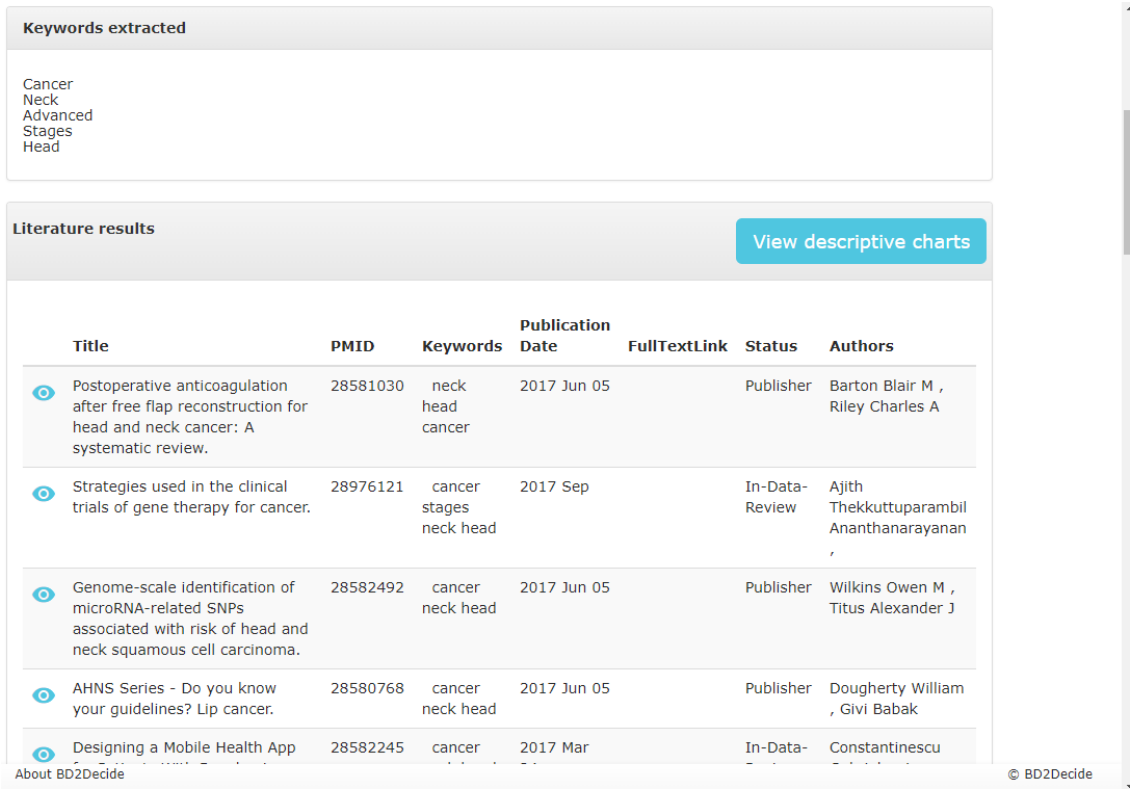

Figure 57: Literature analysis in VAT- articles suggested.

