

Mapping Between CLRC-MD and CERA

(Shoaib Sufi - s.a.sufi@dl.ac.uk)

1. ABBREVIATIONS

CLRC-MD	CCLRC Metadata format
CERA	

2. MAPPING NOTES

2.1 Introduction

CLRC-MD is an XML based standard and the CERA metadata model is a relational database model.

Naturally in the XML model we have an Element Hierarchy – with attributes and in the CERA model we have Columns and Tables with relational links.

2.2 Columns in The Mapping

In the mapping we show:

- Parent - which is the parent of the XML Element – or if an attribute is stated then it is the Element to which the attribute belongs
- Element - is an XML element
- Attribute - is an XML attribute (naturally of an element)
- Table - is the Table name in the CERA Meta information Database
- Column - is the Column name in the CERA metadata database which is the counterpart of the Element or attribute from the CLRC-MD model
- Key_Table - If the Column is a foreign key this value holds the table name holding the primary key
- Key_Column - If the Column is a foreign key this is the key column in the Key_table which is matched to the Column
- Required_Column - If the Column matches onto a Key_Column in the Key_Table the Required_Column is a column in the Key_Table whose actual value is required – i.e. maps to the CLRC-MD Element (or attribute)
- Value - For the case where the meta data in one model may in fact be the data in another model

2.3 Occurrence Constraints

This can be one of:

Entry Value	Explanation
1	Mandatory and there is only one instance
0..n	Optional and the maximum amount is unbounded (i.e. there is no maximum)

1..n	Mandatory and the maximum number of occurrences is unbounded
0..1	Optional but there can only be 0 or 1 Occurences
i..j	where 'i' and 'j' are +ve integers where 'i' >= 0 and 'j' >= 1 and 'i' <= 'j'

If there is more then 1 element that maps to a particular column in a table we may have to construct the conditional part of the query with something like :

... WHERE COLUMN IN ('ELEMENT_VAL1', 'ELEMENT_VAL2' ...)

2.4 Type

Entry Value	Explanation
simple	i.e. element has an atomic type (e.g. string, int)
complex	element contains other elements but no direct simple content of its own
mixed	a complex element with simple content of its own
empty	element contains no content e.g. just attributes

2.4.1 When mapping types

An element with simple content will usually map onto a particular column in a particular table.

An element with complex content will not usually map onto a particular column but may in due course have most of its simple children mapped onto columns from one particular (or a closely related set of tables).

Mixed element - don't foresee many of these.

Empty content - these usually won't have a mapping unless the attributes map onto columns somewhere (e.g. sequence numbers in the database perhaps or even foreign keys).

Also, all attribute values are by definition expressed by simple types.

2.5 What is a CERA Entry

On deciding whether the CERA information specifies a Programme | Experiment | Measurement | Simulation - it would appear that Measurement is the natural choice.

2.6 Column Values

Unless otherwise stated the Element (or Attribute) actual value is mapped to the Column (or Required_Column) actual value and vice-a-versa.

The 'Value' column is included for situation where there is an implicit constant or the actual column name is the value which is mapped from/to the XML element (e.g. in situation where there is metadata in CLRC-MD stored inside the data elements in CLRC-MD but maps to metadata in the CERA Model e.g. see the mapping of Condition->ParamName in Section 3).

2.7 Date format

Format	Meaning
YYYY	4 digit year (e.g. 1971)
MM	months 01-12 (January-December)
DD	days 01-31
24HH	24 hour clock 0000-2359
MI	minutes 00-59
SS	seconds 00-59
TTT	Timezone e.g. +00 for GMT -05 for EST(I think)

2.8 Foreign Keys

In some situation it is necessary to trace the details of a foreign key to the source table and extract another related column:

e.g. in CLRC-MD we need Condition/Units, in CERA the relational model has to be explored to actually extract the required data.

in the Table SPATIAL_COVERAGE there is a field MIN_ALT_UNIT_ID which refers to UNIT_ID in the UNIT Table and there is another column in this table called UNIT_NAME which we actually want to map to Condition/Units

Say we were doing some query to find out the UNIT_NAME just from Condition/Units (e.g. as part of some conditional statement within a query – i.e. select x from y where z=(select a from b)).

