CsvCollectionResource

+ resource_type[1]: String{"csv-observations-collection"}
+ schema_descriptor_version[1]: Decimal{0.3}

1..*   + members

CsvResourceMember

+ resource_name[1]: String[]
+ resource_type[1]:
String{observed_geometries|observations |
observations_with_geometry}
+ header_rows[0..1]: Integer
+ encoding[0..1]: String
+ crs[0..1]: EpsgCode

1..*   + fields

CsvField

+ field_id[1]: String
+ short_name[0..1]: String
+ long_name[0..1]: String
+ description[0..1]: String
+ field_type[1]: DataTypeCode
+ field_role[0..1]: FieldRoleCode
+ uom[0..1]: UomCode
+ phenomenon[0..1]: String
+ date_format[0..1]: IsoDateTimeFormat
+ time_zone[0..1]: TimeZoneString
+ no_data[0..1]: String|Number|null

resource_name: An unorderd list of CSV Resources (files) that share the following CsvField definition.

resource_type: „platforms" is used for resources/files that contain location data, such as station definitions. „observation" is used for resources/files that contain measurements, but not locations.

TODO: define a resource type for resources/files that contain both. Maybe „platforms" is sufficient in this case?

header_rows: Optional. Defaults to „1" if ommitted. Number of header rows in CSV resource/file. Header rows contain column names and metadata.

encoding: Optinal. Defaults to UTF8 if ommitted. Character set of the CSV resources/files. Values shall be valid CharSet codes.

EpsgCode: conditional use if the CSV resource/file contains geographic coordinates. Defaults to EPSG:4326 if not specified.

FieldRoleCode: conditional; use if field is not a phenomenon.

UomCode: conditional use if the field has a unit. If possible, UCUM unit String should be used here.

phenomenon: No policy yet. Use UCUM „kind of quantity" term?

IsoDateTimeFormat: conditional; use if the field is a date/time String. Must comply with standard ISO date/time format:
Y – year; M – month; d – day; H – hour; m – minute; s – seconds and milliseconds
Examples:YYYYMMdd; YYYY-MM-dd; YYYY-MM-dd:HH:mm:ss.sss

TimeZoneString: conditional use if the field is a date/time String. Defaults to „GMT" if not specified. Must comply with standard ISO time zone format.

no_data: Optional NoData value. Use appropriate encoding according to **field_type**.

TODO: Check with GDAL's CSV descriptor for future versions. http://www.gdal.org/drv_csv.html

TODO: Current geo-coordinate definitiononly works with lat/lon CRSes.

Conceptually, all **CsvResourceMember** constitute a single big table. For practical reasons (storage efficiency, thematic grouping, data splitting) this tabular data can be spread across multiple resources/files.

Interpreters of the data should concatenate / join fields with the same field_id.

<<enumeration>>
DataTypeCode

Boolean
Integer
Float
Decimal
Date
String
Geometry (WKT!)

~~Latitude~~
~~Longitude~~
~~Height~~

Extensible List

TODO: Check the use of *Float*. *Real* might be less ambiguous.

General Question: Do we need a distinction between *Float/Double* (encoding precision errors) and *Decimal*? Or is a *Real* type sufficient?

<<enumeration>>
FieldRoleCode

latitude
longitude
height
timestamp
valid_time_start
valid_time_end

Extensible List