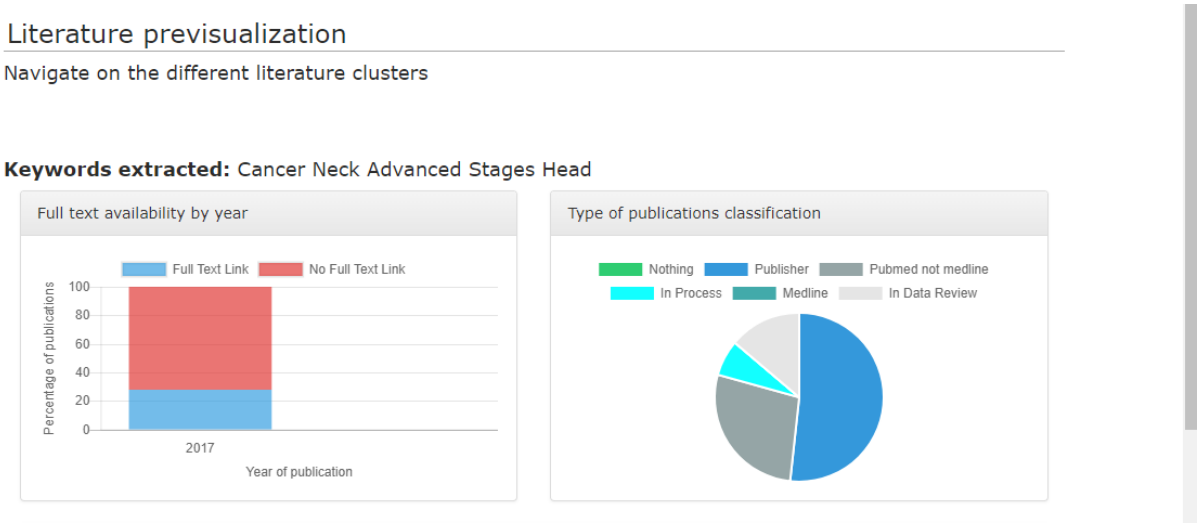

Figure 58: Literature previsualization in VAT.

The lecture of the article is also possible through VAT in two different ways:

- If the article has the full text available in PubMed database, it will be presented in VAT.
- If the full text is not available, then the external link to the article will be provided.

The abstract is always available through this module (Figure 59).

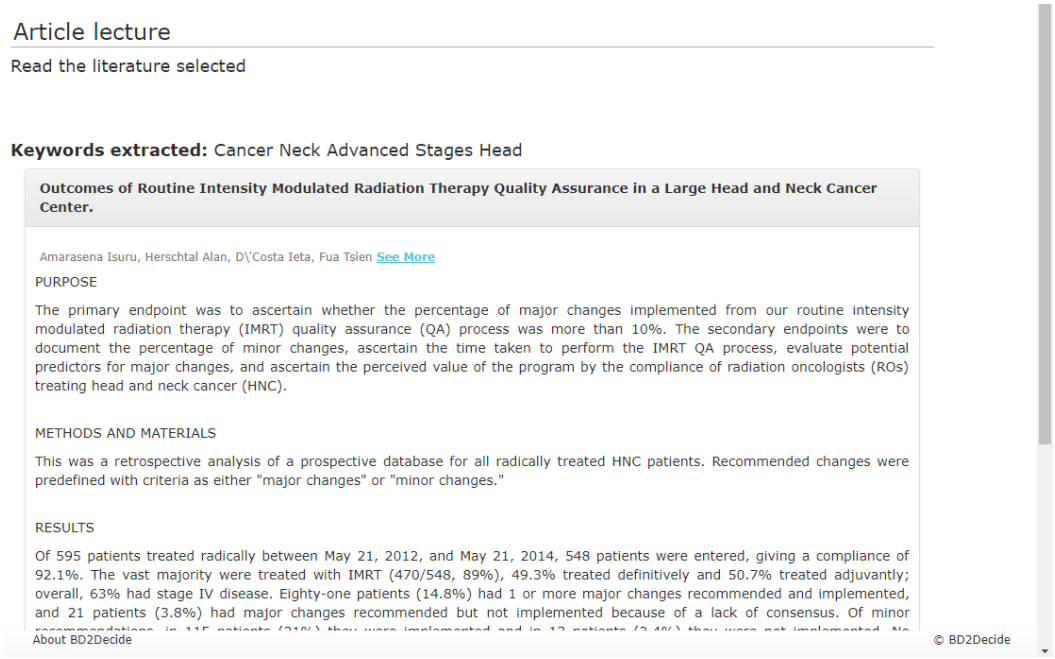

Figure 59: Article lecture presentation in the in VAT.

2. Statistical models

The provided statistical models are based on the selected data. For instance, the models for oral cavity will not be provided if the selected cohort includes only oropharynx tumour regions.

Figure 60: Visualising the results of Statistical modelling in VAT (mocked data). In this example the pathological model for oral cavity is presented: on the left the overall survival curves for a different groups of the patient cohort selected; on the right the comparison among the curve of reference, the baseline and the curve of the model refit. This last graph allows the users to validate the results.

The VAT will provide the possibility to select one or more models and to compare the results, based on the survival curves. Figure 60 presents a proposal for visualization.

3. Big data analysis

The ***Big Data analysis*** considers not only the data, but also the research problems and the outcome of interest (Figure 61). This workflow has been presented and discussed with clinicians in face-to-face meetings, in order to better support them in executing and understanding data analytics functionalities.

