CRC Cohort mapping and conversion to HL7 FHIR resources

Revision history

Revisions	Authors	Reviewers	Main changes
V0 - 14th Dec 2022	Cecilia Mascia, Vittorio Meloni, Alessandro Sulis and Francesca Frexia (CRS4)		First version: extended description of the mapping and the implementation choices made for the openEHR to HL7 FHIR resources conversion of of the dataset

Scope and Objectives of this document

In the context of the EOSC-Life WP1 Demonstrator "Cloudification of BBMRI-ERIC CRC-Cohort and its Digital Pathology Imaging" (APPID 1228), it has been developed a demonstrator to improve the access and re-use to the ColoRectal Cancer-Cohort (CRC-Cohort)[1] dataset, providing clinical and imaging data, with related annotations, in documented formats using common and open approaches and tools.

To improve the integrability and re-usability of the dataset, the demonstrator enables the conversion of the CRC-Cohort clinical data to the openEHR[2] format, in order to save them in an openEHR-based repository, from which they can be exported in OMOP[3] and HL7 FHIR[4], in order to support the future integration into the BBMRI-ERIC Federated Platform, as shown in Figure 1.

All the fields in the CRC-Cohort Data Model have been mapped onto an openEHR template[5] and the dataset has been converted to a set of openEHR compositions stored in an openEHR-compliand backend (EHRBase). Considering this dataset as the data source, the mapping of the minimum data fields required for the federated search to the HL7 FHIR Profiles and OMOP CDM tables has already been done and described in a previous document. This document instead describes more extensively the mapping and the current implementation of the conversion tool of the CRC Cohort data model to the Sample Locator[6] FHIR Profiles.

[1] https://www.bbmri-eric.eu/scientific-collaboration/colorectal-cancer-cohort/
[2] https://www.openehr.org/
[3] https://www.ohdsi.org/data-standardization/the-common-data-model/
[4] https://hl7.org/FHIR/
[5] https://github.com/crs4/crc_cohort_modelling
[6] https://www.bbmri-eric.eu/scientific-collaboration/colorectal-cancer-cohort/

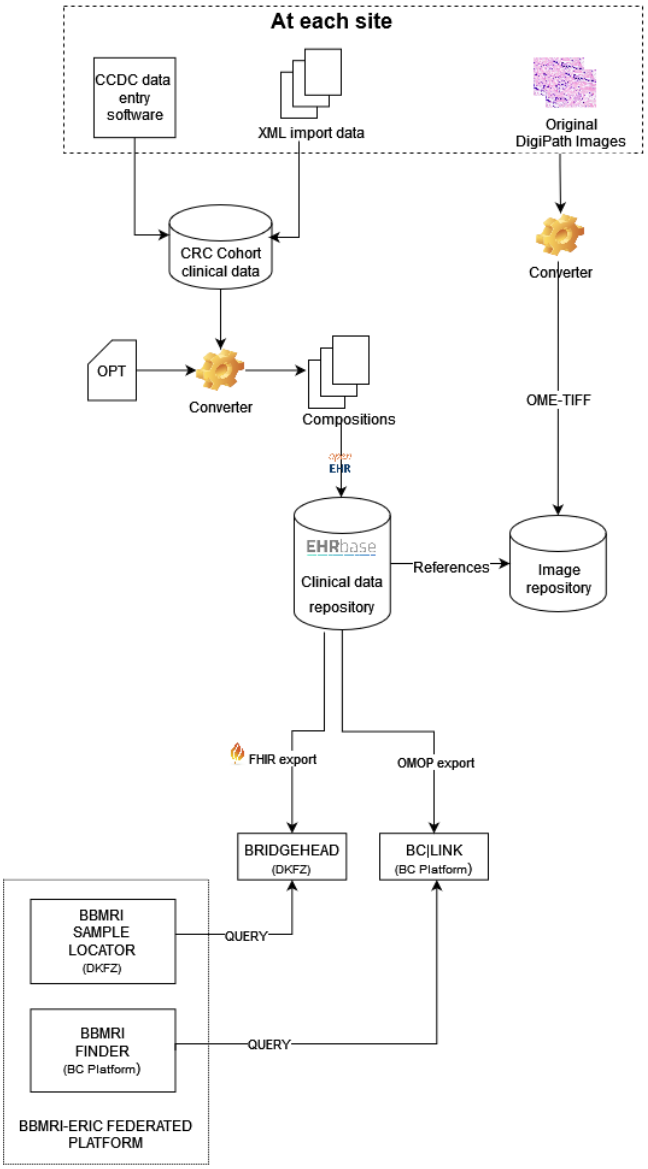

Figure 1 - Architectural diagram of the demonstrator

The set of developed tools[7] enables the extraction and the transformation of the data needed to fill in the following resources, according to the proposed mapping:

