

IDREFs Explained

Version: **1.0 Final**

Date: **04/04/2011**

Distribution: **Industry**

DISCLAIMER

Origo Services Limited believes it has employed personnel using reasonable skill and care in the creation of this document. However, this document is provided to the reader 'as is' without any warranty (express or implied) as to accuracy or completeness and Origo Services Limited cannot be held liable for any errors or omissions in this document, nor for any losses, damages or expenses arising consequent to the use of this document by the reader.

CHANGE HISTORY

Date	Version	Change(s) Incorporated
01/04/2011	1.0 Draft A	Initial version.
04/04/2011	1.0 Final	Include Schema text as Appendices; Add disclaimer note to Introduction; Amend capitalisation in schemas; Amend 'Important note' text; Keep Summary formal.

TABLE OF CONTENTS

1. INTRODUCTION 3

2. A SIMPLE DATA MODEL..... 4

3. SCHEMA 1 4

4. SCHEMA 2 9

5. SUMMARY 13

6. APPENDIX A – SCHEMA 1 LISTING..... 14

7. APPENDIX B – SCHEMA 2 LISTING..... 15

1. INTRODUCTION

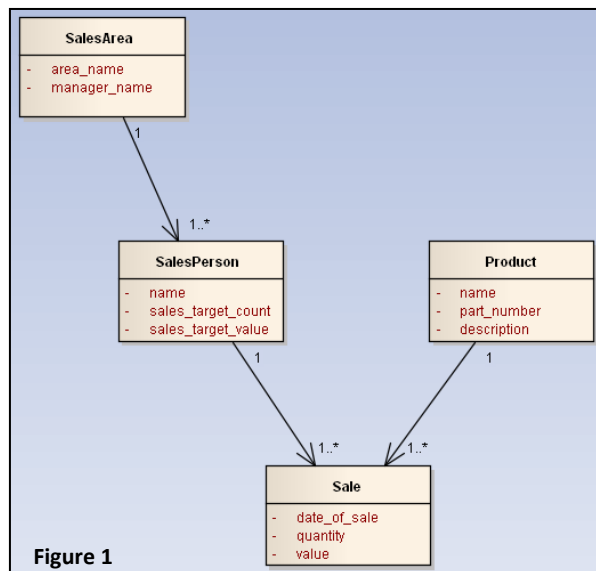
ID/IDREFs are used frequently throughout Origo Standards. This document aims to explain the purpose and use of this construct.

The explanation is made by way of simple examples. Readers need a basic knowledge of Data Models, XML and XML Schema.

Note: This document is a description of IDREFs. The schemas described within it are for illustrative purposes only and should not be regarded as 'best practice', the basis of any current or future Origo Standard, or an indication of how Origo might design standards in the future.

2. A Simple Data Model

For this document we will use the data model in **Figure 1** below as a basis for the explanation. The model describes a simple sales situation where a sales person works within a sales area and sells products. Each product sale is recorded separately. (Note, to keep the model simple we don't know, or care, who the product was sold to.)



For the purposes of this explanation we will assume there are two sales areas, Area North and Area South, managed by different people. In Area North there are three sales persons: Albert Allis, Bertram Barnard and Cerys Cowan. In Area South there are two sales people: Dawn Dove and Edward Evans. The sales team is responsible for selling two different products, the Super X4 and the Dyno V3. A single sale may be for more than one item, but only one product type may be sold in a single sale.

Albert has made three Super X4 sales and one Dyno V4 sale, Bertram has made one Super X4 sale and one Dyno V4 sale, Cerys has made one Super X4 sale, Dawn has made two Super X4 sales and one Dyno V3 sale and Edward has made two Super X4 sales and two Dyno V3 sales.

3. Schema 1

Suppose there is a requirement to produce an XML message organised by Sales Area, Sales Person and Sale showing which Products have been sold. A simple schema (listing in Appendix 1) to validate such a message could look something like the one in **Figure 2** below.