We can see that the flow of information to the source is

TABLE->COLUMN->KEY_TABLE->KEY_COLUMN->REQUIRED_COLUMN

i.e.

```
SELECT  REQUIRED_COLUMN FROM KEY_TABLE, TABLE WHERE
TABLE.COLUMN = KEY_TABLE .KEY_COLUMN
```

in our case

```
SELECT  UNIT_NAME FROM UNIT, SPATIAL_COVERAGE WHERE
SPATIAL_COVERAGE.MIN_ALT_UNIT_ID = UNIT.UNIT_ID ;
```

So in our mapping we have added KEY_TABLE, KEY_COLUMN and REQUIRED_COLUMN to capture this translation in the mapping.

When reading the mapping if the KEY_TABLE is filled the KEY_COLUMN and REQUIRED_COLUMN must also be filled and it is the REQUIRED_COLUMN that decided which value should be mapped to/from the CLRC-MD Parent_Element/Element or Element/Attribute value.

3. THE MAPPING

num	Parent	Element	Occurrence Constraints	Attribute	Type	Notes about mapping	Table	Column	
	CLRCMetadata (root element)				complex				
	CLRCMetadata (root element)	MetadataRecord	0..n		complex				
	MetadataRecord			metadataID			ENTRY	ENTRY_ID	
	MetadataRecord	Topic	1		complex				
	Topic	Discipline	0..1		simple		TOPIC	TOPIC_NAME	
	Topic	Source	0..1		simple				
	Topic	Subject	1..n		simple		GENERAL_KEY	GENERAL_KEY	
	MetadataRecord	Programme		subStudies					
	MetadataRecord	Programme or <u>Experiment</u> or Measurement	1		complex	This was changed from Measurement to Experiment – as this would mean that that the existing XSL scripts should work fine			

		or Simulation							
	Programme	StudyName	1		simple				
	Programme	StudyID	0..n		complex				
	StudyID			Studyid		Changed to ENTRY->ENTRY_NAME as PROJECT->PROJECT_NAME now Measurement/StudyName	ENTRY	ENTRY_NAME	
	StudyID			Institutionref		could make this the same as Institution->InstitutionID perhaps – need to discuss with kevin – perhaps, later, chosen institute_name for now			
	Programme	Investigator	1..n		complex				
	Programme	StudyInformation	1		complex				
	Programme	Notes	0..1		simple				
	StudyID	Institution	0..1		simple	for the cera ‘inhouse’ contact cera ‘distributor’ has more in common with ‘DataManager’ in CLRC-MD	INSTITUTE	INSTITUTE_NAME	
	Institution			institutionID		same mapping for StudyID and DataManager	INSTITUTE	INSTITUTE_ID	
	Institution			institutiontype					
	Investigator	Name	1		complex				

	Investigator	Status	1		simple		PERSON	TITLE	
	Investigator	Institution	1		simple		INSTITUTE	INSTITUTE_NAME	
	Investigator	ContactDetails	1		complex				
	Investigator	Role	1		simple	may have to switch on the actual value to determine whether this person is the investigator or the datamanager (having them the same at the moment is that acceptable?)	CONTACT_TYPE	CONTACT_TYPE	
	Name	Surname	1		simple	Investigator->Name same for DataManager->Contact->Name	PERSON	LAST_NAME	
	Name	Initials	1		simple	need a conversion function here to extract the first letter	PERSON	SECOND_NAME	
	Name	Forename	0..n		simple	Investigator->Name same for DataManager->Contact->Name	PERSON	FIRST_NAME	
	Name	PersonTitle	0..1		simple				
	ContactDetails	Address	1		complex	same for Investigator->ContactDetails and DataManager->Contact-			

						>ContactDetails			
	ContactDetails	Telephone	1		simple		PERSON	TELEPHONE	
	ContactDetails	Email	0..1		simple		PERSON	EMAIL	
	ContactDetails	Fax	0..1		simple		PERSON	FAX	
	Address	Addressline1	1		simple		INSTITUTE	STREET	
	Address	Addressline2	0..1		simple				
	Address	Town	1		simple		INSTITUTE	PLACE	
	Address	Region	0..1		simple				
	Address	Postcode	0..1		simple		INSTITUTE	POBOX_POSTAL_CODE	
	Address	Country	1		simple		INSTITUTE	COUNTRY	
	Country			countryabbrev					
	StudyInformation	Funding	1		simple	Cera does not appear to have who funded the entry (which is one of many in a programme perhaps) in the actual entries			
	StudyInformation	TimePeriod	1		complex				
	StudyInformation	Purpose	1		complex				
	StudyInformation	StudyStatus	1		simple				
	StudyInformation	Resources	0..1		simple				
	TimePeriod	StartDate	1		complex				

