ADOPT BBMRI-ERIC
GRANT AGREEMENT NO. 676550

DELIVERABLE REPORT

Deliverable no	D3.3
Deliverable Title	User interface for collection of the colon cancer cases and database
Contractual delivery month	M12 (September 2016)
Responsible Partner	DKFZ
Author(s)	Rumyana Proynova

USER INTERFACE FOR COLLECTION OF THE COLON CANCER CASES AND DATABASE

Executive Summary

The BBMRI ADOPT project's goal is to accelerate the implementation and adoption of BBMRI infrastructure and services. The BBMRI infrastructure will include, among other initiatives, datasets of high quality data describing available biomaterial. The first such dataset is the colon cancer data collection CCDC. This report describes the ADOPT D3.3 M12 deliverable, which consists of system components needed to create the CCDC, specifically an interface for data entry. We created a data dictionary, which serves both as specification for the system and as the basis for a future ontology of colon cancer. We also implemented and set up the software for entering data about colon cancer cases. With this software, it is now possible to start gathering the data which will comprise the CCDC.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

Copyright notice

Document log

Issue	Date (yyyy-mm-dd)	Comment	Author/partner
D3.3_Rev1	2017-09-09	Revised to comply with GA art 29.4	Outi Törnwall

Table of Contents

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

1. Introduction

In recent years, biobanks have emerged as an important research resource in both the clinical and translational context. With the advance of medical knowledge, the demand for rare, specialized samples in good condition increases. A growing number of researchers is seeking access to high quality biomaterial outside of their institution. This poses new challenges for the biobanks' information infrastructure, which has to support the new modes of collaboration.

The goal of the ADOPT project is to enable biobanks to combine their efforts and provide a common, highly efficient platform to researchers and further stakeholders. One of ADOPT's deliverables is the Colon cancer data collection (CCDC), a pilot project which demonstrates the benefits of the new platform. It also provides the opportunity to test the newly developed software tools and fine-tune their performance. Setting up the CCDC includes creating the infrastructure needed to represent colon cancer samples, gathering this information from participating biobanks, and providing search tools for browsing the collection.

This report describes one of the early results needed for CCDC – an interface for manual data entry of colon cancer cases. With this milestone, several important components of the CCDC infrastructure are made available. These include the data entry forms, the database which will hold the collection, as well as the metadata repository which describes the semantics of the data structure used in CCDC. The data structure represents an important work result in its own right. It was created by a group of domain experts as part of the semantic ontology mapping task, and it represents a step towards harmonizing the available biobank data and creating a common data structure searchable from a single interface.

Currently, the interface is completed and running in a quality assurance environment. BBMRI members can test it online. When the remaining components needed for the CCDC have been developed (planned for later project stages), the complete system will be deployed in a production environment maintained by BBMRI CS-IT. The participating biobanks will then provide the data needed to complete the collection and make it available for use by the research community.

2. Background

Preliminary evaluations revealed that participating biobanks can provide data on roughly 10 000 colon cancer cases for the ADOPT colon cancer data collection. This data is currently available in heterogeneous systems running at each biobank. The majority of the data is stored in systems which have a sufficient degree of interoperability, such that their data can be transferred into the CCDC in a semi-automated ETL (extraction, transformation, loading) process. However, a substantial part of the data (estimate: 3000 cases) is stored in old, isolated systems and cannot be exported. For this data, manual data entry is the only way to utilize it for ADOPT.

The interface we developed is intended for the manual data entry of these 3000 cases. It can be used by medical documentation specialists without specialist IT knowledge.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

Use of the manual interface is not mandatory. If a biobank can provide the data electronically, it will be merged into the CCDC, bypassing the manual entry step.

The completed CCDC will be searchable in the same way as a biobank which is connected to the CS-IT infrastructure. It will thus serve as an example of the end state of the ADOPT project, and give researchers and biobankers a feeling of how the system will work after the biobanks are connected. However, the process of its creation is not the same as the process of connecting a biobank, so it should not be understood as a prescription of how biobanks are to be connected.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

3. Description of work and efforts

The CCDC uses a multi-tiered architecture compatible with other BBMRI CS-IT systems (Alexandre et al. 2016). Figure 1 is a schematic representation of the finalized architecture.

