Oracle *inter*Media Life Sciences Image Metadata Management

January 2005

Oracle Database Manages Life Sciences Image Metadata

Introduction

Life Sciences applications face a couple of image data management problems that make it difficult to query, search, reuse, and manage images and associated contextual data in the discovery process.

When a Life Sciences image is created, additional contextual data (application-related, analytical, and physically descriptive data) is associated with the image. This contextual data is commonly referred to as *metadata* and is an integral part of the image.

The first problem relates to the association between an image and its metadata. Because the image and metadata are stored separately, the association is not maintained when the image is processed, shared between applications, or transmitted across a network. An application should be able to easily obtain metadata for a given image and the reverse should work as well.

The second problem relates to searching. If the association between an image and metadata is not maintained then metadata cannot be used to retrieve a combination of image and metadata elements (and other related relational data).

Common approaches to metadata management are the spreadsheet approach and the TIFF file approach. In the first, metadata is stored separately from the image in one or more spreadsheets. In the second, metadata is stored in the TIFF file header using special tags. The spreadsheet approach does not maintain the association between the image and the metadata. The TIFF approach maintains the association, but does not address the problem of searching.

Kinds of Metadata

Metadata is broadly defined in information technology as data about data, or data used to describe data. More specifically, it includes these three types of information:

- **Technical metadata** includes the height, width, mime type, color depth, file and compression format of an image.
- **Content metadata** is created by lab equipment and software and includes:
    a. **Identification data** such as date, lab id., experiment id., and experimenter id.
    b. **Annotations,** which are text and graphics added by technicians and scientists, based on their observations of the image. An annotation can be associated with the whole image, a 2-D region within the image, a layer of a 3-D volumetric or temporal image or a particular degree of resolution. For example, in the ultra-structural view of a brain section a scientist might annotate at the level of large scale features, then at the level of specific brain regions, and finally at the cellular level.

c. **Analyses** of an image, such as color histograms and intensity values that are data derived from an image by applying image analysis tools. They can change as new analysis techniques are applied to the image.

Metadata Management and Oracle *inter*Media

Technical Metadata: Oracle *inter*Media can extract image property metadata from an image. The metadata is stored in attributes of the *inter*Media image object (if objects are used; otherwise the application can store them in columns in the database). This feature is supported for all the image formats *inter*Media supports.

Content Metadata: Oracle *inter*Media 10g Release 2 (currently in Beta) provides new capabilities for application developers to read and write user-defined metadata in the headers of TIFF, JPEG, and GIF image files. Storing metadata with image data in the same container file provides encapsulation. With encapsulation, both the image and the metadata can be shared and exchanged reliably as one unit. The schema for storing the metadata is determined by the application, giving applications tremendous flexibility in defining the metadata they store. *inter*Media uses XMP, the Extensible Metadata Platform from Adobe Corporation, as a framework for the storage of metadata. XMP makes use of the Resource Data Framework (RDF) standard that is defined as part of the Semantic Web framework by the World Wide Web Consortium (W3C). XMP is used in the desktop publishing and applications domain. A schema can be defined within the XMP framework, and metadata based on the schema is instantiated as an XML document, enabling tools that understand XML to understand the metadata. The metadata methods associated with these capabilities are below:

putMetadata() for XMP: This method accepts a valid XMP document and creates a binary packet suitable for embedding in an image file. This binary packet is written to the image, replacing any existing packets.

getMetadata() for XMP: This method extracts the metadata from the image header, and returns it to the application as a valid XML document.

Advantages of Using Oracle *inter*Media Metadata Management Features

- *inter*Media makes images more meaningful and easier to share by maintaining the association between the image and metadata. For example, when an image is downloaded from a Website, the metadata is contained within the image and hence is not “lost”.
- *inter*Media makes images easier to search and reuse by extracting metadata into an XML document, and storing it in a database column of XMLType, which can be searched and processed using Oracle XML DB capabilities.
- Images and metadata stored using *inter*Media are protected and managed by Oracle Database, can be easily associated with other application data, and kept in synch with other data.
- *inter*Media enables images and metadata to be natively stored in the database and have all the advantages of other application data in the Oracle Database, including

backup, recovery, security, transaction control, scalability, shared access across multiple applications, and integration with other Oracle tools.