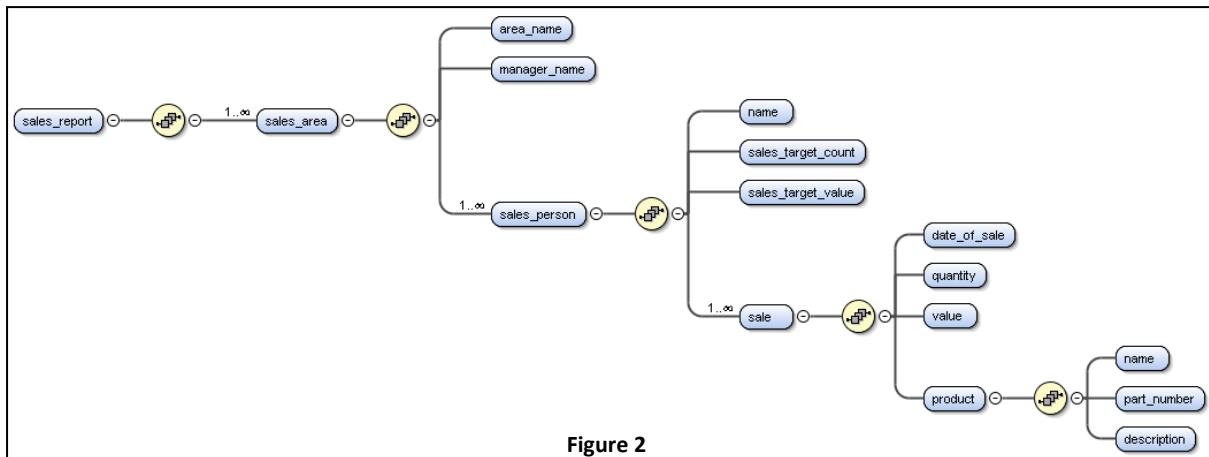


Figure 2

Using this schema, the corresponding XML message for the sales described above would be as shown in **Listing 1** below.

```
<?xml version="1.0" encoding="UTF-8"?>
<sales_report>
  <sales_area>
    <area_name>Area North</area_name>
    <manager_name>Norman Noaks</manager_name>
    <sales_person>
      <name>Albert Allis</name>
      <sales_target_count>24</sales_target_count>
      <sales_target_value>3400.00</sales_target_value>
      <sale>
        <date_of_sale>01/03/2011</date_of_sale>
        <quantity>4</quantity>
        <value>560</value>
        <product>
          <name>Super X4</name>
          <part_number>SX9467675893</part_number>
          <description>Super X4 in green and red</description>
        </product>
      </sale>
      <sale>
        <date_of_sale>14/03/2011</date_of_sale>
        <quantity>2</quantity>
        <value>300</value>
        <product>
          <name>Super X4</name>
          <part_number>SX9467675893</part_number>
          <description>Super X4 in green and red</description>
        </product>
      </sale>
      <sale>
        <date_of_sale>20/03/2011</date_of_sale>
```

```
<quantity>1</quantity>
<value>400</value>
  <product>
    <name>Dyno V3</name>
    <part_number>DNV3-48953-FGGHEJ</part_number>
    <description>Dyno V3 with large cover</description>
  </product>
</sale>
<sale>
  <date_of_sale>14/03/2011</date_of_sale>
  <quantity>2</quantity>
  <value>350</value>
  <product>
    <name>Super X4</name>
    <part_number>SX9467675893</part_number>
    <description>Super X4 in green and red</description>
  </product>
</sale>
</sales_person>
<sales_person>
  <name>Bertram Barnard</name>
  <sales_target_count>8</sales_target_count>
  <sales_target_value>2100.00</sales_target_value>
  <sale>
    <date_of_sale>06/03/2011</date_of_sale>
    <quantity>3</quantity>
    <value>100</value>
    <product>
      <name>Dyno V3</name>
      <part_number>DNV3-48953-FGGHEJ</part_number>
      <description>Dyno V3 with large cover</description>
    </product>
  </sale>
  <sale>
    <date_of_sale>10/03/2011</date_of_sale>
    <quantity>4</quantity>
    <value>650</value>
    <product>
      <name>Super X4</name>
      <part_number>SX9467675893</part_number>
      <description>Super X4 in green and red</description>
    </product>
  </sale>
</sales_person>
<sales_person>
  <name>Cerys Cowan</name>
  <sales_target_count>14</sales_target_count>
  <sales_target_value>1900.00</sales_target_value>
  <sale>
    <date_of_sale>12/03/2011</date_of_sale>
    <quantity>2</quantity>
    <value>380</value>
    <product>
      <name>Super X4</name>
      <part_number>SX9467675893</part_number>
      <description>Super X4 in green and red</description>
    </product>
  </sale>
</sales_person>
```

```
</sales_person>
</sales_area>
<sales_area>
  <area_name>Area South</area_name>
  <manager_name>Susan Smith</manager_name>