For instance, if the research problem is to forecast an outcome from the selected cohort, it means that the required analysis is a supervised learning type. Furthermore:

- If the users want to make a survival analysis, the tool provides the outcomes of models that identify factors with potential predictive utility, or that support the prediction of a given outcome (for instance Univariate algorithms, see Figure 62 as an example of use).
- If, on the contrary, the user wants to find relations between different cohorts, which is a classification problem, the tool will execute techniques such as Support Vector Machine to determine the association between different variables. Figure 63 provides an example of this option.

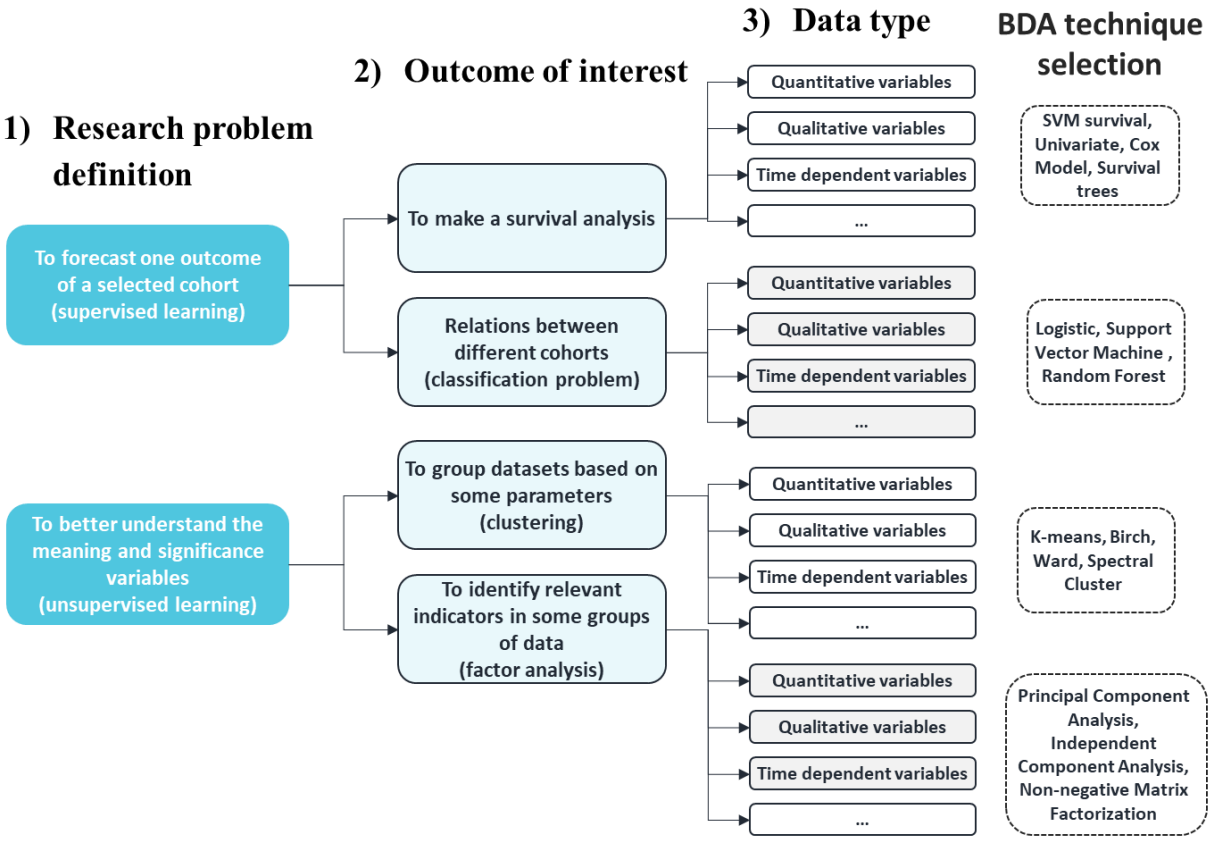

Figure 61: Guided analytics analysis based on researcher interests in the VAT tool.

