



## Best Practices for Publishing RIXML

Alan Francis  
+1 (212) 849-4109

Version 0.2  
2 June 2009

### Executive Summary

RIXML provides a standard for comprehensive tagging of research content. This document assumes familiarity with RIXML (as documented at [www.rixml.org](http://www.rixml.org)). The purpose of this document is to identify best practices associated with implementing RIXML. It is intended to help contributors avoid common mistakes and to ultimately produce well-tagged research in conformity with the RIXML specification.

In this document, we use the term “contributors” broadly to refer to publishers of research. We use the term “vendors” to refer to firms that aggregate (i.e., collect) research content from multiple contributors. We refer to those who are the intended audience for the research content as “consumers”.

Although RIXML provides a framework for tagging a variety of research content (e.g., reports, estimates, models, events, etc.), this document only covers reports.

## Core Principles and Concepts

### Background

In this section we present what we believe to be some core principles and key concepts with regard to RIXML. When in doubt, parties on the producing or receiving end of RIXML should implement and interpret RIXML in the way that is most faithful to these principles and concepts.

#### 1. Write Once, Send Everywhere

This principle comprises the following:

- (1) A contributor should produce only one RIXML file for any given research report
- (2) This single RIXML file should contain all the necessary information to serve the needs of all intended recipients
- (3) No recipient may demand or expect customization on the part of contributors that would either render the RIXML file unfit for some recipients, render the RIXML file non-compliant with the RIXML schema, or require the contributor to create more than one version of the RIXML file

#### 2. Vendors Are Not Required to Support Publisher-Defined Values

While a contributor may include publisher-defined values for the benefit of some recipients or to support internal applications, contributors should not expect or demand that vendors incorporate these non-standard values into their products serving consumers.

#### 3. Tag in the Best Interest of Consumers

This principle comprises the following:

1. The goal of tagging is to help consumers filter reports to arrive at precisely the set of reports of interest to them – no more, no less
2. Any doubt about how to populate a tag should be resolved in the way that best serves consumers
3. Reports should not be over-tagged (i.e., tagged with values that aren't truly relevant to cause the report to undeservedly show up in more search results)

#### 4. Tagging Should Identify What a Report is About

- (1) Tag values should identify what a report is primarily about, not merely what it mentions. (e.g., if the report is about a merger, it should be tagged with the

Subject tag “MergerAcquisitionDivestiture”, but if it merely mentions a merger, this tag value should not be applied)

- (2) Tags should distinguish between what a report is primarily about versus what it is secondarily about. For example, a company report about XYZ Corporation discusses the ramifications of a big new contract with key parts supplier Acme Corporation. The report also includes a table of all of XYZ Corporation’s suppliers. XYZ Corporation should be tagged as a primary issuer because the report is primarily about it. Acme Corporation should be tagged as a non-primary issuer because a consumer looking for useful information about Acme Corporation (e.g., big new contract with XYZ Corporation) will find it in this report. The report should not be tagged with any of the issuers listed in the table of parts suppliers because the report contains no significant information about them.

## 5. Smart Version Control

The RIXML organization should make revisions to the specification few and far between. When revisions to the specification are made, they should be backwards-compatible (i.e., revisions should not make earlier RIXML instance files non-compliant). Lastly, contributors and vendors should upgrade to each new revision as soon as possible.

## Good RIXML is Built in Stages

### Step 1: Minimal RIXML

The first step is to produce a minimal RIXML file that contains the absolute minimum number of required tags.

### Step 2: Well-Formed XML

Every RIXML file is at its core an XML file. The ‘L’ in XML stands for “Language”. Every language has some syntax rules. Written English has syntax rules such as: the first letter of a sentence should be capitalized, an opening parenthesis should be matched by a closing parenthesis, and so on. XML also has a defined syntax. When an XML file conforms to the XML syntax rules, it is said to be “well-formed” XML.

Once you have created a minimal RIXML file, you must ensure that it is well-formed. The best way to do this is to use a software tool like XMLSpy ([www.altova.com](http://www.altova.com)). XMLSpy (or a comparable tool) will check your RIXML file and identify any syntax issues.

### Step 3: RIXML Compliance

Once you have determined that your RIXML file is well-formed (i.e., adheres to the rules of XML), you must ensure that it is also RIXML-compliant (i.e., it adheres to all the rules of RIXML). The best way to check for RIXML compliance is to use a software tool like XMLSpy ([www.altova.com](http://www.altova.com)).

RIXML-compliance is governed by the RIXML schema – a blueprint/template for what a proper RIXML file should contain. The RIXML schema governs such things as the names of the tags, the order in which they appear, which tags are optional and which are required, and the acceptable values for each tag. When you run your RIXML file through XMLSpy (or a comparable tool), it will identify any RIXML-compliance issues.

