

Standard operating procedure for minimal data exchange for the BBMRI-ERIC Colorectal Cancer Cohort.

Dataset Name (acronym):	BBMRI-ERIC colorectal cancer cohort (CRC-cohort)
Description of the Dataset:	The CRC-Cohort collection is a joint long-term European endeavor, with contributions from existing, well-established biobanks. Currently, the CRC-Cohort numbers over 10,000 datasets from across Europe
Data Type(s):	Genomic, Clinical, and Pathology
EOSC4Cancer Use Case:	T4.3
SOP contact person: (name, institution, email)	Kurt Majcen, BBMRI-ERIC (kurt.majcen@bbmri-eric.eu)
Date:	5-12-2023

A. Purpose

This SOP describes the steps required for Genomic, Clinical, and Pathology data from CRC-Cohort to be accessed and used by researchers in the context of the EOSC4Cancer project and beyond (e.g. raw data access and download and/or via analysis and visualization platforms like cBioPortal). This SOP is created for data scientists who want to access the datasets shared within the EOSC4Cancer project. Based on this information, we will define and suggest broader guidelines for data exchange within and beyond the EOSC4Cancer project. Below, you will find an overview of the data models, ontology, vocabulary and dictionaries used by the dataset. Furthermore, data access procedures are described in a step-by-step way.

B. List of definitions and abbreviations

<In case definitions and abbreviations are used, please explain them here>

Definitions	
HNPCC	Hereditary nonpolyposis colorectal cancer (also known as Lynch syndrom)

MSI	Microsatellite instability
-----	----------------------------

C. Data Models and ontology, vocabulary, dictionaries

<Please fill in the detailed information about how your data set is structured. Feel free to expand the table according to the number of data models, ontologies, vocabularies, etc. you have. In case this information was published or released publically, please provide a reference or a link to this material. Please specify if a potential conversion with other data sets (within EOSC4Cancer and otherwise) already exists>

Genomic	Name	Description	References	Is this standard?	Available conversions for harmonization
Data Models	Data model for CRC-Cohort	The data model of the CRC-Cohort includes Molecular markers, like <ul style="list-style-type: none"> • Microsatellite Instability • Mismatch Repair Gene Expression-IHCarrayfordifferentgenes • Risk Situation • RAS mutation status • BRAF, PIC3CA.HER2 mutation status 	data model https://doi.org/10.5281/zenodo.7930536 See "Data Model for CRC-Cohort" (appendix B / section Molecular markers) in BBMRI-ERIC Colorectal Cancer Cohort (CRC-Cohort): Data Protection Policy	no, custom	Converted to format openEHR
Vocabularies/ Ontologies	Revised Bethesda Guidelines for HNCand Microsatellite Instability	Bethesda Guidelines for identifying individuals at risk for HNPCC; and recommend criteria for MSI testing.	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2933058	no	
Dictionaries/ Codebooks	NA				
Formats	-	-	-	-	-

Clinical	Name	Description -custom-made for project?... -based on?...	References	Is this standard?	Available conversions for harmonization
----------	------	--	------------	-------------------	---

Data Models	Data model for CRC-Cohort	<p>The data model of the CRC-Cohort includes Molecular markers, like</p> <ul style="list-style-type: none"> • Sex • Participation In Clinical Study • Age At Primary Diagnosis • Known Risk Factors for CRC • Time Of Recurrence(metastasis) and availability of biological material from recurrence • Vital Status And Survival Information • Surgery • Pharmacotherapy Targeted therapy • Radiation therapy • Response to therapy • Diagnostic Exam 	https://doi.org/10.5281/zenodo.7930536 See "Data Model for CRC-Cohort" (appendix B) in BBMRI-ERIC Colorectal Cancer Cohort (CRC-Cohort) : Data Protection Policy	no, custom	Converted to formats openEHR / OMOP / HL7-FHIR
Vocabularies/ Ontologies	SNOMEDCT	SNOMED Clinical Terms	http://purl.bioontology.org/ontology/SNOMEDCT/423493009	yes	
	Phenotype And Trait Ontology	An ontology of phenotypic qualities (properties, attributes or characteristics)	http://purl.obolibrary.org/obo/PATO_0020000	no	
Dictionaries/ Codebooks	NA				

Formats	-	-	-	-	-
---------	---	---	---	---	---

Pathology	Name	Description <i>-custom-made for project?...</i> <i>-based on?...</i>	References	Is this standard?	Available conversions for harmonization
Data Models	Data model for CRC-Cohort	The data model of the CRC-Cohort includes "Histopathology part".	https://doi.org/10.5281/zenodo.7930536 See "Data Model for CRC-Cohort" (appendix B / section Histopathology part) in BBMRI-ERIC Colorectal Cancer Cohort (CRC-Cohort): Data Protection Policy		
Vocabularies/ Ontologies	NA				
Dictionaries/ Codebooks	ICD-10	The ICD-10 is the 10th version of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list of the World Health Organisation (WHO)	https://icd.who.int/browse10/2019/en	yes	
Formats	dicom	Image format		Work in progress by BigPicture	Compatible with xOpat
	tiff	Image format			

D. Description of the data exchange process

<Describe all steps of the data access process, including timelines and who is responsible. Please include all stages, starting from sending a data access request until the data transfer is complete >

Step	Action	Responsible (Name, email)	Link	Estimated time
0	Public (non-authenticated) search and browsing access to meta-data describing the CRC-Cohort for discovery and select CRC-Cohort in BBMRI-ERIC Directory	Researcher	https://directory.bbmri-eric.eu	
1-5	See BBMRI-ERIC Policy for general process on Access to and Sharing of Biological Samples and Data (appendix H of BBMRI-ERIC Colorectal Cancer Cohort (CRC-Cohort): Data Protection Policy) for a step-by-step description The CRC-Cohort has an expedite procedure directly between BBMRI-ERIC and the requester. This replaces step 3 of the general procedure and the details can be found in Annex I.	Researcher, Access Committee (at BBMRI-ERIC) and biobanks contributing to CRC-Cohort	https://doi.org/10.5281/zenodo.7930536	approx. 1 month
2-1	Public (non-authenticated) search and browsing access to meta-data describing the CRC-Cohort		https://directory.bbmri-eric.eu/#/catalogue	
2-2	Secure authenticated search access by querying the anonymized dataset with a low residual risk of re-identification		https://www.bbmri-eric.eu/services/sample-location/	
2-3	Secure authenticated access to the pseudonymized data (=personal data) controlled by the Access Committee established for Mode-3 will include BBMRI-ERIC and biobank	Jens Habermann		1 month

E. Related documents and links

<If you refer to specific documents or formats, please list them here and include the location of the documents>

e.g. applicable law, guidelines, forms or legal documents

1. BBMRI-ERIC Colorectal Cancer Cohort (CRC-Cohort): Data Protection Policy (V1.5: <https://doi.org/10.5281/zenodo.7930536>)
2. []
3. []