	TimePeriod	EndDate	0..1		complex				
	StartDate	Date	1		simple	(for StudyInformation -> StartDate)	ENTRY	CREATION_DATE	
	StartDate	Time	0..1		simple				
	EndDate	Date	1		simple				
	EndDate	Time	0..1		simple				
	Purpose	Abstract	0..1		simple				
	Purpose	StudyType	0..1		simple				
	MetadataRecord	AccessConditions	1		simple		ACCESS_CONSTRAINT	CONSTRAINT_DESCR	
	MetadataRecord	DataHolding	0..1		complex				
	DataHolding			dataid			DATA_ACCESS	DATA_ACCESS_ID	
	DataHolding	DataName	1		simple	access_structure->access_structure_id is a candidate but seems more like a format			
	DataHolding	TypeOfData	0..1		simple	Does not seem to be specified but has a fixed format			
	DataHolding	Status	0..1		simple		PROGRESS	PROGRESS_DESC	
	DataHolding	LogicalDescription	0..1		simple, complex,				

					or mixed				
	DataHolding	FileFormat	0..1		simple		ACCESS_STRUCTURE	ACCESS_STRUCTURE_NAME	
	DataHolding	DataSet	0..n		complex				
	DataHolding	RelatedStudy	0..n		complex				
	DataSet			dataid		instance 1	STORAGE1	STORAGE_ID	
	DataSet			dataid		instance 2	STORAGE2	STORAGE_ID	
	DataSet			dataid		instance 3	STORAGE3	STORAGE_ID	
	DataSet			dataid		instance 4	STORAGE4	STORAGE_ID	
	DataSet`			dataid		instance 5 – rasdaman	RAS_CONNECT	ENTRY_ID	
	DataSet	DataName	1		simple	instance 1 storage1,storage2,storage3 and storage4 are synonyms for storage	STORAGE	STORAGE_NAME	
	DataSet	DataName	1		simple	instance 5 need for connection to rasdaman – value is ras_collection as this is the name of the data is	(RAS_CONNECT)	(RAS_COLLECTION)	
	DataSet	DataName	1		simple	instance 2	STORAGE	STORAGE_NAME	
	DataSet	DataName	1		simple	instance 3	STORAGE	STORAGE_NAME	
	DataSet	DataName	1		simple	instance 4	STORAGE	STORAGE_NAME	

								ME	
	DataSet	Status	0..1		simple	instance 1	PROGRESS	PROGRESS_DE SCR	
	DataSet	Status	0..1		simple	instance 2	PROGRESS	PROGRESS_DE SCR	
	DataSet	Status	0..1		simple	instance 3	PROGRESS	PROGRESS_DE SCR	
	DataSet	Status	0..1		simple	instance 4	PROGRESS	PROGRESS_DE SCR	
	DataSet	Status	0..1		simple	instance 5 - rasdaman	PROGRESS	PGROGRESS_D ESC	
	DataSet	TypeOfData	0..1		simple				
	DataSet	LogicalDescription	0..1		simple, complex or mixed				
	DataSet	FileFormat	0..1		simple				
	DataSet	File	0..n		complex				
	DataSet	RelatedDataSetRef	0..n		complex				
	DataSet	ParentDataSetRef	0..1		simple				
	DataSet	ChildDataSetRef	0..n		simple				
	DataSet	DerivedFromDataSetRef	0..n		complex				

	File			dataid					
	File	DataName	1		simple				
	File	TypeOfData	0..1		simple				
	File	Status	0..1		simple				
	File	LogicalDescription	0..1		simple, complex or mixed				
	File	FileFormat	0..1		simple				
	File	URI	1		simple				
	File	RelatedFileRef	0..n		complex				
	File	DerivedFromFileRef	0..n		complex				
	RelatedFileRef	FileRef	1		simple				
	RelatedFileRef	RelationType	1		simple				
	DerivedFromFileRef	DataHoldingRef	0..n		simple				
	DerivedFromFileRef	DataSetRef	0..n		simple				
	DerivedFromFileRef	FileRef	1		simple				
	RelatedDataSetRef	DataSetRef	1		simple				
	RelatedDataSetRef	RelationType	1		simple				

