



## ORCA-Registry v1.0 Implementation Notes

### *Document History*

James Blanden	23 October 2007	Version 1.0	Initial document.
James Blanden	14 November 2007	Version 1.1	Restructured.

### *Table of Contents*

<b>Background .....</b>	<b>2</b>
<b>Document Scope .....</b>	<b>2</b>
<b>ISO 2146 Inconsistencies.....</b>	<b>3</b>
<b>ORCA-Registry Data Interchange Schema.....</b>	<b>7</b>
<b>Recommendations .....</b>	<b>8</b>
<b>Dublin Core Collections Application Profile Field Mappings.....</b>	<b>9</b>

## ***Background***

The Online Research Collections Australia (ORCA)-Registry software is a product of the Australian Partnership for Sustainable Repositories (APSR) Collection Services and Infrastructures (COSI) initiatives of 2007. The ORCA-Registry is a PHP/PostgreSQL web application designed to be housed within an instance of the COSI-Framework. More information about ORCA can be found at the APSR website <http://www.apsr.edu.au/orca/index.htm>.

## ***Document Scope***

This document is a loosely structured collection of notes that I made during the development of the ORCA-Registry software. It doesn't cover any functional or technical aspects of the software in any depth. Refer to the COSI-Framework and ORCA-Registry documentation for more information.

The registry was required to be based closely on the ISO 2146 draft standard, and in fact was the first known implementation of that standard. These notes then, also cover implementation and profiling of ISO 2146 for the purposes of building the ORCA-Registry.

## **ISO 2146 Inconsistencies**

When I began translating the XML schema provided in the ISO 2146 documentation into a relational database structure, I noticed inconsistencies between the schema and the descriptive documentation. These are described below, along with the assumptions that I made to continue development.

Documents are referenced from <http://www.nla.gov.au/wgroups/ISO2146/>.

ISO 2146 - Information and Documentation - Registry Services for Libraries and Related Organisations (ISO TC46 SC4 Working Draft, 13 December 2005). Word document. Available: <http://www.nla.gov.au/wgroups/ISO2146/n197.doc>.

Referred to here as N197.

ISO 2146 - Information and Documentation - Registry Services for Libraries and Related Organisations (ISO TC46 SC4 Working Draft, 13 December 2005). XML Schema. Available: <http://www.nla.gov.au/wgroups/ISO2146/n198.xsd>.

Referred to here as N198.

1. *N197 Page 9*, 1 Registry Object, 8.1 Date Range, 0-1 Occurrences.  
*N198 Line 392*, dateRange element has maxOccurs="unbounded".  
Assume error in *N197*.
2. *N197 Page 9*, 1 Registry Object, 8.7 Notes, 0-N Occurrences.  
*N198 Line 396*, notes element has no maxOccurs.  
Assume error in *N197*.
3. *N197 Page 13*, 1.4 Relation references to:  
1.4.2 Relation Type  
1.4.1 Relation Category  
1.4.2 Related Registry Object  
Should be:  
1.4.1 Relation Type  
1.4.2 Relation Category  
1.4.3 Related Registry Object
4. *N197 Page 17*, 1.8 Description is used by 1.10 Event, 3.3 Coverage, 5.3 Product and 6.3 Obligation as well as 1 Registry Object.
5. *N197 Page 19*, 1.10 Event reference to 1.7 Description should be to 1.8 Description.
6. *N197 Page 28*, 5.3 Product reference to 1.7 Description should be to 1.8 Description.
7. *N197 Page 34*, 6.3 Obligation reference to 1.7 Description should be to 1.8 Description.
8. *N197 Page 18*, 1.9 Info Pointer is used by 9.2 Update Event as well as 1 Registry Object.
9. *N197 Page 20*, 2.2 Group, 2.2.1 Group Type, 1-N Occurrences.  
*N198 Line 43*, groupType element has no maxOccurs.  
Assume error in *N197*.
10. *N197 Page 22*, 3.2 Collection Profile reference to 8.10 Resource Type.  
8.10 is Status. Resource Type is not described.