### Step 4: Populate as many tags as you can

Once you have determined that your minimal RIXML file is well-formed and RIXML-compliant, the next step is to grow it from minimally-tagged RIXML to fully-tagged RIXML. Your goal should be to completely populate the full range of tags relevant to research reports. As you add tags, be sure to re-test to assure that your file is still well-formed and RIXML-compliant.

## **Step 5: Q/A Your RIXML**

In the previous step, the goal was to completely populate the tags. The goal of the Q/A step is to ensure that you have also correctly populated the tags.

Let's say that you publish a report authored solely by John Smith. You must ensure that the RIXML file includes the tags for John Smith and for no other authors. There shouldn't be extra spaces before or after the name, there should be one and only one space between the first and last name, and of course any other information related to John Smith (e.g., phone, e-mail) should be correct.

Never assume that just because a file is certified as well-formed and RIXML-compliant that the data is actually correct!

## **Beware of Common Mistakes**

This section lists some common mistakes that contributors often make when first implementing RIXML.

### **RIXML file extension**

RIXML files should have an ‘.xml’ file extension (e.g., “report101.xml”), not ‘.rixml’.

### **Resource file names should match the actual file sent including case**

Make sure the resource file name identified in your <Resource><Name> tag matches exactly (including case) the name of the actual file sent.

### **Tag names are case-sensitive**

Tag names must be spelled exactly as documented (i.e., a tag named “Phone” can’t be spelled “phone” or “PHONE” – only “Phone” is acceptable).

### **Enumerated values are case-sensitive**

Enumerated tag values must be spelled exactly as documented (i.e., a tag whose value must be “Yes” or “No” can only have these values; “YES”, “yes”, “NO”, “no”, “Maybe” and “” are not acceptable).

### **Enumerated values are exclusive**

Enumerated tag values are exclusive – you can’t make up your own values (e.g., “RealEstate” for asset class) for tags that the RIXML specification provides a set list of values. The way to use your own values is via the enumeration value of “PublisherDefined” where applicable.

### **Bloomberg symbols**

Bloomberg symbols consist of a ticker followed by a single space followed by a country/exchange designation (e.g., “IBM US”). When using a Bloomberg symbol in RIXML, Bloomberg expects a ‘@’ instead of a space (e.g., “IBM@US”). Bloomberg codes are available from Bloomberg and should be exact (e.g., “IBM@US”, not “ibm@us”).

### **Reuters (RIC) symbols**

Reuters symbols (also known as RIC codes) consist of a ticker followed by a period followed by an exchange designation (e.g., “IBM.N”). The period and exchange designation may be omitted for U.S.-traded issue (e.g., “IBM”). Reuters symbols are available from

Reuters and should be exact (e.g., “IBM.N”, not “ibm.n” or “IBM N” or “IBM/N”) and may not omit the exchange designation unless a U.S. issue is intended.

## Filename re-use

File names should not be re-used unless the file being published is a replacement for a previously published file. If you publish a daily strategy report called “Strategy Daily”, don’t publish it each day in a file called “Strategy\_Daily.pdf”. Instead, use a unique name each day (e.g., “Strategy\_Daily\_20090512.pdf”, “Strategy\_Daily\_20090513.pdf”, etc.).

## Date/Time values

Date/Time tags, such as createDateTime and publicationDateTime, are very specific with regard to their format. The values you publish in your RIXML must conform to the ISO 8601 standard for “dateTime” values and should look like

**2007-11-05T13:15:30Z**

The year should be 4 digits. The month must be 2 digits (i.e., March must be specified as “03”, not ‘3’). The day of the month must be 2 digits (i.e., the 3<sup>rd</sup> day of the month is ‘03’, not ‘3’).

Your date/time values should include a time zone indicator, preferably Zulu (i.e., GMT). That doesn’t mean placing a ‘Z’ on the end of your local time – it means actually converting your local date/time to GMT. Although a date/time value lacking a time-zone indicator is valid, it is also useless. Vendors aggregate reports from contributors publishing in all time zones around the world. Vendors have no way of knowing what time a report was actually published if the date/time value is missing the time zone designation.

Don’t use a time of 24:00:00. The valid hours are 00 through 23, not 24.

Don’t omit the ‘T’ between the date and time.

Don’t add any other punctuation or spaces.

Don’t use a time of 00:00:00 as a default value when you know the date but don’t know what time a report was actually published. Make sure to identify and specify the actual time. Here’s why... Specifying 00:00:00 (midnight) is a good way to make your research invisible. When published at 7:00 in the morning, it won’t appear on anybody’s scrolling headlines because they’ll be looking at research published at 7:00 and your report will appear with the research published 7 hours earlier at midnight which has long since scrolled off the screen.

## **Tags are order-specific**

Tags must be placed in your RIXML file in exactly the same order as they are specified in the RIXML schema.

## **Required tags are required**

Be sure to include all required tags.

## **3-character language code**

RIXML uses the 3-character ISO standard for identifying languages. For example, the identifier for English is 'eng', not 'en' or 'English').