	DerivedFromDataSetRef	DataHoldingRef	0..1		simple				
	DerivedFromDataSetRef	DataSetRef	1		simple				
	RelatedStudy	StudyRef	1		empty				
	RelatedStudy	RelationType	1		simple				
	StudyRef		studyref						
	MetadataRecord	DataLocation	0..1		complex				
	DataLocation	DataHoldingLocator	1		complex				
	DataLocation	DataSetLocator	0..n		complex				
	DataLocation	FileLocator	0..n		complex				
	DataHoldingLocator			dataidref		for STORAGE1,2,3 and 4	STORAGE	STORAGE_ID	
	DataHoldingLocator			dataidref		for rasdaman	RAS_CONNECT	ENTRY_ID	
	DataHoldingLocator	DataName	1		simple	this is also (usually) a location	STORAGE	STORAGE_NAME	
	DataHoldingLocator	DataName	1		simple	for rasdaman	(RAS_CONNECT)	(RAS_COLLECTION)	
	DataHoldingLocator	Locator	0..n		simple				
	Locator			type					
	Locator	URL	1		simple		STORAGE	STORAGE_NAME	
	Locator	URL	1		simple	for rasdaman – perhaps some location information needs to be here to state the actual location of the data	RAS_CONNECT	RAS_COLLECTION	

						source			
	Locator	DataSourceType	0..1		simple				
	Locator	DataSourceAccess	0..1		simple				
	DataSetLocator			type					
	DataSetLocator	DataName	1		simple				
	DataSetLocator	Locator	0..n		complex				
	FileLocator			dataidref					
	FileLocator	URI	1		simple				
	FileLocator	Locator	0..n		complex				
	MetadataRecord	RelatedMaterial	0..1		simple, complex or mixed				
	Experiment			associated					
	Experiment	StudyName	1		simple				
	Experiment	StudyID	0..n		complex				
	Experiment	Investigator	1..n		complex				
	Experiment	StudyInformation	1		complex				

	Experiment	Notes	0..1		simple				
	Experiment	DataManager	1		complex				
	Experiment	Instrument	1		simple				
	Experiment	Condition	0..n		complex				
	Experiment	Parameter	0..n		complex				
	Measurement			associated					
	Measurement	StudyName	1		simple	note this has to be unique in the UI as this is selected from – due to the way the data in the cera database has been added (i.e. one grib to many entries) to make sense of the situation we have chosen what we have to make it unique yet meaning full	TOPIC ENTRY	ENTRY_NAME. TOPIC_NAME." CERA"ENTRY_ ID	
	Measurement	StudyID	0..n		complex				
	Measurement	Investigator	1..n		complex				
	Measurement	StudyInformation	1		complex				
	Measurement	Notes	0..1		simple				
	Measurement	DataManager	1		complex				
	Measurement	Instrument	1		simple				
	Measurement	Condition	0..n		complex				

	Measurement	Parameter	0..n		complex				
	Simulation			associated					
	Simulation	StudyName	1		simple				
	Simulation	StudyID	0..n		complex				
	Simulation	Investigator	1..n		complex				
	Simulation	StudyInformation	1		complex				
	Simulation	Notes	0..1		simple				
	Simulation	DataManager	1		complex				
	Simulation	Machine	1		simple				
	Simulation	Program	1		simple				
	Simulation	Parameter	0..n		complex				
	Simulation	Data	0..1		simple				
	Condition	ParamName	1		simple	instance 1 – Name of Col could have chosen (LOCATION table also – perhaps)	(SPATIAL_COVE RAGE)	(MIN_LAT)	
	Condition	ParamName	1		simple	instance 2 – Name of Col (LOCATION table also – perhaps)	(SPATIAL_COVE RAGE)	(MAX_LAT)	

	Condition	ParamName	1		simple	instance 3 – Name of Col (LOCATION table also – perhaps)	(SPATIAL_COVERAGE)	(MIN_LON)	
	Condition	ParamName	1		simple	instance 4 – Name of Col (LOCATION table also – perhaps)	(SPATIAL_COVERAGE)	(MAX_LON)	
	Condition	ParamName	1		simple	instance 5 – Name of Col	(SPATIAL_COVERAGE)	(MIN_ALTITUDE)	
	Condition	ParamName	1		simple	instance 6 – Name of Col	(SPATIAL_COVERAGE)	(MAX_ALTITUDE)	
	Condition	ParamName	1		simple	instance 7 – Name of Col	(TEMPORAL_COVERAGE)	(START_YEAR)	
	Condition	ParamName	1		simple	instance 8 – Name of Col	(TEMPORAL_COVERAGE)	(START_MONTH)	
	Condition	ParamName	1		simple	instance 9 – Name of Col	(TEMPORAL_COVERAGE)	(START_DAY)	
	Condition	ParamName	1		simple	instance 10 – Name of Col	(TEMPORAL_COVERAGE)	(STOP_YEAR)	
	Condition	ParamName	1		simple	instance 11 – Name of Col	(TEMPORAL_COVERAGE)	(STOP_MONTH)	
	Condition	ParamName	1		simple	instance 12 – Name of Col	(TEMPORAL_COVERAGE)	(STOP_DAY)	
	Condition	ParamName	1		simple	instance 13 – Name of table in this case as we will put in a complete data time from that table with a timezone	(MOMENT)	(N/A)	