11. *N197 Page 47*, 8.9 Measurement Type states used by Collection Profile and Increment. 6.3.8 Increment does not reference 8.9 Measurement Type.
12. *N297 Page 24*, 3.3 Coverage references to:
  - 3.4.1 Coverage Type
  - 3.4.2 Coverage Area
  - 1.7 Description
  - 3.4.3 Current Strength
  - 3.4.4 Current Strength Notes
  - 3.4.5 Planned Strength
  - 3.4.6 Planned Strength NotesShould be:
  - 3.3.1 Coverage Type
  - 3.3.2 Coverage Area
  - 1.8 Description
  - 3.3.3 Current Strength
  - 3.3.4 Current Strength Notes
  - 3.3.5 Planned Strength
  - 3.3.6 Planned Strength Notes
13. *N197 Page 26*, 4 Activity, 4.1 Activity Type, 1-N Occurrences.  
*N198 Line 91*, activityType element has no maxOccurs.  
Assume error in *N197*.
14. *N197 Page 27*, 5 Service, 5.1 Service Type, 1-N Occurrences.  
*N198 Line 105*, serviceType element has no maxOccurs.  
Assume error in *N197*.
15. *N197 Page 27*, 5 Service reference to 5.9 Policy Combining Algorithm should be to 5.6 Policy Combining Algorithm.
16. *N197 Page 27*, 5.1 Service Type states used by Activity. Activity does not reference Service Type. Should be used by Service.
17. *N197 Page 27*, 5.2 Function states used by Service, Activity, Service Policy. Should be used by Service, Product, Access Policy, and Obligation.
18. *N197 Page 28*, 5.3 Product states used by Service. Should be used by Service and Obligation.
19. *N107 Page 28*, 5.3 Product reference to 1.7 Description. 1.7 is Subject, should be reference to 1.8 Description.
20. *N197 Page 28*, 5.3 Product, 1.8 Description, 0-1 Occurrences.  
*N198 Line 113*, description element has maxOccurs="unbounded".  
Assume error in *N197*.
21. *N197 Page 28*, 5.3 Product, 5.3.3 Input Description, 0-N Occurrences.  
*N198 Line 114*, inputDescription element has no maxOccurs.  
Assume error in *N197*.
22. *N197 Page 28*, 5.3 Product, 5.3.4 Output Description, 0-N Occurrences.  
*N198 Line 115*, inputDescription element has no maxOccurs.  
Assume error in *N197*.

23. *N197 Page 29*, 5.4.1 Service Times Key states used by Service Times. Should be used by Service Times and Access Policy.
24. *N197 Page 29*, 5.4 Service Times references to:
  - 8.13 Description
  - 5.4.3 Date Range
  - 5.4.4 SuspensionShould be:
  - 1.8 Description
  - 8.1 Date Range
  - 5.4.3 Suspension
25. *N197 Page 30*, 5.4.2 Service Hours references to:
  - 5.5.2.1 Day of Week
  - 5.5.2.2 Time Valid From
  - 5.5.2.3 Time Valid ToShould be:
  - 5.4.2.1 Day of Week
  - 5.4.2.2 Time Valid From
  - 5.4.2.3 Time Valid To
26. *N197 Page 31*, 5.5 Access Policy reference to 8.10 Resource Type. 8.10 is Status. Resource Type is not described.
27. *N197 Page 31*, 5.5 Access Policy reference to 6.5.1 Service Description Key should be to 6.1 Service Description Key.
28. *N197 Page 33*, 6.1 Service Description Key states used by Service. Should be used by Service and Access Policy.
29. *N197 Page 34*, 6.3 Obligation references to:
  - 5.3 Function
  - 5.4.1 Product Key
  - 1.7 Description
  - 1.6.5 Contact RoleShould be:
  - 5.2 Function
  - 5.3.1 Product Key
  - 1.8 Description
  - 1.5.1 Contact Role
30. *N197 Page 34* 6.3 Obligation, 1.5.1 Contact Role has occurrences of O-1, should be 0-1.
31. *N197 Page 35*, 6.3.2.2 Utility is data type enumeration. *N198 Line 479*, element utility is of type xs:string. Assume error in *N197*.
32. *N197 Page 37*, 6.3.5 Payment Method reference to 6.3.5.1 Request Method Type should be to 6.3.5.1 Payment Method Type.
33. *N197 Page 38*, 6.3.8 Increment references to:
  - 6.3.8.3 Minimum Quantity
  - 6.3.8.4 Maximum QuantityShould be:

### 6.3.8.3 Maximum Quantity

### 6.3.8.4 Minimum Quantity

34. *N197 Page 47*, 8.10 Status, 8.10.1 Status Value is of type Enumeration/Boolean.  
*N198 Line 472*, element statusValue is of type xs:string.  
Assume error in *N197*.
35. *N197 Page 47*, 8.10 Status, 8.10.3 Date Assigned.  
*N198 Line 474*, element status has element dateRange, not dateAssigned.  
Assume error in *N197*.
36. *N197 Page 48*, 9 Record Details, 8.10 Status mandatory with occurrences 1-N.  
*N198 Line 421*, element ref status has minOccurs="0" and no maxOccurs.  
Assume error in *N197*.
37. *N197 Page 48*, 9.2 Update Event, 1.9 Info Pointer should have Obl O and not 0.

## **ORCA-Registry Data Interchange Schema**

Some changes to the ISO 2146 XML schema were required to support the requirements of the registry software. The resulting XML schema effectively forms a profile of ISO 2146, and is referred to as the *ORCA-Registry Data Interchange Schema*. Some of these changes are listed below:

- Add the top level container element *registryObjects*.
- Made 1.3 Name 1-N instead of 0-N. This makes at least one *name* element mandatory. Objects within the registry are most usefully listed by name in the search and browse results.
- Removed 9.1 Record Identifier and 9.3 Record Relation from 9 Record Details, as these seem to duplicate structures that perform these roles at 1.1 Registry Object Key and 1.4 Relation.
- Removed 9 Record Details from the interchange schema altogether. This is internal administrative metadata that will not be provided or needed for interchange. The structures will remain in the database and modified by the system as required. (Or may be redone.)
- Removed 1.6.1 Address Key. 1.6 Address is contained in a registry object, and there is no way to reference an address that exists externally to that object. A key is required to support the one to many relationship between a registry object and its addresses in the relational database, but this is not required in the XML.
- Made *personType* optional.

In addition to creating a profile of the schema for the registry, there were issues relating to gathering, display, and maintenance of the data:

- Store the *Data Source key* with the registry object. All objects in the registry are gathered from a *Data Source*, and it is required that the registry be able to identify which objects were gathered from which source.
- How to combine data that is broken up (and possibly has multiple language strings) for display in the search results/list view. Resolved by using delimiters (mdash and middot), along with styles, and just displaying all of the data, one after the other.

## Recommendations

The development process has revealed issues with the ISO 2146 draft standard. These comprise issues with the structure described, how it is described, and the lack of guidelines for developers and those wishing to expose their data. A brief list of recommendations (in no particular order) is provided below:

- Administrative metadata should be removed from the specification and left up to implementers (particularly as there are no guidelines provided for how an implementing system should interpret this information). This includes *status*, *notes*, *recordDetails*, and most of the *events* and *date range* structures. Also note that *status* information is transitively dependant on *date range* information, and therefore could be in conflict with it (for example, the structures don't prevent a record from having a *date valid to* that has expired, and at the same time a *status* of 'valid').
- The *date range* structures allow overlapping and therefore conflicting data to be defined. Most of these are used for administrative means and should be removed altogether, but where they might be used (for example to provide provenance information for a collection) they should take the form of *date from* effective dating.
- A *date modified* attribute should be included at the top level of each *registryObject* to support services that need to filter records on this basis (for example RSS feeds and OAI-PMH). A strict, clear, and agreed definition of what *date modified* means in this context will need to be determined.
- An investigation into the *language string* structures, and clarification of which requirements for various levels of internationalisation they are meeting. Character set support, complete locale based translation, and partial translation or alternative language presentation, all make different requirements of the implementing system.
- An overall simplification of the model, and a redo of the XML schema (making use of element attributes where appropriate). A definition of, and focus on, the core use cases would assist implementors—both those building registries, and those wishing to supply data to them. At present the structure is perhaps too flexible, and there is ambiguity about how data should be described. (For example, should a person be described as the contact for a collection, or described separately as a party, and related to the collection?)