## **2-character country code**

RIXML uses the 2-character ISO standard for identifying countries. For example, the identifier for the United States is 'US', not 'USA' or 'U.S.' or 'United States').

## **Not escaping or double-escaping special XML characters**

There are several characters that have special meaning in XML. In English, placing a question mark or exclamation point in the middle of a sentence would likely confuse a reader. Likewise, characters like '&' and '<' are used for XML 'punctuation'. If you need to use such characters within your tag data (e.g., your title includes "Property & Casualty"), you need to 'escape' them. You do this by using '&amp;,' instead of '&' ("Property & Casualty" becomes "Property &amp; Casualty") and by using '&lt;,' instead of '<'.

Common mistakes are to fail to 'escape' these characters or to double-escape them (i.e., to convert an '&' that was already converted to '&amp;,' resulting in '&amp;amp;').

An alternative to escaping XML special characters is to enclose your data within a CDATA section which informs XML parsers that the text within the section is to be interpreted literally.

Conclusion: If you wish to code a tag like the following (which contains '&'):

```
<URL>http://research.xxx.com/control/control.asp?eventID=getDoc&docExt=pdf&ctbDocIDs=L2809027764</URL>
```

You can convert each "&" to "&amp;," which XML parsers will correctly interpret as the ampersand character, as in the following:

```
<URL>http://research.xxx.com/control/control.asp?eventID=getDoc&amp;docExt=pdf&amp;ctbDocIDs=L2809027764</URL>
```



Or, you can enclose your data within a CDATA section which informs an XML parser that the text within the section is to be interpreted literally, as in the following:

```
<URL><![CDATA[http://research.xxx.com/control/control.asp?eventID=getDoc&docExt=pdf&ctbDocIDs= L2809027764]]></URL>
```

## Best Practices

In this section, we list some best practices. Following these practices will ensure that your RIXML is well-formed, compliant with the RIXML schema and provides a set of comprehensive tags that are consistent with the intent of those who designed RIXML, consistent with the way your fellow contributors are implementing their RIXML, and will be most useful to your consumers.

### Perform all transactions via RIXML

Perform all transactions via RIXML. In other words, every new publication, revision, recall and delete should be achieved by sending a RIXML file. Never ask a vendor to manually perform any operation such as an emergency delete by phone or e-mail request. It will solve your immediate problem but circumventing the automated processing, controls, quality checks, and audit trails built into vendor ingestion systems will almost always result in future problems.

### Do NOT send a PDF file along with a RIXML delete request

New publications require transmitting a content file (typically in PDF format) along with a RIXML file. The same is true when you wish to replace the report with a revised version. Do not send a PDF when you are recalling or deleting a report.

### There is a difference between a missing tag and an empty tag

Some tags are required and some are optional. When a tag is optional and you don't wish to provide a value, do NOT include the tag with an empty value (e.g., type=""). Instead, omit the tag completely.

When a tag is required and you don't wish to provide a value, what you should do depends on the RIXML schema. If it is a tag that has a set of enumerated (i.e., pre-defined) choices like "Yes" or "No", you simply must include the tag with one of the acceptable values. If there is no associated set of enumerated values for that tag, then you must still include the tag but may leave the value empty.

### Author identification

Identifying authors has historically been difficult. You either had to make sure that each vendor had an up-to-date list of your authors and the codes you assigned them or that you had an up-to-date list of the codes that your vendors assigned them. You had to do this to ensure that when you provided an author code, your vendors could correctly identify the author from the provided code.

RIXML only provides one tag (personID) for author code. Thus, whatever value you put in that tag must satisfy every recipient vendor of your RIXML. Unless every vendor agrees to standardize on a single code (or RIXML is changed to accommodate multiple codes), you can not provide one code per author that serves all!

The good news is that both contributors and vendors have been getting smarter about identifying authors and technology has helped. Nowadays, every analyst has a unique e-mail address and most have a unique (or relatively unique) phone number. Thus, if you fill out the RIXML author tags well - providing first name, middle name, last name, phone number and e-mail address - most vendors should be able to properly identify the author without requiring any codes.

RIXML version 2.3 addresses this problem by adding a new PersonLabel element which enables you to provide multiple author codes to accommodate the author codes recognized by each vendor.

### **Issuer/security identification**

Most investment professionals use ticker symbols to identify issuers and securities. While tickers may work well for people, they work very poorly for data. The reason is simple – ticker symbols are not unique. Take ticker symbol ‘F’ – that could be Ford or Fiat. How about ‘C’? That’s Citigroup or bcMetals Corporation. And so on.

Tickers are unique when combined with an exchange. However, there is no universally accepted set of codes to identify exchanges. For these and other reasons, identifying issuers and securities by exchange tickers is the most error-prone method. RIXML version 2.3 alleviates this problem by adding the optional attribute tradingExchangeCode to the TradingExchange element with values to be taken from the ISO 10383 Market Identifier Code found at <http://www.iso15022.org/MIC/homepageMIC.htm>.