	Conditon	ParamName	1		simple	instace 14 – Name of Col	(LOCATION)	(LOCATION_N AME)	
	Condition	Units	0..1		simple	instance 1 – implicit or perhaps the could come from UNIT->UNIT_NAME depending on whether these were setup correctly There would have to be many entries in the PARAMETER table for this particular ENTRY_ID for different UNI_ID's			
	Condition	Units	0..1		simple	instance 2 – implicit			
	Condition	Units	0..1		simple	instance 3 – implicit			
	Condition	Units	0..1		simple	instance 4 – implicit			
	Condition	Units	0..1		simple	instance 5 – implicit Note this requires a lookup on PARAMETER->UNIT_ID to reference the UNIT name	SPATIAL_COVER AGE	MIN_ALT_UNI T_ID	UNI
	Condition	Units	0..1		simple	instance 6 – implicit	SPATIAL_COVER AGE	MAX_ALT_UNI T_ID	UNI
	Condition	Units	0..1		simple	instance 7 – implicit			
	Condition	Units	0..1		simple	instance 8 – implicit			
	Condition	Units	0..1		simple	instance 9 – implicit			

	Condition	Units	0..1		simple	instance 10 – implicit			
	Condition	Units	0..1		simple	instance 11 – implicit			
	Condition	Units	0..1		simple	instance 12 – implicit			
	Condition	Units	0..1		simple	instance 13 – implicit			
	Condition	ParamValue	0..1		simple	instance 1	SPACIAL_COVE RAGE	MIN_LAT	
	Condition	ParamValue	0..1		simple	instance 2	SPACIAL_COVE RAGE	MAX_LAT	
	Condition	ParamValue	0..1		simple	instance 3	SPACIAL_COVE RAGE	MIN_LON	
	Condition	ParamValue	0..1		simple	instance 4	SPACIAL_COVE RAGE	MAX_LON	
	Condition	ParamValue	0..1		simple	instance 5	SPACIAL_COVE RAGE	MIN_ALTITUD E	
	Condition	ParamValue	0..1		simple	instance 6	SPACIAL_COVE RAGE	MAX_ALTITUD E	
	Condition	ParamValue	0..1		simple	instance 7	TEMPORAL_CO VERAGE	START_YEAR	
	Condition	ParamValue	0..1		simple	instance 8	TEMPORAL_CO VERAGE	START_MONT H	
	Condition	ParamValue	0..1		simple	instance 9	TEMPORAL_CO VERAGE	START_DAY	
	Condition	ParamValue	0..1		simple	instance 10	TEMPORAL_CO VERAGE	STOP_YEAR	

	Condition	ParamValue	0..1		simple	instance 11	TEMPORAL_CO VERAGE	STOP_MONTH	
	Condition	ParamValue	0..1		simple	instance 12	TEMPORAL_CO VERAGE	STOP_DAY	
	Condition	ParamValue	0..1		simple	instance 13 when concatonating the UTC_DIFFERENCE a + or – and left zero padding needs adding	MOMENT	YEAR MONT H DAY HO UR MINUTE SECOND UT C_DIFFERENC E	
	Conditon	ParamValue	0..1		simple	instance 14	LOCATION	LOCATION_N AME	
	Condition	Range	0..1		complex				
	Parameter	ParamName	1		simple				
	Parameter	Units	0..1		simple				
	Parameter	ParamValue	0..1		simple				
	Parameter	Range	0..1		complex				
	Range	Limit	1..n		simple				
	Limit			bound					
	RelatedMaterial	Publications	0..n		simple		CITATION	TITLE	

	RelatedMaterial	References	0..n		simple				
	RelatedMaterial	RelatedInvestigations	0..n		simple				
	RelatedMaterial	CommunityInformation	0..n		simple				
	DataManager	Institution	1		simple (+attribute)	CONTACT_TYPE- >CONTACT_TYPE in cera needs to be 'distributor'	INSTITUTE	INSTITUTE_NAME	
	DataManager	Contact	0..n		complex				
	Contact	Name	1		complex				
	Contact	Status	1		simple				
	Contact	Institution	1		simple (+attribute)				
	Contact	ContactDetails	1		complex				

