

剖面数据的 XML 格式设计

刘海行, 周玉斌

(国家海洋局第一海洋研究所 海洋环境科学和数值模拟重点实验室, 山东 青岛 266061)

摘要: 根据可扩展标记语言 (XML) 的原理, 设计了针对剖面数据的 XML 存储格式, 并给出了一个 XBT 实测剖面数据的实例。该格式可用于海洋环境要素的剖面数据交换, 并可进一步用于断面数据的数据交换。

关键词: 可扩展标记语言 (XML); Marine XML; 剖面; 数据交换

中图分类号: TP31; P7 **文献标识码:** A **文章编号:** 1000-3096(2004)10-0001-05

利用各种海洋观测仪器进行海上实际观测, 是人类了解、研究海洋的重要手段之一。随着科学技术的不断发展, 海洋观测仪器也在不断进步, 人类能够获取的海洋环境要素数据不仅种类越来越多, 精度也在不断提高。因此对海洋环境要素数据的描述也越来越复杂。

一般来说, 利用海洋观测仪器进行海上实际观测, 在仪器投放后需要通过处理软件将传感器数据转换为人类可读的原始观测数据。由于海洋观测仪器多种多样, 原始观测数据的数据格式也多种多样, 这就给对原始观测数据的处理和交换带来了困难。并且随着海洋观测仪器种类的不断增多、同一种海洋观测仪器其核心技术的不断升级, 原始观测数据的存储格式也不断增多。因此如果只是简单地保存转换后的原始观测数据, 无疑会为这些数据的交换和使用带来不便。

近年来, 针对海洋环境数据的交换使用, 国内外都进行了大量的数据交换格式设计的工作。特别是随着 Internet 的不断发展, 更是有许多优秀的数据交换格式出现。可扩展标记语言 (XML) 就是其中的一种。

作者根据 XML 的原理, 讨论了用于海洋环境数据的 XML 格式的一般结构, 设计了针对海洋环境要素剖面数据的 XML 格式标记, 并给出了一个 XBT 剖面实测数据的 XML 格式实例。同时还讨论了断面数据的 XML 格式设计。

1 XML 与 Marine XML

XML^[1] (Extensible Markup Language, 可扩展标记语言) 是由 W3C (World Wide Web Consortium, 万维网联盟) 于 1998 年 2 月发布的一种标准, 与 HTML 一样是 SGML (Standard Generalized Markup Language, 通用标记

语言) 的一个简化子集。它结合了 SGML 的丰富功能与 HTML 的易用性, 以一种开放的、自我描述方式定义了数据结构。在描述数据内容的同时能突出对结构的描述, 从而体现出数据之间的关系。XML 具有以下几个突出特点。

1.1 XML 是一种元标记语言

与 HTML 不同的是 XML 没有一套预定义的标记, 而是使用 DTD (Document Type Definition)。DTD 列出所有合法标记, 并且指定文档在什么位置以及如何包含该标记。因而可以像 SGML 那样作为元语言来定义其它标记语言。如果把标记语言分为元标记语言和实例标记语言的话, SGML 和 XML 都是元标记语言, 而 HTML 则是实例标记语。

正是由于这一特点, 使得用户可以创建自己的 DTD。换言之, 可以自己定义一种适用于自己业务领域的标记语言。

1.2 XML 是一种语义/结构化标记语言

XML 描述文档的结构和含义而不描述页面上元素的格式。虽然格式可以通过样式表增加到文档中, 但文档本身仅包含描述文档内容的标记, 而不是文档的外观。因此在 XML 中可以使用有含义的标记名称从而

收稿日期: 2004-08-04 修回日期: 2004-08-16
基金项目: 国家自然科学基金青岛海洋科学资料共享服务中心建设
作者简介: 刘海行 (1961-), 男, 山东青岛人, 研究员, 主要从事海洋环境信息系统工作, 电话: 0532-8967412 E-mail: liuhx@fio.org.cn

使得人类通过阅读源代码就可以了解文档的含义。

XML 标记名称也可以使用特殊含义, 如数据库的字段名。显然, 这就使得 XML 对于各种应用更加灵活和更容易被接受。

1.3 XML 可以做为一种通用的数据交换格式

从低级观点来看, XML 是一种极其简单的数据格式。XML 文档是纯 ASCII 码文本文件, 因此可以适合于各种平台环境的数据交换。XML 支持 unicode, 因此文本形式不会因为文字代码的不同而造成不能阅读的问题。

从高级观点来看, XML 是一种自描述语言。它的有含义的标记使得即使不熟悉 XML 的人也可以基本理解 XML 文档所描述的内容。同时, 这些有含义的标记可以有效地用于元数据的描述。