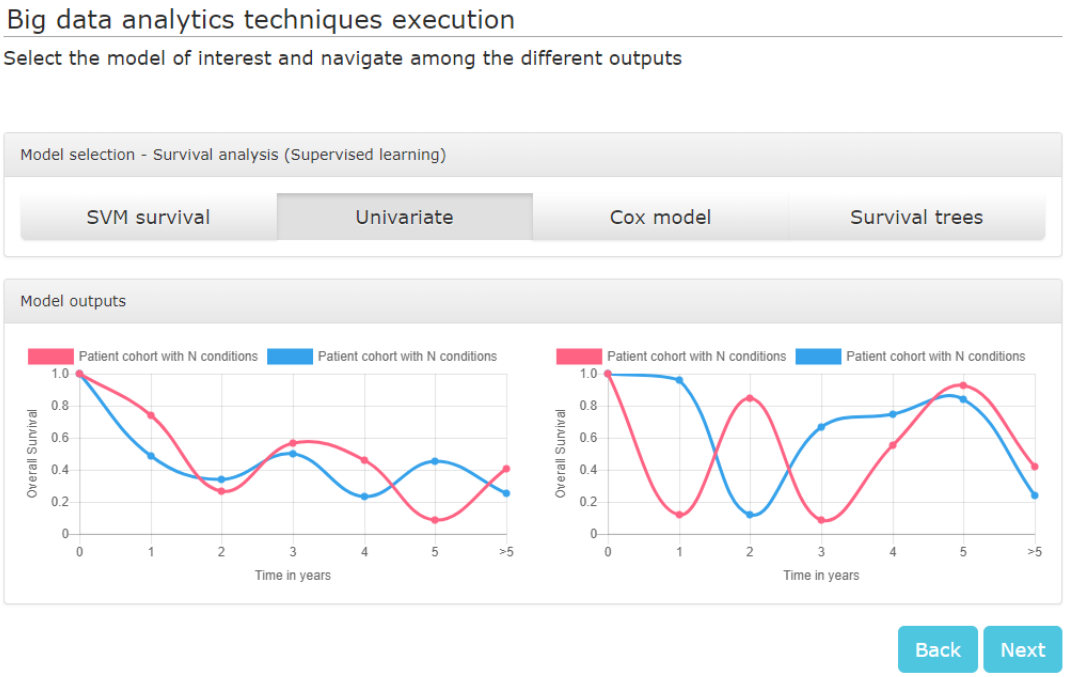

Figure 62: Visualising the results of big data analytics in VAT (mocked data). This example present the survival curves of the variables selected through the 'Inclusion criteria' in the project creation. The result are survival curves because the research questions are focused on survival analysis. In the final implementation, also the accuracy and other validation metrics will be provided.