Your best bet is to identify issuers and securities in your RIXML using the most universally accepted identifiers – CUSIP, SEDOL, ISIN, Bloomberg, and RIC (Reuters) codes. There are 2 good reasons to include as many of these identifiers as you can – (1) not every vendor accepts every identifier type so including more identifier types ensures that you have all the vendors covered, and (2) in case one of your identifiers is incorrect or out-of-date, you will be covered by the others.

### **Encoding**

The concept of encoding is highly technical. We will try to explain it as simply as possible. Here goes...

Computers store, transmit and process information digitally utilizing electricity, magnetism and optics. Both numbers and text are represented by patterns of binary digits (i.e., bits) each having a value of 0 or 1. For example, the letter 'A' is represented in the standard ASCII system as '01000001'. As computer usage grew globally, the basic ASCII system which only standardized the basic letters, numbers and punctuation (e.g., A-Z, a-z, 0-9, !@#\$%^&\*()+) became increasingly inadequate for global use. It did not include binary patterns to represent accented letters, Asian characters, Euro and Yen symbols, and so on.

A new standard known as UTF (Unicode Transformation Format) grew from the Internet age. The UTF standard provides standardized binary patterns encompassing the characters used in the various languages around the world. The most commonly used variant of UTF is known as UTF-8. If you send a RIXML file where each character is represented by the binary pattern for that character as specified by the UTF-8 standard, then your RIXML file is said to be encoded in UTF-8.

In order to accommodate global research content, vendors are standardizing around UTF-8. The first line in a RIXML file (or any XML file) identifies the encoding used. XML files encoded in UTF will begin with the following line:

```
<?xml version="1.0" encoding="UTF-8"?>
```

Vendors will want contributors to encode their RIXML files in UTF-8. Unfortunately, Windows does not encode in UTF-8 but rather in an encoding standard known as Windows-1252. If you copy and paste text from a Word document into a RIXML file, you may introduce characters encoded in Windows-1252 into your RIXML document which you have tagged as being encoded in UTF-8. The result is that some characters will appear as gibberish on the vendor system. There are two solutions to this problem. The first is to convert all text into UTF-8 before placing it in your RIXML file. The other is to specify that your RIXML file is actually encoded using Windows encoding by starting your RIXML file with

```
<?xml version="1.0" encoding="windows-1252" ?>
```

Another encoding scheme used for Japanese language content is known as Shift-JIS. Without elaborating further, the bottom line is that the first line in your RIXML file must correctly identify the encoding scheme in use, and that if you mix encodings within your RIXML or identify the encoding scheme incorrectly, you can expect tags like title, synopsis and abstract to be displayed on vendor systems either containing or consisting entirely of gibberish characters.

## Formatting phone numbers

There is no universally accepted standard for formatting phone numbers. For example, the phone number 555-1212 in New York can be expressed in myriad ways including

+1 (212) 555-1212

1.212.555.1212

1-212-555-1212

The closest thing to a standard is ‘+’ followed by the country/region code followed by the area/city code (where applicable) enclosed in parentheses followed by the local phone number. The aforementioned may be punctuated by spaces, periods or dashes. Of the 3 formats illustrated above, only the first meets the standard just described (because it is the only one that begins with ‘+’ and surrounds the area/city code with parentheses).

RIXML divides the phone number into CountryCode and Number tags. It does not specify whether you should include the ‘+’ in the CountryCode value. Some contributors divide the full number between the CountryCode and Number tags. Others place the entire number in the Number tag and leave the CountryCode tag empty.

What vendors typically do is slap together whatever you put in these two tags. That’s why you should never put the country code into both the CountryCode and Number tags.

Although it tends to run against the implied intent of the framers of RIXML (which is to separate the country code from the rest of the phone number), the simplest for both contributors and vendors is probably to place the entire number in the Number tag and leave the CountryCode tag empty. You should format the phone number like the first of the three sample phone numbers above is formatted using all spaces, all periods or all dashes to punctuate the local number resulting in one of the following:

+1 (212) 555-1212

+1 (212) 555.1212

+1 (212) 555 1212

## Default values

Contributors often encounter a classic dilemma. They use an automated authorship system. Their system may not collect all the information about each report that is necessary to fully populate all the RIXML tags; or it may use a non-RIXML tagging scheme which doesn’t map exactly to the RIXML tagging scheme; or the contributor may rely on analysts to fill in the tags (but unfortunately analysts tend to be unreliable when expected to perform such tasks).

The obvious solution is to use a default value. So let's say a contributor wants to populate the rating action and estimate action tags. They determine that the most common value is "Reiterate". They therefore decide to automatically fill the rating action and estimate action for every security in every report with "Reiterate" unless the analyst changes it.