- Patient
- Condition
- Specimen

Two further resources are generated by the tool statically, filled with a set of known information[8]. In particular:

[7] URL: https://github.com/crs4/crcc-fhir-importer/tree/main/fhir_importer
[8] URL: https://directory.bbmri-eric.eu/menu/main/app-molgenis-app-biobank-explorer/#/biobank/bbmri-eric:ID:EU_BBMRI-ERIC

- Collection resource represents the CRC Cohort, acting as the Custodian of all the Specimens and part of the Biobank;
- Biobank resource represents the BBMRI Organization, to which the cohort belongs.

This document details how the openEHR fields are expressed according to the HL7 FHIR Profiles, in order to **support both the sharing of the mapping/conversion approach and the validation of the mapping/conversion itself with domain experts**, which is **required to verify the quality of the content**.

FHIR Profiles mapping

This mapping relates the current openEHR data model[9], expressed in the form of a template, to the elements of the FHIR Profiles of the German Biobank Node, Version: 1.2.0[10].

The mapping has the FHIR resources as a starting point, and in particular, all the elements labelled as "MustSupport". For each of them, an analogous element was sought in the openEHR model that was similar in terms of semantics, data type and constraints (e.g., allowed choices for a coded textual element). We define the mapping of each element as "valid" when these aspects are considered to be sufficiently overlapping or when it is possible to adapt the source element (openEHR) to the intended form of the target element (FHIR resources fields). However, given the difference in the content structure of the two approaches, there are some special cases for which a direct mapping is not so obvious or even not possible. These cases, marked as "uncertain", "not applicable" or "not done", are dependent on specific choices to be made in the ETL phase.

In the following paragraphs, we will give an **overview of the current mapping** and the **implementation/conversion choices** made so far in the tools to extract data from the openEHR database and convert them to FHIR resources. In particular, for each resource, we present a mapping table structured as follows:

- the *FHIR/openEHR* columns and subcolumns contain, respectively, the details of the destination and the source of the data transformation for each data element;
- the *Notes/implementation choices* column contains any remaining critical point, the choices made (*Current version*) and the actions to be taken for the continuation of the work (*Action required*);
- the *Mapping confidentiality* column contains the level we assign to the mapping of each element, with the marks: **V** - valid; **U** - uncertain; **NA** - not applicable; **ND** - not done.

When relevant, a specific table containing the mapping of the values to the FHIR terminology value sets and code systems is also included.

All the following tables contain only those fields that are currently involved in the ETL process and will be expanded in later versions of this document when the mapping will be extended.

[9] CRC model: https://github.com/crs4/crc_cohort_modelling/blob/main/templates/opt/crc_cohort_rev.opt
[10] BBMRI.de/GBA Implementation Guide : https://samply.github.io/bbmri-fhir-ig/index.html

Patient

The Patient resource is the resource representing a sample donor.

HL7 FHIR				openEHR			Notes/implementation choices	Mapping confidentiality[11]
Field Name	**Cardinality**	**MustSupport ?**	**Datatype**	**node name(XSD label)**	**occurrences**	**datatype**		
Patient.id	0..1		string				**Current version:** the converter assigns to this field the identifier of the patient.	V
Patient.identifier.value	0..*		string	Patient pseudonym (XSD label: Identifier)	Mandatory	TEXT	This field holds the original pseudonym.	V
Patient.gender	0..1	**Y**	code	Biological sex (XSD label: Dataelement_85_1)	Mandatory	TEXT	In FHIR the value is bound with AdministrativeGender	V
Patient.birthDate	0..1	**Y**	date	_ _ _	_ _ _	_ _ _	This information is missing in the data source. **Current version:** during the ETL, the year of birth is calculated as: year of diagnosis (XSD label: Dataelement_51_3) - age at diagnosis (XSD label: Dataelement_3_1)	U
Patient.deceased[x]	0..1	**Y**	dateTime	_ _ _	_ _ _	_ _ _	This information is missing in the data source.	NA