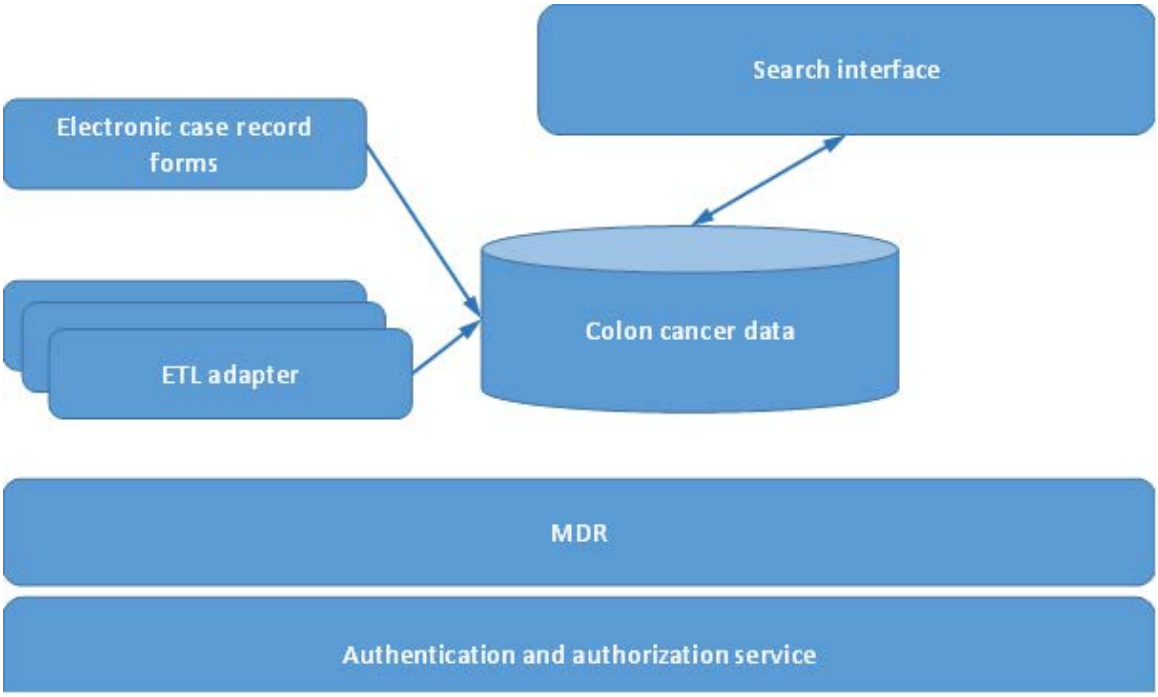

Figure 1 Architecture of the CCDC

The central element is the database which stores the colon cancer data. The data model is described in the metadata repository, to allow for harmonization with other data sources. As the original data is unlikely to be structured according to our data model in the source biobanks' systems, ETL adapter scripts are needed for semi-automated data transfer. For biobanks whose systems do not have the needed interfaces for a semiautomatic ETL, their data is manually entered into the database via a specialized interface. When the CCDC is deployed, users will be able to access and search the data via a search interface. Privacy and security issues are addressed through the use of a dedicated authentication and authorization service, which allows role based access control.

The current deliverable encompasses following artefacts:

- **Data dictionary**, a set of defined common data elements to be used for data entry via the manual data entry interface (Linarsson and Wigertz 1989)
- **Metadata repository**, a collection of metadata records corresponding to the data elements of the data dictionary (Strausberg et al. 2009)
- **Electronic case record forms (eCRF)**, a software component which allows a data specialist to enter colon cancer data
- **Data storage**, a database instance which can save the data entered in the data entry interface
- **Role concept**, a description of the roles and their permissions needed for creating and maintaining the collection

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

The ETL adapters, the search interface and the authentication and authorization software component will be delivered in a later stage, as agreed in the ADOPT project proposal.

3.1 Data dictionary

A group of domain experts and IT experts created the data dictionary which defines the data elements to be used for the CCDC. The goal was to create a comprehensive data dictionary for annotating the available biomaterial, such that the resulting data collection can be used for answering a wide range of research questions. The data dictionary not only serves as a specification for the data entry interface, but also provides a solid basis for developing a colon cancer ontology and providing a terminology service for semantic ontology mapping.

A *data dictionary* is defined as a set of defined *common data elements* (CDE) for data entry and retrieval by search criteria. Its data elements are systematically ordered in a hierarchical structure similar to a taxonomy and identified by a unique global key common to all participating institutions. The key and its designation constitute a defined data element, which is an IT representation of a medical attribute. To create the CCDC data dictionary, the expert group started by identifying a list of medically relevant attributes.