  <sales_person>
    <name>Dawn Dove</name>
    <sales_target_count>14</sales_target_count>
    <sales_target_value>1900.00</sales_target_value>
    <sale>
      <date_of_sale>12/03/2011</date_of_sale>
      <quantity>2</quantity>
      <value>380</value>
      <product>
        <name>Super X4</name>
        <part_number>SX9467675893</part_number>
        <description>Super X4 in green and red</description>
      </product>
    </sale>
    <sale>
      <date_of_sale>06/03/2011</date_of_sale>
      <quantity>3</quantity>
      <value>100</value>
      <product>
        <name>Dyno V3</name>
        <part_number>DNV3-48953-FGGHEJ</part_number>
        <description>Dyno V3 with large cover</description>
      </product>
    </sale>
    <sale>
      <date_of_sale>10/03/2011</date_of_sale>
      <quantity>4</quantity>
      <value>650</value>
      <product>
        <name>Super X4</name>
        <part_number>SX9467675893</part_number>
        <description>Super X4 in green and red</description>
      </product>
    </sale>
  </sales_person>
  <sales_person>
    <name>Edward Evans</name>
    <sales_target_count>14</sales_target_count>
    <sales_target_value>1900.00</sales_target_value>
    <sale>
      <date_of_sale>06/03/2011</date_of_sale>
      <quantity>3</quantity>
      <value>100</value>
      <product>
        <name>Dyno V3</name>
        <part_number>DNV3-48953-FGGHEJ</part_number>
        <description>Dyno V3 with large cover</description>
      </product>
    </sale>
    <sale>
      <date_of_sale>06/03/2011</date_of_sale>
      <quantity>3</quantity>
      <value>100</value>
```

```
<product>
  <name>Dyno V3</name>
  <part_number>DNV3-48953-FGGHEJ</part_number>
  <description>Dyno V3 with large cover</description>
</product>
</sale>
<sale>
  <date_of_sale>10/03/2011</date_of_sale>
  <quantity>4</quantity>
  <value>650</value>
  <product>
    <name>Super X4</name>
    <part_number>SX9467675893</part_number>
    <description>Super X4 in green and red</description>
  </product>
</sale>
<sale>
  <date_of_sale>12/03/2011</date_of_sale>
  <quantity>2</quantity>
  <value>380</value>
  <product>
    <name>Super X4</name>
    <part_number>SX9467675893</part_number>
    <description>Super X4 in green and red</description>
  </product>
</sale>
</sales_person>
</sales_area>
</sales_report>
```

Listing 1

The point of interest in **Listing 1** as far as our purposes are concerned is that the inherent hierarchical nature of XML documents that do not utilise IDREFs means that the details for the two products are repeated numerous times, once for each sale.

Next we shall see that it is possible to eliminate data duplication by using IDREFs.

4. Schema 2

In order to utilise IDREFs it is first necessary to change the schema. A possible schema (listing in Appendix 2) that will support the requirement to report on sales and eliminate data duplication is shown in **Figure 3**.