## Dublin Core Collections Application Profile Field Mappings

At one stage it seemed that it might be useful to investigate some kind of cross walking of the Dublin Core Collections Application Profile to the structures in ISO 2146, and then to the structures in the database built for the ORCA-Registry. This rough mapping is shown in the table below. (Note that there are simple types in the DC Collections that only have complex equivalents in ISO 2146, and so there is no clear or simple translation.)

Dublin Core Collections Application Profile	ISO 2146	db_s_orca
Type [dc:type]	3.1 Collection Type [complexType]	tbl_collections.collection_type
Collection Identifier [dc:identifier]	1.2 Identifier [complexType]	tbl_registry_object_identifiers
Title [dc:title]	1.3 Name [complexType]	tbl_registry_object_names
Alternative Title [dcterms:alternative]	1.3 Name [complexType]	tbl_registry_object_names
Description [dcterms:abstract]	1.8 Description [complexType]	tbl_registry_object_descriptions
Size [dcterms:extent]	3.2.2 Number of Units [positiveInteger]	tbl_collection_profiles.number_of_units
Language [dc:language]	8.5 Language	tbl_registry_object_languages
Item Type [cld:itemType]	?.? Resource Type	tbl_collection_profiles.resource_type
Item Format [cld:itemFormat]	NO MAPPING	NO MAPPING
Rights [dc:rights]	NO MAPPING	NO MAPPING
Access Rights [dcterms:accessRights]	5.5 Access Policy [complexType]	tbl_access_policies
Accrual Method [dcterms:accrualMethod]	3.2.7 Accrual Method [complexType]	tbl_collection_profiles.accrual_method
Accrual Periodicity [dcterms:accrualPeriodicity]	3.2.6 Accrual Periodicity [complexType]	tbl_collection_profiles.accrual_periodicity
Accrual Policy [dcterms:accrualPolicy]	3.2.6 Accrual Policy [complexType]	tbl_collection_profiles.accrual_policy
Custodial History [dcterms:provenance]	NO MAPPING	NO MAPPING
Audience [dcterms:audience]	NO MAPPING	NO MAPPING
Subject [dc:subject]	1.7 Subject [complexType]	tbl_registry_object_subjects
Spatial Coverage [dcterms:spatial]	3.3 Coverage [complexType]	tbl_coverage
Temporal Coverage [dcterms:temporal]	3.3 Coverage [complexType]	tbl_coverage
Date Collection Accumulated [dcterms:created]	1.10 Event [complexType]	tbl_events
Date Items Created [cld:dateItemsCreated]	1.10 Event [complexType]	tbl_events
Collector [dc:creator]	1.4 Relation [complexType] to a Party	tbl_registry_object_relations
Owner [marcrel:OWN]	1.4 Relation [complexType] to a Party	tbl_registry_object_relations
Is Located At [cld:isLocatedAt]	8.6 Location [complexType]	tbl_registry_object_locations
Is Accessed Via [cld:isAccessedVia]	1.4 Relation [complexType] to a Service	tbl_registry_object_relations
Sub-Collection [dcterms:hasPart]	1.4 Relation [complexType] to a Collection	tbl_registry_object_relations
Super-Collection [dcterms:isPartOf]	1.4 Relation [complexType] to a Collection	tbl_registry_object_relations
Catalogue or Index [cld:catalogueOrIndex]	NO MAPPING	NO MAPPING
Associated Collection [cld:associatedCollection]	1.4 Relation [complexType] to a Collection	tbl_registry_object_relations
Associated Publication [dcterms:isReferencedBy]	NO MAPPING	NO MAPPING