Unlike a collection developed to answer a specific research question, the CCDC is a general-purpose collection whose prospective users display a wide variety of information needs. The data dictionary reflects this by defining a basic set of data elements which cover the information typically recorded about a colon cancer case, thus enabling flexibility in servicing diverse queries. The medical experts ensured that the list of items to be included in the dictionary contains items with high clinical and scientific relevance.

The IT experts created an entity relationship model based on the medical experts' list, using input from the medical experts to determine the relationship between the entities and their attributes. They ensured that all data elements are disjoint, unambiguous, and that for each data element, there is a well-defined range of values.

The data elements were based on typical data gathered during the course of a colon cancer case, describing both diagnostic and therapeutic measures. Where possible, the items were based on existing classification systems, such as the ICD-10 for diagnosis and the UICC scale for tumor grading. For the remaining answers, the vocabulary was chosen such that it reflects typical clinical use and minimizes potential ambiguity.
The data dictionary can be extended, if user feedback determines that this is necessary.

3.2 Metadata repository

A metadata repository is a database that contains the metadata on all data elements to be used for a given data collection. The BBMRI ERIC architecture requires that the data model of all data collections is described in a metadata repository as opposed to being implemented as a data schema in the database management system itself. This allows the semantic mapping of data coming from different

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

sources to the data model chosen for the collection and is a best practice when pursuing the goal of semantic interoperability.

We used the open source product Samply.MDR to create the metadata repository. This is a specialized metadata repository system which fulfills the requirements of the ISO 11179 Metadata Registry standard (ISO 11179) and is compatible with other software products employed in BBMRI CS-IT. It supports the definition of common data elements with a unique designation, one or more definitions, and a value range. Its versioning mechanism allows tracking semantics even if the data model changes in the future. It is under active development, and further features can be added as needed to support the goals of BBMRI.

The data elements defined in the data dictionary were entered in the metadata repository, using a hierarchical structure provided by the expert group. The result is a human- and machine-readable representation of the data dictionary which can be used for multiple purposes, including 1) generation of eCRFs, 2) harmonization of data gathered from heterogeneous source systems, and 3) validating data contained in local biobank and clinical information systems.

3.3 Electronic case record forms

The actual data entry process will be carried out using Web based entry forms. Their structure is based on the data model defined in the data dictionary.

The expert group identified two relevant entities for colon cancer cases, a *patient* and a *sample*. Each sample is donated by one patient, while each patient can donate many samples. As the CCDC is biomaterial driven, we will not have information about patients who have not donated. This results in the cardinality 1:n for the patient-sample relationship. The attributes of each entity are defined in the data dictionary and described as data elements in the metadata repository.

We used the MDRFaces software tool to generate two web forms consistent with the metadata repository described in the previous section. The first of the forms contains fields for the attributes of the patient, the other contains the fields for the sample. Data elements whose value range consists of a list of predefined strings are represented as drop-down boxes, while other elements (e.g. numeric ones) are validated on form submission. The validation helps mitigate data entry errors.

Each patient in CCDC is identified by a pseudonym only. The interface does not foresee the entry of personally identifying data. The data specialist (who enters the data) is expected to provide a pseudonym used in their own biobank information system, which will allow employees of the biobank to trace a patient and its samples if requested.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

3.4 Data storage

The electronic case record forms provide a front end for creating the data collection. The corresponding back end is a database intended for storing the CCDC data. It is based on a PostgreSQL database management system, an open source product used throughout BBMRI. In future, it will hold all data in CCDC – both the records entered through the eCFRs and the records transferred semi-automatically.

The back end is generated together with the front end. The data schema does not correspond to the data model, which is described in the MDR. Rather, it makes use of PostgreSQL's JSON storage capabilities to allow on-the-fly data model changes driven by the MDR.

The CCDC only supports the storage of harmonized data, without mappings. Any data entered either through the forms or through an import has to be in the goal schema.

3.5 Role concept

We created a simple role concept which covers the necessary roles for the CCDC. It contains the three roles of data specialist, administrator and researcher. This is a fairly standard model based on experience from previous projects. In case some deviations are needed, our software product allows changes during runtime without new development effort

- **Data specialist**
  This role is intended for the person who enters data. A person with this role can create new patients and edit the data in the patient form and the sample form.
- **Researcher**
  This role has read-only access to the data entered about patients and their samples.
- **Administrator**
  This person maintains the data collection and can provide some technical support. Permissions include creating new user accounts, editing the roles, and deploying new forms. The administrator does not have access to the patient data.

If the biobanks who provide the data will desire to keep it confidential, it is possible to configure it in a way that it is only visible to the data specialist who entered it.