[11] Mapping confidentiality levels: **V** - valid; **U** - uncertain; **NA** - not applicable; **ND** - not done

Condition

HL7 FHIR				openEHR			Notes/implementation choices	Mapping confidentiality [12]
Field Name	**Cardinality**	**MustSupport?**	**Datatype**	**node name(XSD label)**	**occurrences**	**datatype**		
Condition.id	0..*		string	– – –	– – –	– – –	**Current version:** the converter assigns to this field a string made of the original identifier of the patient + '-condition' (e.g., '1234-condition').	V
Condition.code	0..1	**Y**	CodeableConcept	Localization of primary tumor (XSD Label: Dataelement_92_1)	Mandatory	CODED TEXT		V
Condition.code.coding.code	0..1		code	– – –	– – –	– – –	**Current version**: we assigned values to these elements considering the primary tumour diagnosis code as shown in the value mapping table A (third and fourth columns).	V
Condition.code.coding.display	0..1		string	– – –	– – –	– – –		V
Condition.code.coding.system	1..1		uri	– – –	– – –	– – –	**Current version**: the value is fixed to "http://hl7.org/fhir/sid/icd-10"	V
Condition.subject	1..1	**Y**	Reference(Patient)	– – –	– – –	– – –	Reference to the Patient.id (e.g., 'Patient/1234')	V
Condition.encounter	0..1	**Y**	Reference(Encounter)	– – –	– – –	– – –	This is not applicable to this use case.	NA
Condition.onset[x]	0..1	**Y**	dateTime	Date of diagnosis (XSD Label: Dataelement_51_3)	Optional	DATETIME		V

[12] Mapping confidentiality levels: **V** - valid; **U** - uncertain; **NA** - not applicable; **ND** - not done

Condition.recordedDate	0..1		dateTime	Date of diagnosis (XSD Label: Dataelement_51_3)	Optional	DATETIME	**Current version**: we assigned to this field the same value of the date of diagnosis, assuming it matches the one in which the data is recorded. **Action required**: Check if it is acceptable or if it is preferable to remove this field from the resource (as it is not mandatory)	U

Mapping for "Condition.codes" values - Value mapping table A

openEHR		HL7 FHIR	
Source Code (OMOP codes)	**Source Text**	**Code**	**Display**
432837	C 18.0 - Caecum	C18.0	Localization of primary tumor - C18.0
433143	C 18.1 - Appendix	C18.1	Localization of primary tumor - C18.1
4247719	C 18.2 - Ascending colon	C18.2	Localization of primary tumor - C18.2
438979	C 18.3 - Hepatic flexure	C18.3	Localization of primary tumor - C18.3
432257	C 18.4 - Transverse colon	C18.4	Localization of primary tumor - C18.4
437798	C 18.5 - Splenic flexure	C18.5	Localization of primary tumor - C18.5
441800	C 18.6 - Descending colon	C18.6	Localization of primary tumor - C18.6
436635	C 18.7 - Sigmoid colon	C18.7	Localization of primary tumor - C18.7
438699	C 19 - Rectosigmoid junction	C19	Localization of primary tumor - C19
74582	C 20 - Rectum	C20	Localization of primary tumor - C20