Is this a good solution? The answer is No! From a data perspective, the defect rate (the percentage of incorrect values) will probably exceed 50%. This is the data equivalent of publishing a research report with every other word misspelled! From a consumer's perspective, numerous reports will falsely turn up in their search results. When consumers notice a pattern of a particular contributor's reports constantly turning up falsely in their search results, they learn to ignore that contributor's reports and never open them (even when they turn up legitimately in search results), or worse - they actively exclude that contributor's reports in their search filters.

The bottom line is that you should never use default values unless you can intelligently apply them. For example, if you only publish equity reports, you can safely automate every report to be tagged with asset class "Equity". Or, if all of your company reports are written by fundamental equity analysts, you can safely automate every company report to be tagged with research approach "Fundamental". Or, if every report authored by a particular analyst is an economics report, you can safely automate every report drafted by that analyst to be tagged with discipline type "Economics".

### **Publisher-defined values**

The framers of the RIXML standard recognized that there would be times when the standard values provided by the schema would not fully meet user needs. They therefore left room for RIXML users to provide their own custom ("publisher-defined") values in a way that remains RIXML-compliant.

When it comes to use of publisher-defined values, the popular adage comes to mind – "Just because you can, doesn't mean you should!" Customization runs directly counter to the goal of standardization! There are 2 legitimate instances where publisher-defined values should be used – (1) to support internal use of RIXML, and (2) to support values that are expected by particular vendors such as an organization ID or entitlement code required by a particular vendor.

Other than these two instances, your published-defined values will generally be ignored by vendors and consumers alike, taking up space and time yielding no benefit.

What is particularly interesting is that much real-life usage of publisher-defined values could have and should have been coded within the RIXML schema. We see instances like a publisher-defined Subject code of "Initiate Coverage" for which there is a perfect good

RIXML coverage action tag to represent an initiation of coverage. We see instances for which a contributor uses a publisher-defined region of “NorthAmerica” when RIXML already includes a region value of “NorthAmerica”.

Bottom line: Only use publisher-defined values to support internal use of RIXML or to support values that are expected or required by particular vendors.

### **Don't embed HTML in Synopsis and Abstract**

Some contributors embed HTML markup in their Synopsis and Abstract tags in an attempt to control the display formatting of the text on vendor systems. HTML encoding can sometimes impact the well-formedness of XML files and will look like gibberish when displayed on platforms that don't support the HTML markup.

## Classic Problems Addressed by RIXML 2.3

In this section, we list some classic issues and omissions that are addressed in RIXML version 2.3.

### No way to identify a ticker symbol

One can specify that an Issuer/Security identifier value represent a CUSIP, SEDOL, ISIN, Bloomberg, RIC, etc. but there is no choice so specify that it represents an exchange ticker. RIXML 2.3 addresses this oversight by adding the value ExchangeTicker to the list of possible values for the IssuerSecurityIDTypeEnum enumeration.

### No way to provide a standard exchange code

We have noted previously that while tickers are not unique, they are unique when combined with an exchange. In addition to providing a means to provide an exchange ticker, RIXML 2.3 adds the ability to associate the ticker with an exchange by placing a standardized ISO 10383 Market Identifier Code (<http://www.iso15022.org/MIC/homepageMIC.htm>) in a new optional attribute tradingExchangeCode in the TradingExchange element.

### No way to omit the OrganizationID

OrganizationID is currently a required element. Since it is not needed when a RIXML file is used internally or when sent to a destination that does not require OrganizationID to identify the organization, RIXML 2.3 renders this element optional.

### "Reiterate" missing from the WeightingActionEnum and TargetPriceEnum enumerated lists.

RIXML 2.3 adds "Reiterate" to the enumerated choices for the WeightingActionEnum and TargetPriceEnum enumerated lists.

### No counterpart to rating action "Rating Withdrawn"

A rating action of "RatingWithdrawn" is provided in the RatingActionEnum enumerated list but no corresponding value is provided for when a rating is restored. RIXML 2.3 adds a "RatingRestored" value to the RatingActionEnum enumerated list.

### No recognition for some widely used region designations

In the world of research, there is actually a region hierarchy. The top level - Americas, EMEA and AsiaPacific - comprises the lower-level regions as follows:

Americas = NorthAmerica + LatinAmerica

EMEA = Europe + MiddleEast + Africa



AsiaPacific = AsiaExJapan + Japan + Australasia

RIXML 2.3 adds Americas, EMEA and AsiaPacific to the RegionTypeEnum enumerated list. It also adds Oceania and Caribbean.

### **No way to associate an entitlement code with a particular vendor**

Vendor-specific entitlement codes have typically been passed using publisher-defined audience type codes. To facilitate this usage and ensure that codes intended for one vendor don't get inadvertently applied by a different vendor, RIXML 2.3 adds an optional attribute entitlementContext to the existing AudienceTypeEntitlement element.