4. Results

The outcome of our efforts is twofold, consisting of the data entry interface and the data dictionary. The data entry interface is a self-contained part of the CCDC infrastructure which can be deployed in a live environment and used for entering data. The data dictionary is not only the basis for the data entry interface, but also a preliminary step towards creating a colon cancer ontology.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

4.1 Data entry interface

For the user, the eCFRs in the last section are the most salient part of the data entry interface. We present here partial screenshots of the two forms. The data in the screenshots describes a fictional patient with a realistic colon cancer case and was entered for demonstration purposes. It does not correspond to a case actually present in a participating biobank.

Figure 2 contains a screenshot of the top part of the patient form. The patient is identified by the pseudonym 5D846PmBvWxN, shown in the upper left corner. Below that, there is a list of the "episodes", each episode representing an event at which samples were donated. It is used for navigation to the samples form. Below that, the patient form is displayed, with its fields corresponding to the data elements defined in the MDR. For example, the "biological sex" field is a dropdown with two possible values (female and male) and the "participation in clinical study" field is a checkbox, since it can only be true or false.

When an episode is selected, the user sees the sample form pertaining to that donation episode. Figure 3 shows a screenshot of that form. It describes the histopathology of the samples as defined in the data dictionary. The data specialist can enter unlimited episodes per patient, and fill the sample form for each of them.

4.2 Data dictionary

The data dictionary can be seen as a first step towards developing a full-fledged ontology for colon cancer. It provides a basic set of entities and attributes to be included in the ontology and takes into account the actual level of information available in clinical and biobank information systems.

The data dictionary is also suitable for semantically mapping the data from the source biobanks to a standard data model used for the CCDC. The interface described in this report allows this to happen in a manual data entry process, while the future ETL adapters will use the data model to provide semi-automatic mapping. This leads to harmonization of the diverse source data, and allows users to discover comparable data with a single search. The data dictionary is therefore an important intermediate result for creating semantic interoperability of biobank data in ADOPT.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