Specimen

HL7 FHIR				openEHR			Notes/implementation choices	Mapping confidentiality [13]
Field Name	**Cardinality**	**MustSupport?**	**Datatype**	**node name(XSD label)**	**occurrences**	**datatype**		
Specimen.id	0..1		string	Sample ID (XSD Label: Dataelement_56_2)	Mandatory	IDENTIFIER	**Current version:** the converter assigns to this field a "normalised" version of the identifier of the specimen (e.g.: "123/4" becomes "123-4").	V
Specimen.storageTemperature	0..*	**Y**	CodeableConcept	– – –	– – –	– – –	In the original dataset this information is missing. **Action required**: Check if this information can be derived from the biobank protocol for sample management.	NA
Specimen.diagnosis	0..*	**Y**	CodeableConcept	– – –	– – –	– – –	This should be the diagnosis associated to the specimen, and it should be: - the ICD 10 code if the Specimen.type is "tumor tissue" - one of the possible value from the Other Sample Diagnosis code system if the Specimen.type is "normal tissue" or "tissue other" **Action required**: Check if this logic is correct and implement it in the converter.	**ND**
Specimen.custodian	0..1	**Y**	Reference(Collection)	– – –	– – –	– – –	**Current version**: this field points to "Organization/bbmri-eric-ID-EU-BBMRI-ERIC-collections-CRC-Cohort"	

[13] Mapping confidentiality levels: **V** - valid; **U** - uncertain; **NA** - not applicable; **ND** - not done

Specimen.identifier	0..*		Identifier	Sample ID (XSD Label: Dataelement_56_2)	Mandatory	IDENTIFIER	This field holds the original Sample ID	V
Specimen.type	0..1	**Y**	CodeableConcept	Combination of: Material type (XSD Label: Dataelement_54_2) and Preservation mode (XSD Label:Dataelement_55_2)	– – –	– – –	In FHIR the value is bound with Sample material type.	V
Specimen.type.coding.code	0..1		code	– – –	– – –	– – –	**Current version**: We derive this information from the combination of two fields in the original data according to the rules expressed in the value mapping table B.	V
Specimen.type.coding.display	0..1		string	– – –	– – –	– – –		V
Specimen.type.coding.system	0..1		uri	– – –	– – –	– – –	**Current version**: the value is fixed to "https://fhir.bbmri.de/CodeSystem/SampleMaterialType"	V
Specimen.subject	0..1	**Y**	Reference(Patient)	– – –	– – –	– – –	Reference to the Patient.id (e.g., 'Patient/1234')	V
Specimen.collection	0..1	**Y**	BackboneElement					
Specimen.collection.collected[x]	0..1	**Y**	dateTime	Year of sample collection (XSD Label: Dataelement_89_3)	Mandatory	DATE	In the data source there is only the year and not the full date. **Current version**: we set the date with the actual year of acquisition (YYYY) and 01 - 01 for months and days. **Action required**: Check if it is possible to record only the year.	V
Specimen.collection.bodySite.coding	0..1	**Y**	Coding				In the original dataset this information is missing.	NA

Specimen.collection.fastingStatus[x]	0..1	**Y**	CodeableConcept				In the original dataset this information is missing.	NA

Mapping for “Specimen.type” values - Value mapping table B

openEHR				**HL7 FHIR**	
Source Code (OMOP codes)	**Source Text**	**Source Code (OMOP codes)**	**Source Text**	**Code**	**Display**
4134449	Tissue specimen from colon	4338536	Cryopreservation	normal-tissue-frozen	Normal tissue (frozen)
4134449	Tissue specimen from colon	37206850	FFPE	normal-tissue-ffpe	Normal tissue (FFPE)
4134449	Tissue specimen from colon	4190569	Other	tissue-other	Other tissue storage
4122248	Tumor tissue sample	4338536	Cryopreservation	tumor-tissue-frozen	Tumor tissue (frozen)
4122248	Tumor tissue sample	37206850	FFPE	tumor-tissue-ffpe	Tumor tissue (FFPE)
4122248	Tumor tissue sample	4190569	Other	tissue-other	Other tissue storage
NS_4163599	Other specimen type	4338536	Cryopreservation	other-tissue-frozen	Other tissue (frozen)
NS_4163599	Other specimen type	37206850	FFPE	other-tissue-ffpe	Other tissue (FFPE)
NS_4163599	Other specimen type	4190569	Other	tissue-other	Other tissue storage

Biobank

The following table does not contain the three openEHR columns as it is not a proper mapping but more a description on how the resource is built.