XML 是一个非专有格式, 无知识产权的制约, 并且易理解、易读写。因此在不同的应用程序之间用于数据交换是一种极好的选择。

1.4 XML 是一种面向对象的标记语言

由于 XML 通过 DTD 定义文档的结构, 使得开发者可以几乎随心所欲地描述数据及数据之间的关系。因此, 完全可以按照面向对象的数据结构构造文档的结构, 这是到目前为止其他方法所不具备的数据描述特点。

1.5 XML 是一种易于表示的标记语言

XML 的基本原理之一就是把数据的描述和数据

的表示分离, 使得一个 XML 文档只描述数据本身, 而通过 XSL (Extensible Stylesheet Language, 可扩展样式表语言) 表示数据。这就使得数据的表示可以随需求的不同而不同, 并且这个过程和 XML 本身无关。

XML 的上述特点使得 XML 在各个领域等到了广泛的应用。如: MathML (数学)、CML (化学)、SMIL (多媒体演示)、BML (卫星数据传送) 等。

近年来, 随着 Internet 的快速发展和对海洋环境要素数据交换需求的不断增长, 基于 XML 的海洋环境要素数据交换格式 Marine XML 的研究和开发也逐步兴起^[2]。众多海洋科研组织纷纷设计、开发了 Marine XML 的各种格式和工具, 如: IOC, AODC^[3], NASA 等。

海洋环境要素数据由于海洋环境的特殊性, 一般来说都具有时间、空间的属性, 即四维属性。同时, 由于海洋观测手段的多样性和海洋环境的不稳定性, 海洋观测数据的元数据描述具有十分重要的作用。由于 XML 的元标记语言属性, 这些问题在利用 XML 设计用于海洋环境要素数据的 XML 标记语言时都可以得到较好的解决。

2 Marine XML 的基本结构

设计一整套适合海洋环境要素数据的 XML 标记语言超出了该研究的能力, 作者仅就 XML 的原理讨论 Marine XML 的基本结构和适用于剖面数据的 XML 标记及结构。

Marine XML 的基本结构见图 1。

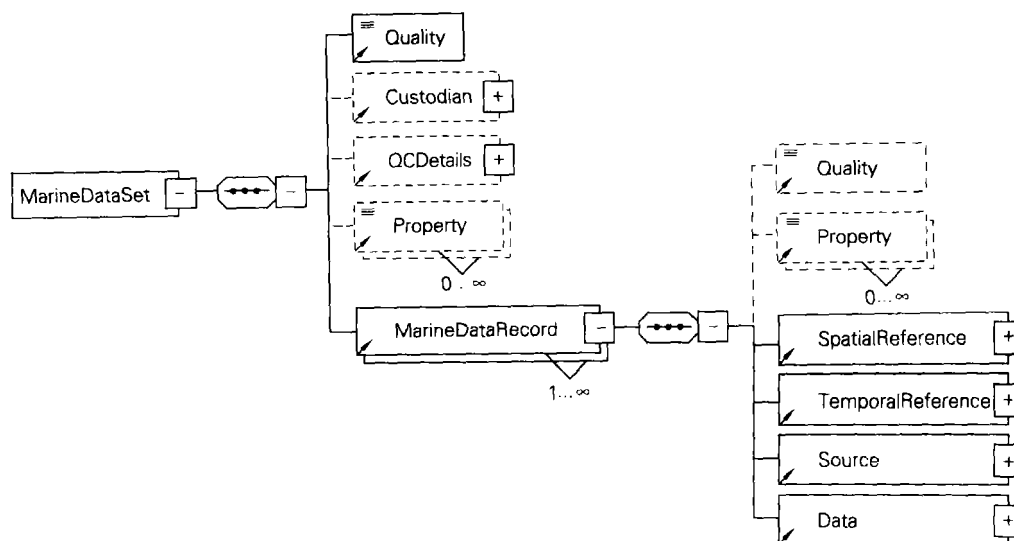


图 1 Marine XML 的基本结构

Fig. 1 Marine XML structure

Marine XML 的主体由根元素和数据记录元素构成。

根元素用 MarineDataSet 表示。在根元素下, 可以加入 Quality(数据集质量) 元素、Custodian(管理人) 元素、QCDetail(质量控制描述) 元素以及 Property(其它属性) 元素。这些元素描述了关于该数据集的一些基本信息。其中值得注意的是 QCDetail 元素, 它可以包含对数据集中的数据进行修改的记录。从而可以既保留原始数据, 又可以保留经不同处理方法处理后的数据。

数据记录元素用 MarineDataRecord 表示, 描述数据集中的数据。每一个 MarineDataRecord 包含了一组海洋环境要素数据。Quality 元素和 Property 元素的含义与根元素相同。

一个数据集中通常含有多个数据记录, 因此, MarineDataRecord 元素应该具有描述记录号的属性, 其一般格式为:

```
< MarineDataRecord id= "0">
```

