Appendix A -- Preferred XML Formats

BizInt Smart Charts can import data in a number of different formats, but our experience is that the best results come from well-implemented XML. This document describes what we mean by "well-implemented".

BizInt Smart Charts software makes every effort to identify the source of every record, and to preserve the branding and use restrictions set forth by our publisher partners. The essential elements to doing this are also set forth in this document.

1. File Structure

    There are three ways to implement a collection of records in XML:

    (a) an XML container element containing a list of record elements

    (b) a packed XML file containing record elements back-to-back, without a container element

    (c) a collection of XML files, one record per file (this format requires the files to be packaged in a ZIP archive)

    As noted above, BizInt Smart Charts needs to identify, unambiguously, the source of the data. For this reason, we expect the XML document element to have a distinguishing name, such as:

      <OurSystemToBizInt>

    One advantage of using this element as a container is that the records within the file can use industry standards for record elements without any confusion.

    Whether there is a single XML file or many, the data may be compressed in a ZIP archive for download. A ZIP archive is required for option (c) above.

    PREFERRED: option (a) with a distinguishing container name compressed in a ZIP archive

2. Encoding

    Multi-byte characters should be encoded using numeric entities (e.g. ⇳).

3. Download Support

    In order to automatically invoke BizInt Smart Charts for Patents, the exported file should have a .BPD extension.

    In order to automatically invoke BizInt Smart Charts for Drug Pipelines, the exported file should have a .BRD extension.

4. Record Structure

    A record is a collection of fields. Empty fields may be omitted, or may be present with empty values. Do not use placemarkers (such as a '-') in empty fields. The value "None" or "N/A" should only be used if it is meaningful, not to indicate an empty field.

    The order of fields in a record does not matter.

    All fields must have both a label and a value.

    For systems which present several collections of data (Hosts), we prefer a generic field label presentation. In a generic presentation, each field should include a tag and a text label. The text label may be in the user's language. BizInt Smart Charts will use the tag to identify the field, but may use the text label for user presentation.

    Two generic presentations are equivalent:

```
<field tag="tag" label="label">...

<field><tag>tag</tag><label>label</label>...
```

For all other systems, we prefer an XML presentation where the element name is the field label. For example, a Patent Assignee might be presented in an element named <PatentAssignee>.

5. What constitutes a Field?

    It is probably best to explain what we consider a field by way of an example. A Patent Family consists of a list of publications. Each publication consists of a publication number, and a publication date among other information. Each publication number consists of a granting authority, possibly a year, a serial number, and a kind code. For purposes of export to BizInt Smart Charts, we consider the entire Patent Family to be a field. The development effort to construct fields from component XML elements will delay implmentation of a database.

6. Field Value Structure

    Record fields consist of one or more data items. We support two structures for these lists of items.

    The first approach is to present the items with a separator element between each item. This is similar to placing a line break between items in a text presentation. A suitable separator would be <sep/>.

    ```
    Item 1<sep/>Item 2
    ```

    The second approach is wrap each item in an element. Multiple items would simply be presented back to back, without need for a separator.

    ```
    <item>Item 1</item><item>Item 2</item>
    ```

    Fields containing text flows may use elements to indicate the text structure, such as <heading> and <para>, rather than the generic <item>.

    Within an item in a field, structure may be indicated by elements showing data separation, such as <tab/> or <is/>.

    There are two considerations on where to separate item components. The first is that all databases within a collection should be treated the same. In our example of a Patent Family, if a separator is placed between the authority and the serial number for one database, it should be placed there for all similar databases.

    Second, it is important not to subdivide the field items too finely. An example is the presentation of dates. BizInt Smart Charts prefers a date in CCYY-MM-DD format (2009-08-31), rather than having year, month, and date as separate elements.

    Our example of a Patent Family entry might read

    ```
    <PatentFamily>

      <item>US 5000000<is/>B<is/>1990-6-28</item>   ...
    ```

    When an export format is for a single product, items may have more meaningful element names. For example, a list of Inventors may be shown as

    ```
    <Inventors> <Inventor>first name</Inventor>
    ```

    Of course the generic <item> could also be used in this case.

7. Metadata

    Every record must have a unique identifier (an accession number) that can be used to ensure that only one copy of a record appears in a chart file. For first-level patent data, we often use the application number for this

purpose. For family-level data, an identifier must be provided in the data.

If the system has update dates for records, these should be exported.

If you are creating a document export from scratch, we suggest that you place the accession number and update date in the record element as attributes:

    <record AN=1234abc ED=2009-8-28>

If a system comprises multiple collections of data, each record must indicate which collection it comes from. This may either be implemented as an attribute of the record element, or as an element within the record:

    <record database="DBLABEL">

or

    <record><database>DBLABEL</database>

Copyright statements or other use restrictions should be explicitly presented in the export file. A <copyright> element in the container will serve as a default for the records in the file. Any copyright element within a record will replace the default, if present.

Some systems provide a URL with which the user may return to a

document on the server. A <recordURL> element within the record element may be used to provide this URL. Please present a fully-formed URL.

8.   Images

BizInt Smart Charts handles many sources with images (e.g. chemical structures or clipped images from patents).

Image files should be in a standard format (TIFF, GIF, JPG, PNG). The image files should be placed in the ZIP archive along with the XML file. Images may be stored with a path, but the path information is ignored. The file name for each image should be unique within the collection.

A field within the document should list the image file name. The path within the ZIP archive is optional.

Images should not have any whitespace or padding.

Images should be at a resolution suitable for printing at full size in a record. BizInt Smart Charts will scale the images as needed for the table. Using a high resolution image will result in much better quality display on screen and in print.

Currently only one image is supported for each record. However we are> working towards lifting this restriction. Until this restriction is lifted you may include multiple images per record but only the first image will be loaded.