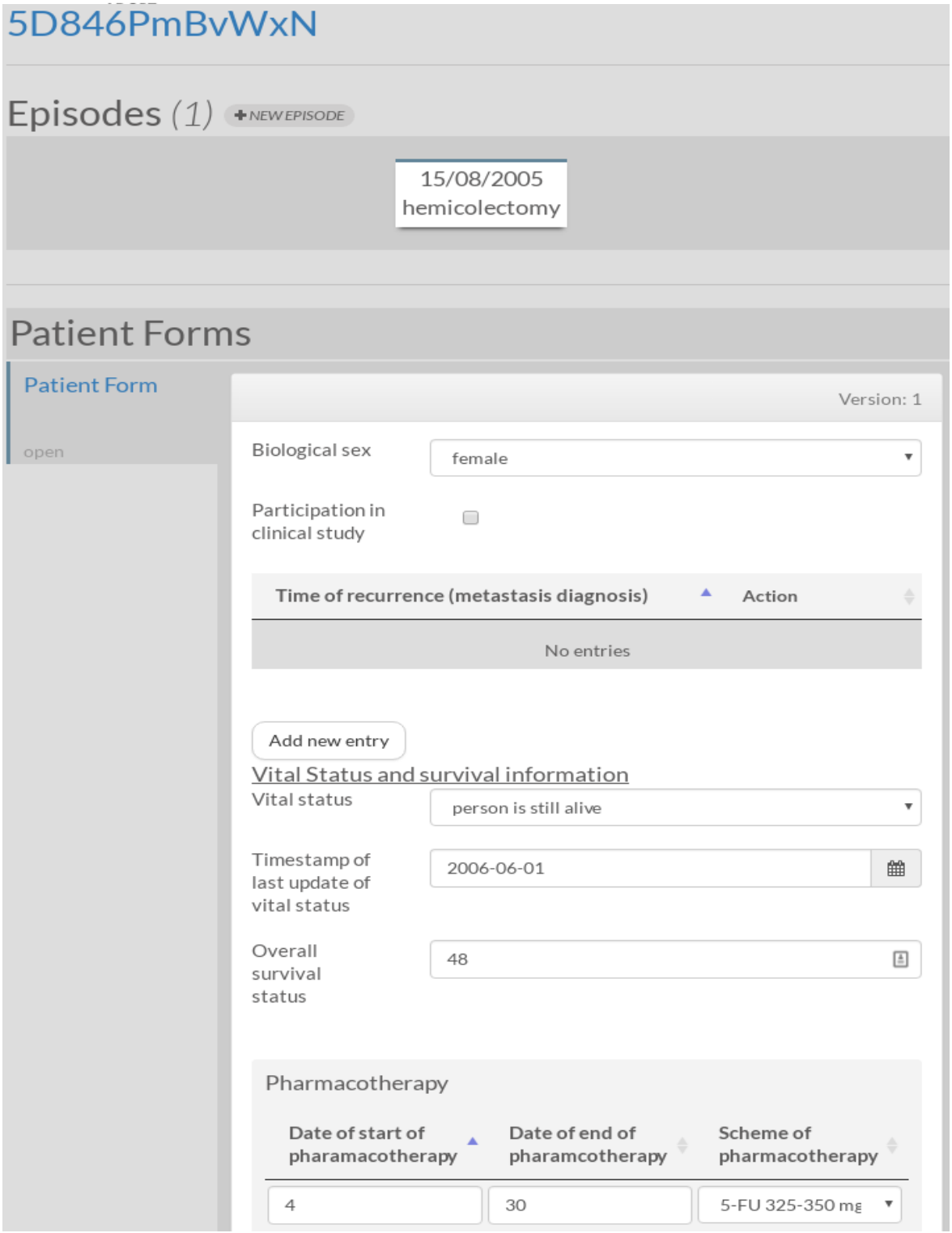

Figure 2 Patient form for the CCDC

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

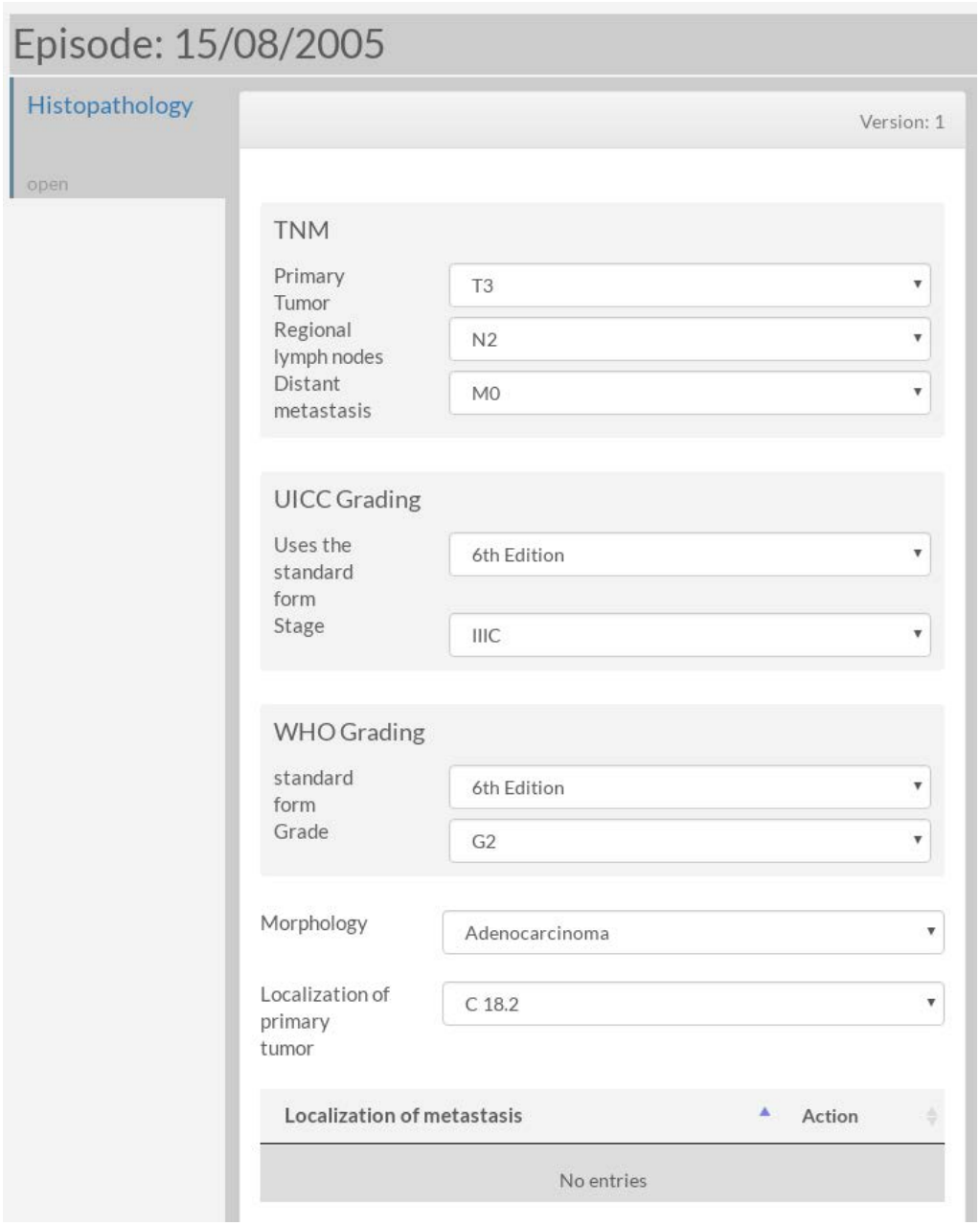

Figure 3 Sample form for CCDC

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

5. Next Steps

The work described in this report is a prerequisite for completing both D3.3 (Ontology-based tool set for mapping of the biobanking terminologies) and D2.2 (Data set on samples from colorectal cancer patients) of the BBMRI ADOPT project. Next steps can be defined for both of these deliverables.

The CCDC deliverable (D2.2) requires the technical infrastructure for creating the data collection. This deliverable included several components of this infrastructure. While they are sufficient for a self-contained system, they do not provide the full functionality needed for creating the CCDC. As described in section “Description of work and efforts”, the final infrastructure also needs to include ETL adapters. Our next development effort will include writing these adapters and extending the system with the appropriate RESTful APIs needed for communication with the adapters and the future search interface. The search interface itself is part of the BBMRI CS-IT project and will also be developed in the next months. Additionally, the system is currently using a Samply based authentication and authorization service for demonstration purposes. BBMRI-ERIC will use a unified authentication and authorization service for all its components, and connecting the system to it is an additional task needed before the data entry can commence.

This project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 676550.

References

ALEXANDRE ET AL. 2016 Alexandre, D; Lablans, M; Ückert, F; Holub, P; Hummel, M; Swertz, M; Proynova, R; van Enckevort, D; Jetten, J. BBMRI-ERIC CS-IT Architecture Notebook. Deliverable report for ADOPT D3.3, submitted September 2016.

ISO 11179 ISO/IEC 11179: Information Technology -- Metadata registries (MDR), 2015-11-05, International Organization for Standardization, Geneva, Switzerland. http://metadata-standards.org/11179/