HL7 FHIR				Notes/implementation choices	Mapping confidentiality[1] [4]
Field Name	**Cardinality**	**MustSupport?**	**Datatype**		
Organization.id	0..1		string	**Current version:** we assign to this field the string "bbmri-eric-ID-EU-BBMRI-ERIC"	V
Organization.OrganizationDescription	1..1	**Y**	string	**Current version:** we assign to this field the following string "BBMRI-ERIC does not act as a biobank itself but in special cases it acts as custodian of data collected centrally on European scale."	V
Organization.OrganizationjuridicalPerson	1..1	**Y**	string	**Current version:** we assign to this field the string "BBMRI-ERIC"	V
Organization.OrganizationqualityStandard	0..*	**Y**	CodeableConcept		V
Organization.OrganizationqualityStandard.coding.code (1)	0..1		string	**Current version:** we assign to this field the string "oecd-guidelines"	V
Organization.OrganizationqualityStandard .coding.system (1)	0..1		uri	**Current version:** we assign to this field the following string "https://fhir.bbmri.de/CodeSystem/QualityStandard"	V
Organization.OrganizationqualityStandard.coding.code (2)	0..1		string	**Current version:** we assign to this field the string "iso-15189"	V
Organization.OrganizationqualityStandard .coding.system (2)	0..1		uri	**Current version:** we assign to this field the string "https://fhir.bbmri.de/CodeSystem/QualityStandard"	V
Organization.identifier:Bbmri-EricId	1..1	**Y**	Identifier		V
Organization.identifier.system	0..1	**Y**	uri	**Current version:** we assign to this field the value "http://www.bbmri-eric.eu/"	V

[14] Mapping confidentiality levels: **V** - valid; **U** - uncertain; **NA** - not applicable; **ND** - not done

Organization.identifier.value	0..1	**Y**	string	**Current version:** we assign to this field the string "bbmri-eric-ID-EU-BBMRI-ERIC"	v
Organization.name	1..1	**Y**	string	**Current version:** we assign to this field the string "Central data collections collected by BBMRI-ERIC"	v
Organization.alias	0..1	**Y**	string	**Current version:** we assign to this field the string "bbmri-eric:ID:EU_BBMRI-ERIC"	v
Organization.telecom	0..*	**Y**	ContactPoint	This element is bound to the ContactPointSystem value set.	v
Organization.telecom.system	0..1		code	**Current version:** we assign to this field the code "url"	v
Organization.telecom.value	0..1		string	**Current version:** we assign to this field the string "https://directory.bbmri-eric.eu/"	v
Organization.address	1..*	**Y**	Address		v
Organization.address.city	0..1		string	**Current version:** we assign to this field the string "Graz"	v
Organization.address.country	1..1	**Y**	string	**Current version:** we assign to this field the string "AT"	v
Organization.address.line	0..*		string	**Current version:** we assign to this field the string "Neue Stiftingtalstrasse 2/B/6"	v
Organization.address.postalCode	0..1		string	**Current version:** we assign to this field the string "8010"	v
Organization.contact.purpose	1..1	**Y**	CodeableConcept	This element is bound to the ContactEntityType value set	v
Organization.contact.purpose.coding.code	0..1		string	**Current version:** we assign to this field the string "ADMIN"	v
Organization.contact.purpose.coding.display	0..1		string	**Current version:** we assign to this field the string "Administrative"	v
Organization.contact.purpose.coding.system	0..1		uri	**Current version:** we assign to this field the string "http://terminology.hl7.org/CodeSystem/contactentity-type"	v
Organization.contact.name.family	1..1	**Y**	string	**Current version:** we assign to this field the string "Petr"	v
Organization.contact.name.given	1..*	**Y**	string	**Current version:** we assign to this field the string "Holub"	v
Organization.contact.name.prefix	0..*	**Y**	string	**Current version:** we assign to this field the string "Dr."	v

Organization.contact.telecom.system	0..1	**Y**	code	This element is bound to the ContactPointSystem value set **Current version:** we assign to this field the code "phone"	V
Organization.contact.telecom.value	0..1	**Y**	string	**Current version:** we assign to this field the string "+433163499170"	V

Collection

The following table does not contain the three openEHR columns as it is not a proper mapping but more a description on how the resource is built.