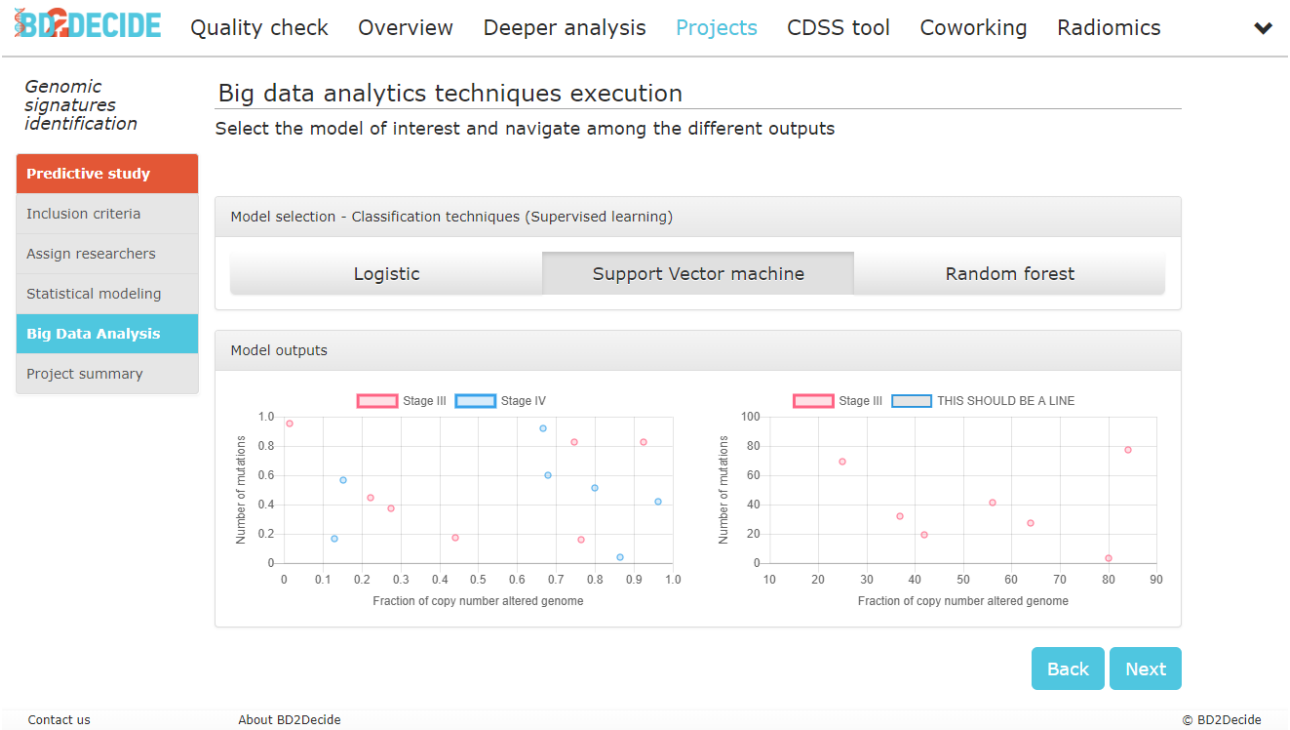

Figure 63: Big data analytics example of classification techniques representation (mocked data is used). On left the distribution of the patient cohort divided by Stage, and on the right the margin line to segregate different classes will be provided. In the final implementation, the results will be accompanied by model parameters to validate the selected algorithm.

4. Population data

Among all the features, the user is able to exploit other external data sources: a) to analyse the quality of care in different countries, b) to execute an epidemiology study, and c) to evaluate the impact of risk factors and context data in HNC. All these features, except the quality of care in different countries, are provided along the different activities during the project execution. Quality of care has its own section in the tool. Epidemiology studies provides the incidence and mortality rates of a population divided by country or region (Figure 64). Risk factors supports the understanding of the patient cohort context and because of that these kind of data is also combined within the different functionalities of the tool. Cancer registries are merged directly with the BD2Decide data to be used also in the statistical models and big data techniques.

5. Project summary

Finally, all the results of the research will be reported in a customizable summary (see Figure 65).