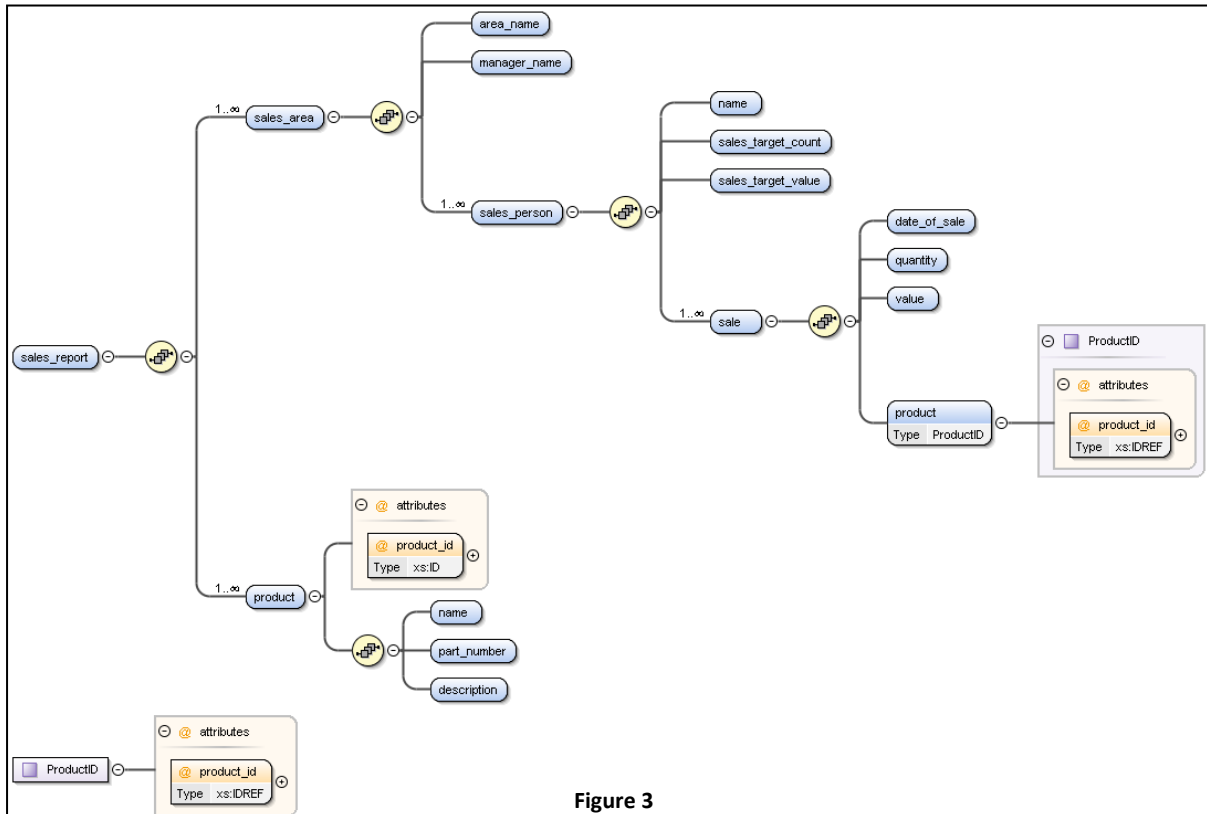


Figure 3

It can be seen from **Figure 3** that a new type, **ProductID**, has been introduced which has one attribute of type IDREF. IDREF is one of the data types available in XML.

The **Product** structure has been promoted in the schema hierarchy to be directly under the `sales_report` node and given an unbounded multiplicity. **Product** has also been given one attribute of type ID. ID is one of the data types available in XML.

The original **product** structure that existed under `sale` has been replaced by an element of type **ProductID** this creates a pointer in the document that will reference another element in the document with the same corresponding attribute name, `product_id` in this case. The actual instance in the document pointed to is determined by the value of the `product_id` attribute in the **product** element. For the document to conform to the schema there must be a **Product** element instance with the same value in its `product_id` attribute. All the `product_id` attribute values in the **product** must be unique within the document.