### **Provide better support for identifying people**

The means to identify a person (e.g., an analyst) is via the personID attribute. As an attribute (as opposed to an element), only one value can be provided. A contributor is thus unable to provide the ID by which Vendor A recognizes the person and also the ID by which Vendor B recognizes the person and so on. RIXML version 2.3 addresses this problem by adding a new PersonLabel element which enables you to provide multiple author codes.

### **No way to specify more than one estimate action**

The estimateAction tag is an attribute of the RIXML Security element. As an attribute, only one estimate action can be specified for a security. However, it's legitimate (and common) for more than one estimate action to apply to a security. A typical case is where the analyst reiterates his current fiscal year estimate and increases his next fiscal year estimate (i.e., two actions - Reiterate and Upgrade). RIXML 2.3 adds the optional attribute estimateAction to the FinancialValue element enabling multiple estimate actions.

### **Problematic periodicalIndicator**

The periodicalIndicator tag has been a required attribute of the popular RIXML ProductDetails element. Since contributors want to provide the important tags within the ProductDetails element, they have little choice but to include and populate the periodicalIndicator tag. Since the value of the periodicalIndicator tag can only be "Yes" or "No", what happens when a contributor doesn't have accurate information about whether their reports are periodicals or not? They tend to either set periodicalIndicator to "Yes" for all reports or "No" for all reports. Either way, that yields an incorrect tag about half the time – an unacceptable defect rate! RIXML 2.3 makes the periodicalIndicator attribute optional.

## **Not Every Report Has a Focus**

ProductFocus is another required tag with a fixed set of enumerated values. The RIXML specification assumed that the focus of every report can be selected from the fixed set of enumerated values. But that isn't so. Take for example a morning call report. It's not about any single issuer, industry, country, region, etc. – it's an unfocused report. RIXML 2.3 renders the ProductFocus tag optional and adds the value "Unfocused" to the enumerated list to identify unfocused reports or reports not easily categorized by the other focus choices.

## **The Enumerated List for Indexes**

IndexEnum represents the list of codes to identify an index discussed in a report. The list has 2 key problems – (1) the list is far from complete, and (2) many of the values are vague (i.e., the intended index is not clear from the mnemonic code) or out-of-date (e.g., the index no longer exists). RIXML 2.3 enables contributors to identify any index via CUSIP or SEDOL rather than from an enumerated list.

## **Support for Selling Research**

In order to support the sale of research, RIXML 2.3 adds a Price tag to the Resource element.

## **Upgrade the RIXML User Guide**

The RIXML 2.3 User Guide adds new clear definitions for enumerated values.

## Best Practices for Populating the RIXML Tags

### Background

The RIXML specification leaves room for differing interpretations. As a result, there is a bit of variance on how certain tags are being populated by RIXML contributors. In this section, we discuss how we believe certain tags should be populated and why.

### Title

There is great inconsistency in how contributors construct their titles/sub-titles. Some contributors believe that considering the huge volume of reports, the best way to get their report read is to use catchy titles. So they use cryptic but catchy titles (e.g., “A Bird in the Hand is Worth Two in the Bush!”). This approach is reasonable in theory but works poorly in practice.

The goal is not to get your report opened by as many consumers as possible – it’s to get your report opened by those most interested in its contents. You want to be rewarded by consumers who can easily find the research that matters to them. You don’t want your consumers to feel spammed by your research. Getting your report opened by those not truly interested in its contents only conditions them not to bother opening your reports in the future.

Consumers are mining for nuggets of new useful information. They don’t have the time to open every report looking for those nuggets. If the headline does not clearly identify what the report is about, they will skip it.

The best titles identify 3 things – (1) the focus, (2) the specific item of focus, (3) the thrust of the report. For example, a company report on XYZ Corp. might have a title like “Company Update – XYZ Corp. – Downgraded on Worse-than-Expected Results”. It’s not catchy but it tells the reader that the report is about a company, which company, and the key thrust of the report. This title will attract interested consumers whereas a catchy but vague title will be passed over.

### Primary versus non-Primary Issuers

There is a single purpose in identifying the primary issuer, namely to identify the single issuer/security that a report is about. For example, a company report about Toyota may also mention other companies that are competitors, customers or suppliers to Toyota. You may also want to identify all such companies with significant mention in the report, however, only the company that the report is about (i.e., Toyota in our example) should be listed as a primary issuer. The other companies with significant mention should be included but not marked as primary.

If a company merely appears in a table or list within a report, do not include it as an issuer. Mere mention of a company in a report (e.g., “XYZ Corp. is one of the largest employers in the state”) also does not warrant inclusion as an issuer.

Generally, a report should not have more than one company listed as a primary issuer. An example of a rare exception would be where a report is about the merger between two companies and is equally about both companies. Sector/Industry reports should generally not have any companies listed as primary issuers (because the report is not primarily about any one of the companies mentioned in the report). Likewise, a compilation report such as a morning call report should never have any companies listed as primary issuers.