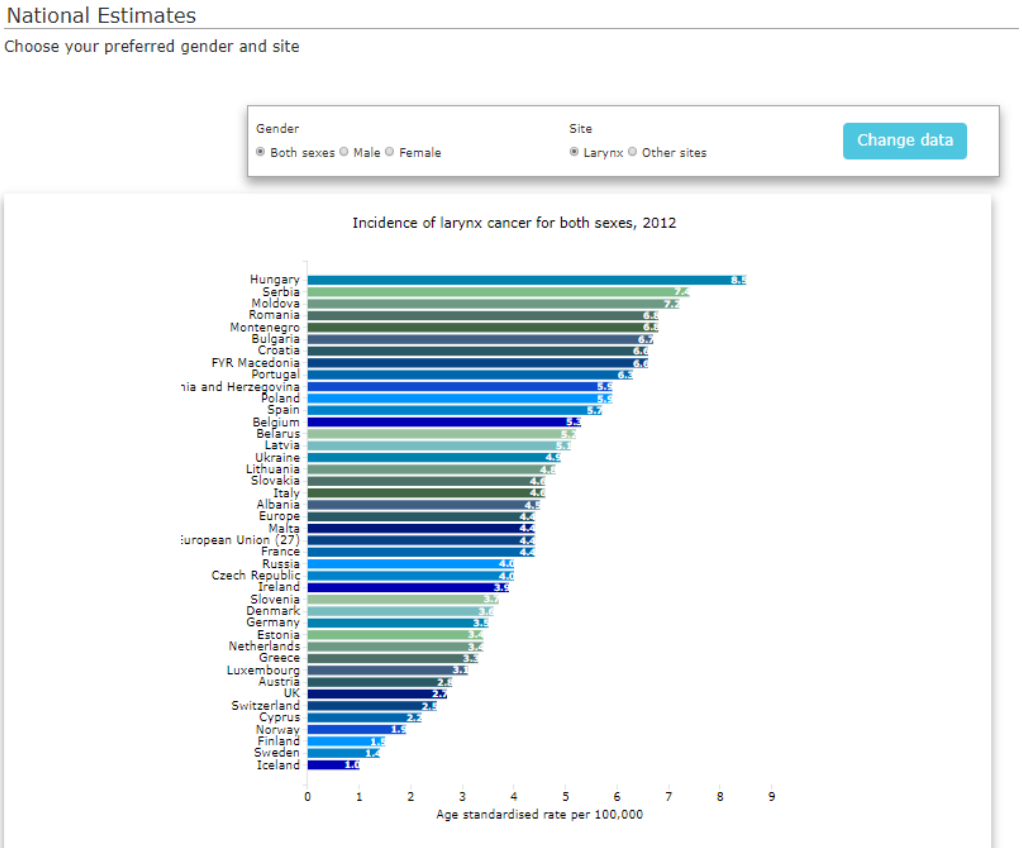

Figure 64: Epidemiological analysis example in the VAT tool.

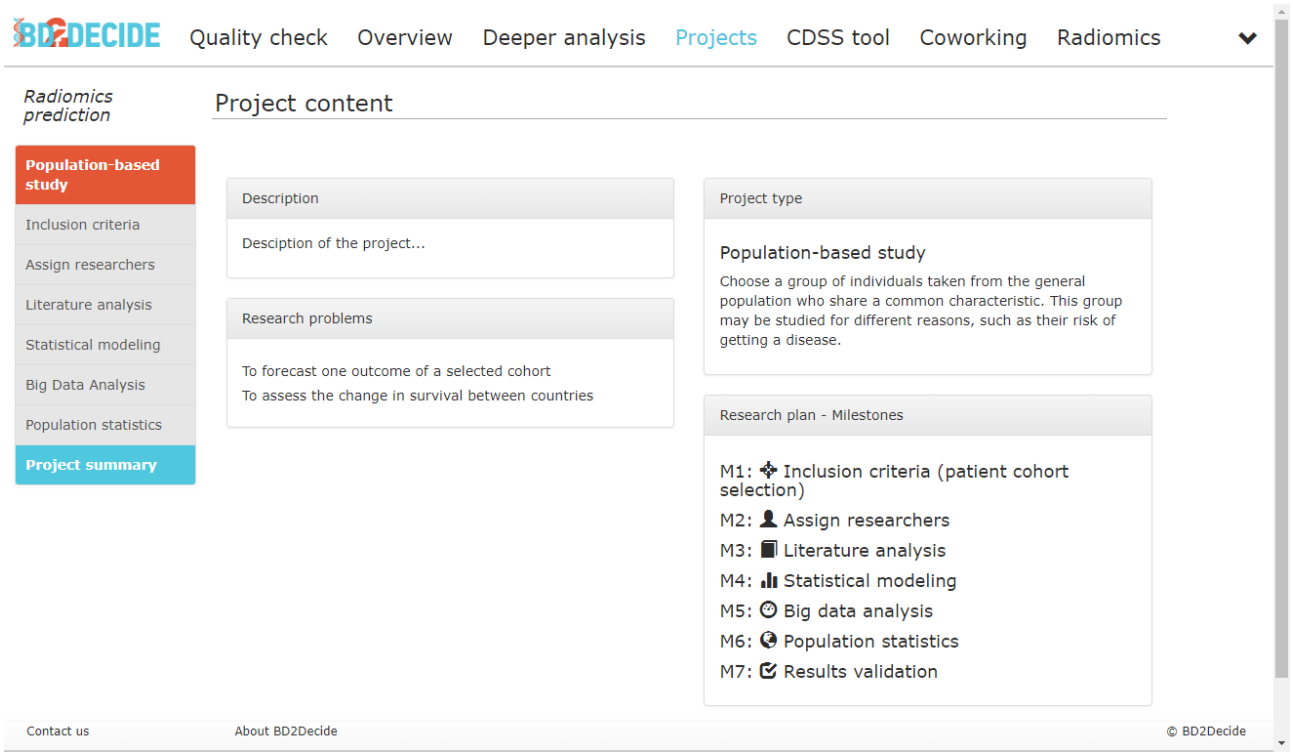

Figure 65: Visualising a project summary.