With the schema shown in **Figure 3** the corresponding message for the sales described above would be:

```
<?xml version="1.0" encoding="UTF-8"?>
<sales_report>
  <sales_area>
    <area_name>Area North</area_name>
    <manager_name>Norman Noaks</manager_name>
    <sales_person>
      <name>Albert Allis</name>
      <sales_target_count>24</sales_target_count>
      <sales_target_value>3400.00</sales_target_value>
      <sale>
        <date_of_sale>01/03/2011</date_of_sale>
        <quantity>4</quantity>
        <value>560</value>
        <product product_id="p1"/>
      </sale>
      <sale>
        <date_of_sale>14/03/2011</date_of_sale>
        <quantity>2</quantity>
        <value>300</value>
        <product product_id="p1"/>
      </sale>
      <sale>
        <date_of_sale>20/03/2011</date_of_sale>
        <quantity>1</quantity>
        <value>400</value>
        <product product_id="p2"/>
      </sale>
      <sale>
        <date_of_sale>14/03/2011</date_of_sale>
        <quantity>2</quantity>
        <value>350</value>
        <product product_id="p1"/>
      </sale>
    </sales_person>
    <sales_person>
      <name>Bertram Barnard</name>
      <sales_target_count>8</sales_target_count>
      <sales_target_value>2100.00</sales_target_value>
      <sale>
        <date_of_sale>06/03/2011</date_of_sale>
        <quantity>3</quantity>
        <value>100</value>
        <product product_id="p2"/>
      </sale>
      <sale>
        <date_of_sale>10/03/2011</date_of_sale>
        <quantity>4</quantity>
        <value>650</value>
        <product product_id="p1"/>
      </sale>
    </sales_person>
    <sales_person>
      <name>Cerys Cowan</name>
      <sales_target_count>14</sales_target_count>
      <sales_target_value>1900.00</sales_target_value>
      <sale>
        <date_of_sale>12/03/2011</date_of_sale>
        <quantity>2</quantity>
        <value>380</value>
        <product product_id="p1"/>
      </sale>
    </sales_person>
  </sales_area>
</sales_report>
```

```
<sales_area>
  <area_name>Area South</area_name>
  <manager_name>Susan Smith</manager_name>
  <sales_person>
    <name>Dawn Dove</name>
    <sales_target_count>14</sales_target_count>
    <sales_target_value>1900.00</sales_target_value>
    <sale>
      <date_of_sale>12/03/2011</date_of_sale>
      <quantity>2</quantity>
      <value>380</value>
      <product product_id="p1"/>
    </sale>
    <sale>
      <date_of_sale>06/03/2011</date_of_sale>
      <quantity>3</quantity>
      <value>100</value>
      <product product_id="p2"/>
    </sale>
    <sale>
      <date_of_sale>10/03/2011</date_of_sale>
      <quantity>4</quantity>
      <value>650</value>
      <product product_id="p1"/>
    </sale>
  </sales_person>
  <sales_person>
    <name>Edward Evans</name>
    <sales_target_count>14</sales_target_count>
    <sales_target_value>1900.00</sales_target_value>
    <sale>
      <date_of_sale>06/03/2011</date_of_sale>
      <quantity>3</quantity>
      <value>100</value>
      <product product_id="p2"/>
    </sale>
    <sale>
      <date_of_sale>06/03/2011</date_of_sale>
      <quantity>3</quantity>
      <value>100</value>
      <product product_id="p2"/>
    </sale>
    <sale>
      <date_of_sale>10/03/2011</date_of_sale>
      <quantity>4</quantity>
      <value>650</value>
      <product product_id="p1"/>
    </sale>
    <sale>
      <date_of_sale>12/03/2011</date_of_sale>
      <quantity>2</quantity>
      <value>380</value>
      <product product_id="p1"/>
    </sale>
  </sales_person>
</sales_area>
<product product_id="p1">
  <name>Super X4</name>
  <part_number>SX9467675893</part_number>
  <description>Super X4 in green and red</description>
</product>
<product product_id="p2">
  <name>Dyno V3</name>
  <part_number>DNV3-48953-FGGHEJ</part_number>
  <description>Dyno V3 with large cover</description>
</product>
</sales_report>
```

Listing 2

The two products now only appear in the document once and have been given IDs of **p1** and **p2**. At all the points in **Listing 1** where a product element appeared it has been replaced with a reference to either **p1** or **p2**.