Example

This is an example of an XMP schema and a sample instantiation of an XML document containing metadata based on that schema. This example was put together in collaboration with the Sanger Institute.

The XMP namespace prefix is lsAD. The XMP namespace is http://ns.oracle.com/im/ls/1.0/ApplicationData

Property	Value Type	Description
Creator	bag Text	List of creators of the image
Organization	bag Text	The creators' organization
ExpName	Text	The name of the experiment
ExpDate	Date	The name of the experiment
Device	Text	The device used to acquire the image
DeviceSetting	Text	The device setting used to acquire the image
Software	Text	The software used to process the image
SoftwareVersion	Text	The version of the software used to process the image
SubImageRelation	Text	The relationship among the sub-images

The XMP namespace prefix is lsCD. The XMP namespace is http://ns.oracle.com/im/ls/1.0/ContentDescription

Property	Value Type	Description
ImageIdx	Integer	Image index
Dimension	Dimensions	Image dimension
Resolution	Rational	Number of pixels
ColorSpace	Text	Color space of the image, such as RGB, CMYK
NumberOfColors	Integer	Number of colors in the image
AvgColor	Color	Image average color
ColorHistogram	seq Integer	Color histogram of the image
MeanIntensity	Integer	Mean intensity of the image

StdDevIntensity	Real	Standard deviation of the image intensity
MeanGrayLevel	Integer	Mean gray level of the image
AvgTexture	seq Integer	Image average texture
ImageOps	Text	The operations that were applied to generate the image

Sample Document

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<rdf:Description
xmlns:lsAD="http://ns.oracle.com/im/ls/1.0/ApplicationData">

  <lsAD:Creator>Melli Annamalai</lsAD:Creator>
  <lsAD:Organization>Oracle Corporation</lsAD:Organization>
  <lsAD:ExpName>Color Ratio Analysis</lsAD:ExpName>
  <lsAD:ExpDate>10-Mar-2004</lsAD:ExpDate>
  <lsAD:Device> Oracle Scanner </lsAD:Device>
  <lsAD:Software> Oracle's Image Analysis Tool
</lsAD:Software>

</rdf:Description>

<rdf:Description
xmlns:lsCD="http://ns.oracle.com/im/ls/1.0/ContentDescription" >

   <lsCD:ImageIdx> 0 < /lsCD:ImageIdx>
   <lsCD:Dimension>
        <stDim:w> 69 </stDim:w>
        <stDim:h> 93 </stDim:h>
   </lsCD:Dimension>
   <lsCD:Resolution> 6417 </lsCD:Resolution>
   <lsCD:ColorSpace> RGB </lsCD:ColorSpace>
   <lsCD:AvgColor>
        <stColor:R> 180 </stColor:R>
        <stColor:G> 124 </stColor:G>
        <stColor:b> 154 </stColor:B>
   </lsCD:AvgColor>
   <lsCD:NumberOfColors> 7 </lsCD:NumberOfColors>
</rdf:Description>

</rdf:RDF>
```

Conclusion

Life Sciences applications typically associate a rich variety of descriptive metadata with images. Oracle Database 10g Release 2 increases the return on an organization's valuable image assets (ROA). It adds Oracle *inter*Media support to read and encapsulate user-defined metadata in the headers of TIFF, JPEG, and GIF image files, and extract that metadata into an XML document for persistent storage, indexing, and searching.

Oracle interMedia Life Sciences Image Metadata Management
May 2004
Authors: Melli Annamalai, Dong Lin

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
www.oracle.com