LINNARSSON AND WIGERTZ 1989 Linnarson R; Wigartz O. The data dictionary – a controlled vocabulary for integrating clinical databases and medical knowledge bases. *Methods of information in medicine* 28.2 (1989): 78-85.

STRAUSBERG ET AL. 2009 Stausberg, J., Löbe, M., Verplancke, P., Drepper, J., Herre, H., & Löffler, M. (2009). Foundations of a metadata repository for databases of registers and trials. *Studies in health technology and informatics* 150 (2008): 409-413.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

Appendix I

This appendix contains the data dictionary created for this deliverable. It is represented in two tables, one for the data items relating to the entity *patient*, another for the data items relating to the entity *sample*. The items in each are grouped under headings. Each heading has a name and cardinality, for example the heading "Surgery (0...n)" means that there can be between zero and infinity surgeries per patient.

The columns in the table have following meaning:

- Label: Unique label of the variable.
- Cardinality: How many times the element can appear, related to the element group as defined by the heading.
- Short description of the variable - to be used in forms.
- Semantics = definition of meaning.
- Syntax
    - including data type (elementary types such as boolean, float, integer, free text, specifically structured text, etc., array or lists of elementary types),
    - including coding (e.g., IEEE 754 for floats, regular expressions for structured text).
- List of allowed units
    - including their conversion algorithms (with "non-existent" and "unknown" interim options)
- Level: REQUIRED, OPTIONAL, RECOMMENDED
    - REQUIRED means the data can't be entered at all without this item being provided,
    - OPTIONAL means data may or may not be provided, but the item will be ready for inputting the data in as part of the data model,
    - RECOMMENDED is a special subclass of the OPTIONAL, which is highly-recommended to be filled in (intended for items where we need the data but where we know that some sources won't be able to fill this in and we still want such data not being discarded as invalid).