The new XML document at **Listing 2** is some 46 lines (26%) shorter than the original at **Listing 1** and data redundancy has been reduced as the definitions for the two products are not repeated within the document.

Important note: The ID end of the ID/IDREF pairing must have a unique value within the **whole** document (**p1** and **p2** in the example above). This applies regardless of the name given to the ID/IDREF pairing (**product_id** in the example above). Origo usually gives all IDs the name, **id**, with the associated IDREFs being named specifically according to their use e.g. **product_id**.

5. Summary

It should be noted that using IDREFs is not necessarily appropriate in all circumstances. The purpose and nature of XML documents valid for a particular schema design should be taken into account before deciding how much, if any, normalising should be done. The use of IDREFs will usually slow down the validation of an XML document against its schema and they are also quite likely to slow down the production and consumption of XML documents.

Most commercially available XML processors are usually optimised to work in either a stream mode or non-stream mode. In stream mode an XML document is processed without holding the whole document in memory. Stream mode processing can generally be done very rapidly and is useful for processing large XML documents in a serial manner. In stream mode IDREFs can become a hindrance as backtracking or programmer defined solutions may be required to resolve IDREF pointers.

Non-streaming XML processors load an entire XML document into memory before processing it. For this type of processor IDREFs can be very useful as they reduce the overall size of a document and can therefore reduce the amount of memory required by a program to process a document. Navigating a document wholly held in memory by using IDREF pointers can be very efficient.

6. Appendix A – Schema 1 Listing

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="sales_report">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" name="sales_area">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="area_name"/>
              <xs:element name="manager_name"/>
              <xs:element maxOccurs="unbounded" name="sales_person">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element name="name"/>
                    <xs:element name="sales_target_count"/>
                    <xs:element name="sales_target_value"/>
                    <xs:element maxOccurs="unbounded" name="sale">
                      <xs:complexType>
                        <xs:sequence>
                          <xs:element name="date_of_sale"/>
                          <xs:element name="quantity"/>
                          <xs:element name="value"/>
                          <xs:element name="product">
                            <xs:complexType>
                              <xs:sequence>
                                <xs:element name="name"/>
                                <xs:element name="part_number"/>
                                <xs:element name="description"/>
                              </xs:sequence>
                            </xs:complexType>
                          </xs:element>
                        </xs:sequence>
                      </xs:complexType>
                    </xs:element>
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

7. Appendix B – Schema 2 Listing

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="sales_report">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" name="sales_area">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="area_name"/>
              <xs:element name="manager_name"/>
              <xs:element maxOccurs="unbounded" name="sales_person">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element name="name"/>
                    <xs:element name="sales_target_count"/>
                    <xs:element name="sales_target_value"/>
                    <xs:element maxOccurs="unbounded" name="sale">
                      <xs:complexType>
                        <xs:sequence>
                          <xs:element name="date_of_sale"/>
                          <xs:element name="quantity"/>
                          <xs:element name="value"/>
                          <xs:element name="product" type="ProductID"/>
                        </xs:sequence>
                      </xs:complexType>
                    </xs:element>
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element maxOccurs="unbounded" name="product">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="name"/>
        <xs:element name="part_number"/>
        <xs:element name="description"/>
      </xs:sequence>
      <xs:attribute name="product_id" type="xs:ID" use="required"/>
    </xs:complexType>
  </xs:element>
  </xs:sequence>
</xs:complexType>
</xs:element>
<xs:complexType name="ProductID">
  <xs:attribute name="product_id" type="xs:IDREF" use="required"/>
</xs:complexType>
</xs:schema>
```