Although the foregoing discussion refers only to company reports, it also applies to other issuers/securities such as REITs, mutual funds, and preferred stocks (e.g., if a report is about a single mutual fund, that mutual fund should be tagged as the primary issuer).

#### **Which Analysts to Include in the RIXML File**

Research consumers develop preferences for reports drafted by particular analysts. Good author tagging lets users seek out your research by author and help your analysts develop a following.

A primary author is one who primarily drafted the report. Publishers of research often want to associate multiple analysts with a report such as the member of a team. Some reports also list salespeople, traders, editors or research assistants.

Only the analyst(s) who actually authored a report should be listed as primary authors. Other analysts who have contributed can be included but should not be marked as primary. Other team members who did not contribute or editors or research assistants should not be included. Including such secondary people adds little value to consumers (because they use the author info to seek out content drafted by analysts they follow) and creates clutter.

There is no reason to send analyst tags populated with department names like “Global Equity Research” or “Latin America Strategy Team”. Such information provides no value for consumers and only clutters screens.

Compilation reports, such as a morning call report, typically encapsulate the work of many unrelated authors, none of which can be deemed to have primarily authored the entire report. On such reports, contributing analysts should not be marked primary.

#### **Issuer Focus Versus Sector/Industry Focus**

By definition, an industry is a collection of companies in the same business. When a report is primarily "about" one or perhaps two companies, it should be considered a company

report with focus set to “Issuer”. When a report is about a group of companies in the same business (but not primarily about one or two companies in that group), it should be considered an industry report with focus set to “SectorIndustry”.

There is no focus value appropriate for morning call (and similar) reports because they are not about a company or a sector/industry (because the companies discussed range randomly across varying unrelated businesses). Until RIXML.org develops a good way to set the focus for morning call reports, the industry practice has been to set the focus for morning call reports to “SectorIndustry”.

### **When to set primaryIndicator to “Yes” and when to set it to “No”**

The primaryIndicator attribute is used throughout RIXML. Contributors need greater clarity in understanding when to set primaryIndicator to “Yes” and when to set it to “No”. Contributors seem to take one of 3 approaches:

- (1) The entity that a report is primarily about is marked as primary. For example, if a report is about an issuer, then that issuer is marked as primary. All other instances of the primaryIndicator attribute in all product classification elements are set to “No”.
- (2) The entity that a report is primarily about is marked as primary and all element instances related to the primary entity are also marked as primary. All other instances of the primaryIndicator attribute in all product classification elements are set to “No”.
- (3) Chaos characterized by setting all primaryIndicator attributes to “Yes”, all to “No” or some combination thereof unsupported by any consistent methodology.

Under the first approach, a report about IBM would have issuer/security IBM marked as primary. Since the report is not about a sector/industry, region, country, etc., all sector/industry, region, country, etc. elements would be marked non-primary.

Under the second approach, a report about IBM would have issuer and security IBM marked as primary. Elements identifying IBM’s sector/industry, region, country, etc. should be included and marked primary. Any other sector/industry, region, country, etc. elements should be marked non-primary.

It should be noted that the second approach seems much more in sync with the RIXML documentation and sample use cases. We favor this approach because it would seem to serve consumers better.

Based on the foregoing, we have developed some rules that we believe serve consumers best. We would like to see these rules clearly documented and encouraged.

1. Company reports (and reports about a single mutual fund, REIT, etc.) should have one and only one issuer/security marked as primary. (Rare exceptions might include a report discussing the merger of Company A and Company B where the report is truly primarily about two companies.) Elements identifying the primary company's sector/industry, region, country, etc. should be included and marked primary. Other companies (e.g., competitors, suppliers, customers, vendors) significantly discussed in the report in the report should be included in the RIXML file but should not be flagged as primary. Companies that appear only in tables should not be included in the RIXML file.
2. Sector/Industry reports should not have any issuer/security marked as primary (because if the report is primarily about an issuer/security, then the report is a company report – not a sector/industry report). All companies significantly discussed in the text of the report should be included in the RIXML file but should not be flagged as primary. Other companies in the sector/industry but not significantly discussed in the report or only appearing in tables should not be included in the RIXML file.
3. Morning call / morning note (and similar compilation) reports are not company reports because they are not primarily about a single company. They are also not sector/industry reports because an industry is a set of companies in the same business and morning call reports cover a random array of companies across unrelated businesses. Thus morning call reports should not have any companies or industries tagged as primary.

### **Synopsis**

The Synopsis is a very brief overview of the topics covered in the report. This should be no longer than a few lines and would be appropriate for use in a highly abbreviated display such as when the user's mouse is positioned over the report's headline.

### **Abstract**

The Abstract tag should highlight the key points covered in the report and should be no longer than a few paragraphs or bullets. If your report begins with a summary, highlights, or key points section, the simplest way to provide a report summary is to copy and paste this section of your report into the Abstract tag.