The combination of the Level column and headings works in the following way:
if the heading says "0…n", then it is possible that none of the following fields are entered. E.g. if a biobank has no information on surgeries, or a patient did not undergo surgery, no surgery information is added at all. But if at least one field under that heading is added, then all REQUIRED fields from that heading must be filled once. So for a surgery, it is impossible to add the time and radicality but leave out the surgery type.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

Common data elements for the Patient entity

Label	Cardinality	Description	Semantics	Syntax	Units	Level
Patient information (1)						
SEX	1	Biological sex	Biological sex of the person, defined by chromosomes.	male, female (only 2 values allowed)	n/a	REQUIRED
CLINICAL_STUDY_PARTICIPANT	1		Participant of clinical study	boolean	n/a	RECOMMENDED
AGE_AT_PRIMARY_DIAGNOSIS	1	Age at diagnosis (rounded to years)	Age at initial histopathological diagnosis (biopsy or surgical specimen of the primary tumor) rounded to years.	integer	years since birth	REQUIRED
TIME_OF_RECURRENCE_RELATIVE	0...n	Time of recurrence (metastasis diagnosis)	Weeks between primary diagnosis (AGE_AT_PRIMARY_DIAGNOSIS) and diagnosed recurrence.*	integer	weeks since primary diagnosis	OPTIONAL

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

Vital status and survival information (1)						
VITAL_STATUS	1	Vital status	living or deceased	list (ALIVE= … person is still alive, DEATH_COLON_CANCER = death due to colon cancer, DEATH_OTHER = death due to other reasons, DEATH_UNKNOWN_REASON = death for unknown reasons, UNKNOWN = unknown)	n/a	REQUIRED
VITAL_STATUS_TIMESTAMP	1	Timestamp of last update of vital status	Timestamp of last update of vital status	timestamp compliant to ISO 8601, date part without time	n/a	RECOMMENDED
OVERALL_SURVIVAL_STATUS	1	Overall survival status	Weeks after first colon cancer therapy started for the given person*	integer	weeks	REQUIRED
Surgery (0...n)						
SURGERY_START_RELATIVE	1	Time difference between initial diagnosis and surgery	Weeks between initial diagnosis and date of surgery. *	integer	weeks	REQUIRED
SURGERY_RADICALITY	1	Surgery radicality	Whether the surgery removed the entire tumor was removed during surgery.	list (RX, R0, R1, R2)	n/a	REQUIRED

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

SURGERY_TYPE	1	Type of surgery	"OTHER" value may allow for optional "please specify" free text option	list (RIGHT_HEMICOLECTOMY, LEFT_HEMICOLECTOMY, TRANSVERSE_COLECTOMY, SIGMOID_COLECTOMY, TOTAL_COLECTOMY, PAN-PROCTO_COLECTOMY, LOW_ANTERIOR_COLON_RESECTION, ANTERIOR_RESECTION_OF_RECTUM, ABDOMINO-PERINEAL_RESECTION, ENDO-RECTAL_TUMOR_RESECTION, OTHER)	n/a	REQUIRED
Pharmacotherapy (0...n)						
PHARMACOTHERAPY_START_RELATIVE	1	Date of start	start of the drug intake in weeks since initial diagnosis.	integer	weeks	REQUIRED
PHARMACOTHERAPY_END_RELATIVE	1	Date of end	end of the drug intake in weeks since initial diagnosis.	integer	weeks	REQUIRED

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

PHARMACOTHERAPY_SCHEME	1	Scheme of pharmacotherapy	A scheme of administering the drugs, as specified in literature	list(5-FU 325-350 mg/m2 + LV 20 mg/m2 i.v. bolus, day 1-5, weeks 1 and 5; 5-FU 400 mg/m2 + LV 100 mg i.v. bolus, d. 1,2,11,12,21,22; 5-FU 1000mg/m2 i.v. continuous infusion, day 1-5, weeks 1 and 5; Capecitabine 800-825 mg/m2 bid po, day 1-5, together with radiation or continuously until end of radiation; UFT (300-350 mg/m2/day) and LV (22.5-90 mg/day) po continuously, 5(-7) days per week, together with radiotherapy; Only preoperatively (no standard): 5-FU 250 mg/m2 i.v. continuous infusion on days 1-14 and 22-35 and oxaliplatin 50 mg/m2 i.v. day 1,8,22 and 29	n/a	REQUIRED
Targeted therapy (0...n)						
TARGETED_THERAPY_START_RELATIVE	1	Date of start	start of the drug intake in weeks since initial diagnosis.	integer	weeks	REQUIRED
TARGETED_THERAPY_END_RELATIVE	1	Date of end	end of the drug intake in weeks since initial diagnosis.	integer	weeks	REQUIRED