其中 id 属性的数字表示记录号。

如前所述, 海洋环境要素数据一般具有四维属性, 因此在 MarineDataRecord 元素下设置了 SpatialReference 元素描述空间信息和 TemporalReference 元素描述时间信息。同时为了保留元数据信息, 设计了 Source 元素, 用于描述数据源的有关信息。Data 元素则封装该数据记录的数据部分。

对于剖面数据, 考虑一个剖面的情况, 其 Marine XML 的主体结构如下:

```
< ?xml version= "1.0" standalone= "yes"?>
< MarineDataSet creationDate= "2004- 07- 31 17: 00: 00">
< MarineDataRecord id= "0">

    < ! -- SpatialReference, TemporalReference and
    Data -- >
    < /MarineDataRecord>
< /MarineDataSet>
```

3 剖面数据的 SpatialReference 元素结构

一般来说, 海洋环境数据有站点数据、剖面数据、断面数据和区域数据等方式。从空间分布特征考虑, 应该有单点坐标、航线坐标和区域(规则或不规则)坐标。相应地, SpatialReference 元素应包括 GeoPoint, GeoLine, GeoBox 和 GeoPolygon 4 个元素。

对于构造 GeoPoint 元素、GeoLine 元素和 GeoBox 元素而言, 需要一个或多个坐标对, 因此其子元素只需要坐

标元素 Coordinates 即可。考虑到水体及大气的三维特征, Coordinates 元素应包括水深/海拔元素。

对于剖面数据而言, 使用 GeoPoint 即可, 其 XML 代码如下:

```
< SpatialReference>
< GeoPoint>
    < Coordinates datum= "WGS84">
        < Longitude> 115.00< /Longitude>
        < Latitude> - 15.00< /Latitude>
    < /Coordinates>
< /GeoPoint>
< /SpatialReference>
```

注意到, 由于是剖面数据, 所以这里并没有水深信息。

GeoPolygon 元素要相对复杂一些。由于区域的构成可能会包括空洞, 仅考虑区域的外边界是不够的。因此, GeoPolygon 元素应由外边界线元素和内边界线元素构成, 而边界线则由多个闭合曲线构成, 闭合曲线由坐标对构成。限于篇幅, 作者不对 GeoPolygon 元素坐进一步的描述。

4 剖面数据的 TemporalReference 元素结构

对于海洋环境要素数据而言, 其时间属性一般包括单一时刻(如 XBT)、连续时间段(如 ADCP)、周期性时间段(如每年的某一时间或时段)。

单一时刻用 Instant 元素表示, 其子元素只有一个 Date 元素, Date 元素使用属性描述一个确定的时间。

时间段使用 Period 元素表示, 对于连续时间段, 使用子元素 Start 和 End 表示开始时间和结束时间; 对于周期性时间段, 通过在 Start 元素和 End 元素之间嵌入一个或多个 Period 元素表示。限于篇幅, 作者不做详述。但应该看到, 对周期性时间段的描述方法体现了 XML 的优越性, 这种嵌入 Period 的方法是其它格式设计方法很难做到的。

对于剖面数据而言, 使用单一时刻即可, 其 XML 格式如下:

```
< TemporalReference>
< Instant> < Date year= "2003" month= "10" day= "3"
```

hour = "16" minute = "11" second = "58" time-
Zone= "UTC" /> < /Instant>
< /TemporalReference>

5 剖面数据的 Data 元素结构

由于 Data 元素中可以包括一组以上的数据, 因此

Data 元素中使用多个 DataObject 元素装入数据, 使用 numberOfDataObjects 属性指明 DataObject 的个数。同时由于海洋环境数据通常包括一个以上的物理要素, 在使用 DataObject 装入数据时必须对物理要素进行区分, 因此使用 Parameter 元素或 ParameterSet 元素分别描述单个物理要素或一组物理要素。其基本结构见图 2。