6. Radiomic feature extractor tool

In order to facilitate the extraction of radiomic features, the VAT will provide a module that extracts radiomics by only selecting the raw images and the segmentation images to be processed.

5.3 Implementation technologies

VAT technical specifications are represented in Figure 66 as layer diagram. The user interacts directly with the presentation layer, which holds the user interface of the tool. Below this layer, the application framework is responsible for providing the structure and data management support in the application.

RESTfull APIs provide access to the data located in the BDI and also handle the integration of the BD2Decide services: imaging, genomics, statistical models and big data analytics.

On top of all layers, there is the cross-cutting one, which orchestrates all the system features, the knowledge and data management, and also the security aspects.

The technical details of the components are summarized in Table 2.

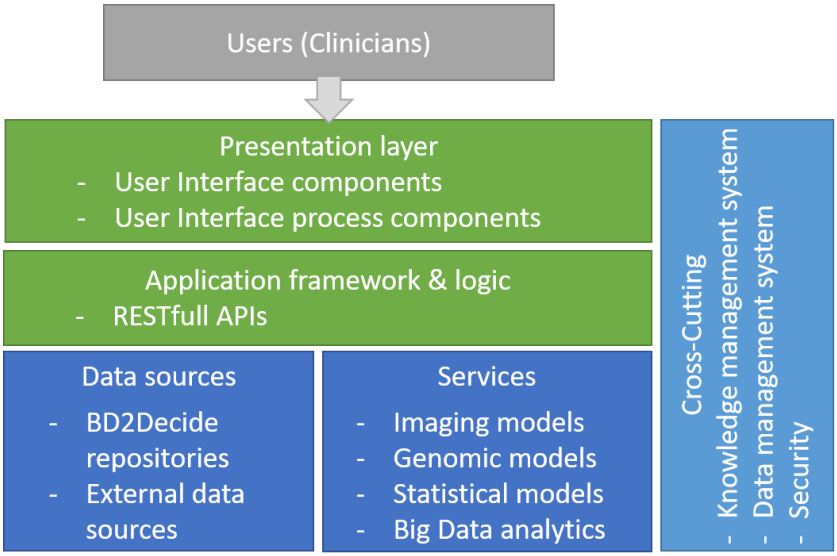

Figure 66: The *VAT* layer diagram.

Table 2. The technical component of the VAT tool.

Component	Formats/languages
Application type	Web
Web browser compatibility	IE, Firefox, Chrome, Safari and Opera
Language of the user interface	English
Front-end technologies and frameworks	HTML5, CSS and Javascript. Boostrap front-end framework. Plotly for data visualization development.
Back-end technology	Python (Django framework)
Database	PostgreSQL
Data and services access	All-in-image APIs PubMed e-utilities

VAT is developed in Python using Django framework[21]. PostgreSQL is the local database selected for the internal logic of the VAT. The access to BD2Decide data and services is available in real time through the APIs and synchronization features.

[21] https://www.djangoproject.com/

For the user interface, HTML5, CSS and Javascript have been used, for the data visualization, Plotly[22] was selected because this open-source library allows to represent data in a modern and powerful way, providing the users an interactive data visualization experience.

5.4 Integration with other tools

As shown in Figure 67, the VAT has integrated RESTfull APIs with different purposes:

- To get the different datasets available in the BDI (clinical, images, radiomics, genomics and population). The VAT uses it to allow clinicians to export the clinical dataset.
- To query the datasets available through SQL and SPARQL queries. SQL APIs are ready to use for the clinical dataset and for the population dataset.
- To store the literature results obtained through PubMed e-utilities services. This includes the tags requested by clinicians and the text analysis results through the Natural Language Toolkit (Python library to natural language processing).
- To access and integrate the statistical models and analytics techniques in the VAT.