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

Response to therapy (0...n)						
THERAPY_RESPONSE	1	Specific response	Therapy response according (to RECIST criteria when possible)	list (PROGRESSIVE_DISEASE, STABLE_DISEASE, PARTIAL_RESPONSE, COMPLETE_RESPONSE)	n/a	REQUIRED
THERAPY_RESPONSE_TIMESTAMP_RELATIVE	1	Time of specific response	Timestamp when the therapy response was obtained, in weeks relative to the initial diagnosis*	integer	weeks	REQUIRED
Diagnostic exam (1)						
DIAG_COLONOSCOPY		Colonoscopy	whether colonoscopy was done	list(NOT_DONE, POSITIVE, NEGATIVE)	n/a	REQUIRED
DIAG_LIVER_IMAGING_DONE		Liver imaging	whether this diagnostics was done	list (NOT_DONE, DONE_DATA_AVAILABLE, DONE_DATA_NOT_AVAILABLE, UNKNOWN)	n/a	REQUIRED

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

DIAG_LUNG_IMAGING_DONE		Lung imaging	whether this diagnostics was done	list (NOT_DONE, DONE_DATA_AVAILABLE, DONE_DATA_NOT_AVAILABLE, UNKNOWN)	n/a	REQUIRED
DIAG_MRI_DONE		MRI	whether this diagnostics was done	list (NOT_DONE, DONE_DATA_AVAILABLE, DONE_DATA_NOT_AVAILABLE, UNKNOWN)	n/a	REQUIRED
DIAG_CT_DONE		CT	whether this diagnostics was done	list (NOT_DONE, DONE_DATA_AVAILABLE, DONE_DATA_NOT_AVAILABLE, UNKNOWN)	n/a	REQUIRED

* If only months are available, conversion factor is 1 month = 4 weeks

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

Common data elements for the Sample entity

Label	Cardinality	Description	Semantics	Syntax	Units	Level
TNM (1)						
HIST_TNM_PRIMARY_TUMOR	1	TNM classification	Guidelines for UICC-TNM classification of tumors	list(TX, T0, Tis, T1, T2, T3, T4a, T4b)	n/a	REQUIRED
HIST_TNM_LYMPH_NODES	1	TNM classification	Guidelines for UICC-TNM classification of tumors	list(NX, N0, N1, N1a, N1b, N1c, N2, N2a, N2b)	n/a	REQUIRED
HIST_TNM_METASTASES	1	TNM classification	Guidelines for UICC- TNM classification of tumors	list(M0, M1, M1a, M1b)	n/a	REQUIRED
UICC Staging (1)						
UICC_STAGE	1	UICC stage	Tumor stage as defined in the UICC guidelines	list((0, I, IIA, IIB, IIIA, IIIB, IIIC, IVA, IVB)	n/a	REQUIRED
UICC_STANDARD	1	UICC standard used	The edition of the UICC guidelines used for staging	List from 1st to 7th edition	n/a	REQUIRED
WHO Grading (1)						

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.

WHO_GRADE	1	WHO grade	Tumor grade as defined in the WHO guidelines	list(G1, G2, G3, G4)	n/a	REQUIRED
GRADE_STANDARD	1	WHO standard used	The edition of the WHO guidelines used for grading	List from 1st to 7th edition	n/a	REQUIRED
Other						
HIST_MORPHOLOGY	1	morphology	Specified in a WHO standard	list (ADENOCARCINOMA, MUCINOUS_CARCINOMA, SIGNET-RING_CELL_CARCINOMA, MEDULLARY_CARCINOMA, HIGH-GRADE_NEUROENDOCRINE_CARCINOMA, LARGE_CELL_NEUROENDOCRINE_CARCINOMA, SMALL_CELL_NEUROENDOCRINE_CARCINOMA, SQUAMOUS_CELL_CARCINOMA, ADENOSQUAMOUS_CARCINOMA, MICROPAPILLARY_CARCINOMA, SERRATED_ADENOCARCINOMA, SPINDLE_CELL_CARCINOMA, MIXED_ADENONEUROENDOCRINE_CARCINOMA, UNDIFFERENTIATED_CARCINOMA, OTHER)	n/a	REQUIRED
HIST_LOCALIZATION	1	localization	As defined in ICD-10	C18.1 to C18.7, C19, C20	n/a	REQUIRED
HIST_METASTASIS	0...n	metastasis	Organ with metastasis	list(NONE, PULMONARY, OSSEOUS, HEPATIC, BRAIN, LYMPH_NODES, BONE_MARROW, PLEURA, PERITONEUM, ADRENALS, SKIN, OTHERS)	n/a	REQUIRED

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 676550.