HL7 FHIR				Notes/implementation choices	Mapping confidentiality[1][5]
Field Name	**Cardinality**	**MustSupport?**	**Datatype**		
Organization.id	0..1		string	**Current version:** we assign to this field the string "bbmri-eric-ID-EU-BBMRI-ERIC-collections-CRC-Cohort"	V
Organization.OrganizationDescription	1..1	**Y**	string	**Current version:** we assign to this field the string "Collection of more than 10,000 European cases of colorectal cancer, collected as a part of the ADOPT BBMRI-ERIC project. Centrally collected data set acts as a proxy for finding the contributing biobanks. Complete data sets as well as biological samples (at least FFPE samples) are available via access procedure, involving the contributing biobanks."	V
Organization.CollectionType	1..*	**Y**	CodeableConcept	This element is bound to the Collection Type value set	V
Organization.CollectionType.coding.code (1)	0..1		string	**Current version:** we assign to this field the code "SAMPLE"	V
Organization.CollectionType.coding.system (1)	0..1		uri	**Current version:** we assign to this field the string "https://fhir.bbmri.de/CodeSystem/CollectionType"	V
Organization.CollectionType.coding.code (2)	0..1		string	**Current version:** we assign to this field the code "DISEASE_SPECIFIC"	V

[15] Mapping confidentiality levels: **V** - valid; **U** - uncertain; **NA** - not applicable; **ND** - not done

Organization.CollectionType.coding.system (2)	0..1		uri	**Current version:** we assign to this field the string "https://fhir.bbmri.de/CodeSystem/CollectionType"	V
Organization.DataCategory	1..*	**Y**	CodeableConcept	This element is bound to the Data Category value set	V
Organization.CollectionType.coding.code (1)	0..1		string	**Current version:** not implemented yet	**ND**
Organization.CollectionType.coding.system (1)	0..1		uri	**Current version:** not implemented yet	**ND**
Organization.identifier:Bbmri-EricId	1..1	**Y**	Identifier		V
Organization.identifier.system	0..1	**Y**	uri	**Current version:** we assign to this field the value "http://www.bbmri-eric.eu/"	V
Organization.identifier.value	0..1	**Y**	string	**Current version:** we assign to this field the string "bbmri-eric-ID-EU-BBMRI-ERIC-collections-CRC-Cohort"	V
Organization.name	1..1	**Y**	string	**Current version:** we assign to this field the string "CRC-Cohort"	V
Organization.alias	0..1	**Y**	string	**Current version:** we assign to this field the string "CRC-COHORT"	V
Organization.telecom	0..*	**Y**	ContactPoint	This element is bound to the ContactPointSystem value set.	V
Organization.telecom.system	0..1	**Y**	code	**Current version:** we assign to this field the code "url"	V
Organization.partOf	1..1	**Y**	Reference (Biobank)	**Current implementation:** this element contains the reference to the Biobank resource: "Organization/bbmri-eric-ID-EU-BBMRI-ERIC"	V
Organization.contact.purpose	1..1	**Y**	CodeableConcept	This element is bound to the ContactEntityType value set	V
Organization.contact.purpose.coding.code	0..1		string	**Current version:** we assign to this field the string "ADMIN"	V
Organization.contact.purpose.coding.display	0..1		string	**Current version:** we assign to this field the string "Administrative"	V

Organization.contact.purpose.coding .system	0..1		uri	**Current version:** we assign to this field the string "http://terminology.hl7.org/CodeSystem/contactentity-type"	v
Organization.contact.name.family	1..1	**Y**	string	**Current version:** we assign to this field the string "Petr"	v
Organization.contact.name.given	1..*	**Y**	string	**Current version:** we assign to this field the string "Holub"	v
Organization.contact.name.prefix	0..*	**Y**	string	**Current version:** we assign to this field the string "Dr."	v
Organization.contact.telecom.system	0..1	**Y**	code	This element is bound to the ContactPointSystem value set **Current version:** we assign to this field the code "phone"	v
Organization.contact.telecom.value	0..1	**Y**	string	**Current version:** we assign to this field the string "+433163499170"	v

Next steps

The next actions we will carry out are the finalisation of the mapping of the missing data elements and the extension of the tools for their extraction and conversion from the openEHR DB into FHIR resources.

In parallel, general feedback on the whole mapping and any input on the following specific task would be useful for the continuation of the work:

- Check each specific question highlighted in the text.
- Each mapping of the values (i.e., tables A and B) should be checked and validated by a domain expert.