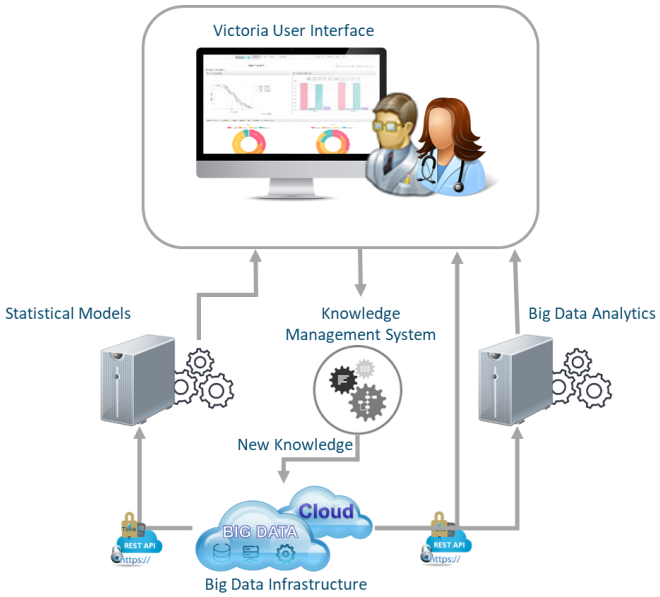

Figure 67: A summary of the VAT integration with other tools.

[22] https://plot.ly/

6 CONCLUSIONS AND NEXT STEPS

This deliverable presented the current implementation of the functional and working prototypes for the Clinical Decision Support System and the Visual Analytics Tool, which comprise the main visualisation tools of the BD2Decide platform. Specifically, the document summarised the results produced in the project the last 18 months following a user-oriented development approach for these tools, which involved the continuous engagement of clinicians, health professionals and researchers in the specification of the functional scenarios and mock-up screens that are expected from these tools.

With respect to CDSS, this deliverable provided the scenarios of use for supporting health professionals, either being physicians in a Head and Neck Cancer (HNC) case or undertaking a specific role in the HNC treatment process, like a surgeon, a medical oncologist, pathologist, a radiologist or a radiation oncologist, in making treatment decisions. The scenarios emphasise on the role of big data analytics for the combination of clinical, imaging, genomic and radiomic variables into aggregated visualisations for making assessments on the patient status and disease evolution. Furthermore, they virtualise the environment for hosting remote tumor boards, in which the health professionals can collaborate remotely in making the most appropriate treatment decisions for their HNC patients.

The document, also, presented the functional prototype screens for the VAT tool. This is used by clinical researchers in accomplishing their research tasks, which involve the exploitation, representation and visualization of information retrieved from large-scale and heterogeneous data sources in the field of HNC treatment. It gives a descriptive analysis of the BD2Decide data and facilitates on-demand queries on a specific HNC related dataset, without any knowledge of technical details. Further to it, the tool integrates services for a deeper investigation of selected HNC cohorts, based on statistical modelling and big data analytics techniques.

Following the outcome of this deliverable, in the next months, we will intensify the evaluation of the visualisation tools, through focus groups and hands-on workshops with clinicians and researchers. Through this process, we expect to calibrate the functions offered to the clinicians for supporting their decision-making on their HNC patients' treatment and the researchers for enhancing their investigation activities in identifying the influential variables in specific HNC cohorts. To this end, this deliverable D5.3 will feed the CDSS and VAT prototype implementation in the BD2Decide Deliverables D5.5 and D5.6 respectively, and it will drive the user validation activities, which will be reported in the in the BD2Decide Deliverable D5.7.

7 REFERENCES

[1] BD2Decide Deliverable D2.2, “User Interaction sketches”, August 2016.

[2] BD2Decide Deliverable D5.2, “The BD2Decide visualisation mockups concepts”, December 2016.

[3] BD2Decide Deliverable D5.4, “The IPDA prototype”, June 2017.

[4] Gregor Hohpe, Bobby Woolf, “Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions”, 1st Edition, ISBN-13: 978-0321200686, ISBN-10: 0321200683, http://www.enterpriseintegrationpatterns.com/.

[5] Baldoni, R.; M. Contenti, and A. Virgillito. "The Evolution of Publish/Subscribe Communication Systems." Future Directions of Distributed Computing”, Springer Verlag LNCS Vol. 2584, 2003.

[6] BDVA position paper, “BDVA’s response to the European Commission’s public consultation on Health and Care in the Digital Single Market”, BDVA Task Force 7 – Healthcare Subgroup, October 2017.

[7] A. Cooper, R. Reimann, D. Cronin, and C. Noessel, About face: The Essentials of Interaction Design, Fourth. John Wiley & Sons, Inc.