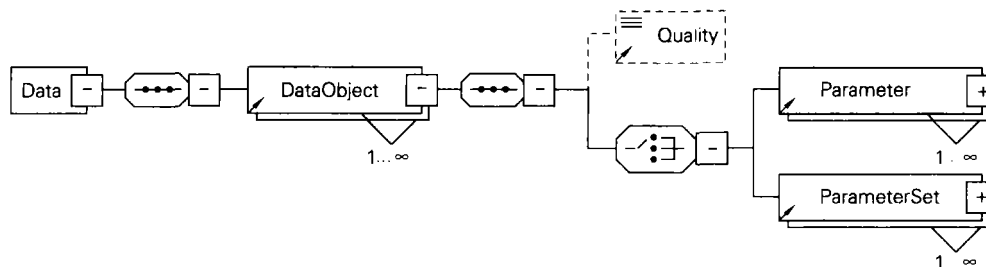


图 2 Data 元素基本结构

Fig. 2 Data element structure

其中, 如果使用 ParameterSet 元素, 则其为一组 Parameter 元素。

对于剖面数据, DataObject 及 Parameter 的设计有多种方法。作者给出一种封装为一个 DataObject 的方法描述 XBT 的一次观测。其物理要素包括水深、温度和声速, 沿水深方向只给出 8 层。其 XML 代码如下:

```
< Data numberOfDataObjects= "1">
  < DataObject index= "0" numberOfParameterSets=
    "1">
    < ParameterSet index= "0" numberOfParameters=
      "3">
      < Sensor name= "XBT" model= "MK21 T- 7"/>
      < Parameter index= "0" name= "depth" long
        name= "sea depth" units= "Metres"/>
      < Parameter index= "1" name= "temp" longname
        = "temperature" units= "Degrees C"/>
      < Parameter index= "2" name= "sv" longname=
        "sound velocity" units= "m/s"/>
      < ValueList numberOfValueSets= "8">
        0. 0, 28. 94, 1518. 48 0. 7, 28. 70, 1535. 39
        1. 3, 28. 80, 1526. 05 2. 0, 28. 40, 1485. 22
        2. 7, 23. 95, 1526. 46 3. 3, 28. 67, 1537. 59
        4. 0, 28. 66, 1537. 59 4. 7, 28. 69, 1537. 66
```

```
< /ValueList>
< /ParameterSet>
< /DataObject>
< /Data>
```

这种设计方式的优点在于一次剖面数据使用一个 ParameterSet, 比较简洁, 形成的文档较小。使用文本编辑器或浏览器 IE 直接阅读 XML 文档时容易理解。缺点是数据过于集中, 并未完全遵守 XML 的规则。

如果将 ParameterSet 分解并使用 Value 元素形成完全分离的数据, 则单层 XBT 剖面数据的 XML 代码如下:

```
< Data numberOfDataObjects= "1">
  < DataObject index= "0" numberOfParameterSets=
    "10">
    < ParameterSet index= "0" numberOfParameters=
      "3">
      < Sensor name= "XBT" model= "MK21 T- 7"/>
      < Parameter index= "0" name= "depth" longname
        = "sea depth" units= "Metres"/>
      < Value> 0. 0< /Value> < /Parameter>
      < Parameter index= "1" name= "temp" longname
        = "temperature" units= "Degrees C"/>
      < Value> 28. 94< /Value> < /Parameter>
      < Parameter index= "2" name= "sv" longname=
        "sound velocity" units= "m/s"/>
      < Value> 1518. 48< /Value> < /Parameter>
```

```
< /ParameterSet>  
< !- more ParameterSet- - >  
< /DataObject>  
< /Data>
```

显然,这种方式会形成较大的文件,并且直接阅读比较困难。但这种方式更好地遵守了 XML 的规则,对于每一个元素或数据解析都会更容易。

7 结语

作者根据 XML 的原理和 Marine XML 的一般结构给出了针对剖面数据的 XML 格式设计。事实上,通过增加 DataObject,这个格式可以进一步用于断面数据。对于 ADCP 的连续观测,该格式也可以通过修改 TemporalReference 而用于连续时间段。

限于篇幅,作者没有给出与该设计相对应的 XSL。一方面,由于 XSL 主要用于表示,随个人喜好不

同;另一方面,目前的浏览器(如:IE)已经支持直接浏览 XML 代码,可以方便地浏览 XML 文件。

通过作者的实例,可以看出 XML 作为一种元标记语言,在数据交换方面有巨大的优势,对于结构复杂的海洋环境数据也是一种优秀的数据交换工具。

参考文献:

- [1] 孔小玲,商艳莉译. XML 技术手册[M]. 北京:中国电力出版社,2001. 11- 34.
- [2] Isenor A W, Keeley J R, Linganti J. Developing an extensible markup language (XML) application for DFO marine data exchange via the web [EB/OL]. [http:// www. marinexml. net](http://www.marinexml.net), 2003- 10- 07.
- [3] Romai B, Sliogeris P, Plater M, *et al.* Development and use of marine XML within the Australian oceanographic data centre to encapsulate marine data[EB/ OL] . [http:// www. marinexml. net](http://www.marinexml.net), 2002- 08- 28.

Design of XML structure for profile data recording

LIU Hai- xing, ZHOU Yu- bin

(Key Laboratory of Marine Science and Numerical Modeling, First Institute of Oceanography, State Oceanic Administration, Qingdao 266061, China)

Received: Aug. 4, 2004

Key words: extensible markup language (XML); marine XML; profile; data exchange

Abstract: This paper introduced a new design of extensible markup language (XML) for marine profile data with an example of XBT observation data. This format provides a general tool to record marine profile data, and also can be used for data exchange.

(本文编辑: 刘珊珊)