## Language

The Language tag is used to identify the language in which the report is primarily written.

## Product Focus

The Product Focus identifies what the report is primarily about. The Product Focus tag choices are Issuer, Sector/Industry, Region, Country, Discipline, Asset Class, Asset Type, Security Type, Index, and Exchange. Thus, the Product Focus is an indication of whether the report is primarily about a company, an industry, a region, a country, etc.

The most common type of report is a single-company report. Tag single-company reports with a Product Focus of “Issuer.” The next most common type of report is a sector/industry report, which you should tag with a Product Focus of “Sector/Industry”.

How do you tag a report that seems to have multiple focuses such as a report on the U.S. Auto industry? Is the focus on a country – U.S., or is the focus on an industry – Autos, or perhaps is there a dual focus on country and industry?

One may make the argument that the primary focus is on the Auto industry with a secondary focus on a country. Such an interpretation would call for including a Product Focus tag of SectorIndustry marked as primary and also including a Product Focus tag of Country marked as non-primary.

Our view at FactSet is that such an interpretation merely confuses the geographic scope of a report with the Product Focus. We view a report on the U.S. auto industry as an industry report, not a country report (i.e., the report is focused on an industry and is merely limited in scope to a country but the report is not focused on the country). Thus, we would only tag it with a Product Focus of SectorIndustry and not with Country. Likewise, we would tag a report on U.S. Equities with a Product Focus of Asset Class.

Here are some additional guidelines...

Tag reports about the markets in a region with a Product Focus of “Region.” Likewise, reports about the markets in a country should be tagged with a Product Focus of “Country.”

Tag strategy, economics, market analysis, quantitative and technical analysis reports with a Product Focus of “Discipline.”

Tag reports that broadly cover an asset class (e.g., Equities, Fixed Income, Currencies, Commodities) with a Product Focus of “Asset Class.” Such reports typically bear titles like “U.S. Equity Update” or “Fixed Income Daily.”

Tag reports that are focused on an asset type (i.e., a category that is less broad than an asset class but encompasses multiple security types) with a Product Focus of “Asset Type.” Such reports typically bear titles like “Derivatives Update” or “Muni Monthly.”

Tag reports that are focused on a security type (e.g., bonds, options) with a Product Focus of “Security Type.” Such reports typically bear titles like “Bond Market Update” or “Futures Monthly”.

### **Sector/Industry**

To understand how to best provide sector/industry tagging, it is important to understand how consumers use the Sector/Industry tag.

Consumers interested in a particular industry will be interested in (a) company reports on companies in that industry, and (b) industry reports on that industry. Therefore, only company and sector/industry reports should be tagged with a sector/industry.

Tag a company report with the sector/industry that the company belongs to. An industry/sector report should be tagged with the sector/industry that the companies discussed in the report belong to. You may have to tag a sector/industry report with more than one sector/industry when the companies discussed span multiple industries. An example might be a report that discusses both life/health insurance companies and property/casualty insurance companies.

### **Subject**

The Subject tag identifies the topic(s) primarily discussed in the report. It is often an indication of an event that triggered the writing of the report (e.g., a company is about to release its earnings; a merger between two companies is announced; a company is planning to redeem its bonds; the CFO of a company resigns, etc.).

Note that mere mention of a subject such as an accounting change or an acquisition does not justify applying its subject code to the report. Apply a subject tag only when that subject is a primary topic of the report.

More than one subject tag can be applied to a report when applicable.

### **Periodicity**

A periodical report is one that is published at a regular fixed interval. For example, a “Tech Weekly” report should be tagged with a Periodicity of “Weekly”. The “Quarterly Coal Review” should be tagged with a Periodicity of “Quarterly.”



A common tagging mistake is to tag company reports as “Daily” because they are published every day. Company reports are published in reaction to events (e.g., results announcement, management change, merger/acquisition, rating change, estimate revision). Although company reports as a class of reports are published each day, a company report on any given company is not published at a fixed time interval but only when a reason to publish arises. A true periodical is always published at the same regular interval. Thus, tag a company report as a periodical only when it is truly published at a regular fixed interval (e.g., “This week at General Motors”) with no relation to any particular event.

Consumers seeking actionable research will often want to filter out periodical reports. Other consumers who want to keep on top of an area of interest but have no need to read every research note when published will seek out periodicals. Likewise, consumers seeking overview, reference and educational reports will seek out periodicals. Proper Periodicity tagging will help you target your various report types to the appropriate audience of consumers who are seeking them.

### **Product Classifications – Asset Class**

The Asset Class classification refers to the broad category of asset that is discussed in the research report.

If a research report discusses the stock of a particular company or industry, or if it discusses how the stock market in a particular country/region is doing in general, then classify the report with asset class Equity.

If a research report discusses fixed income (i.e., securities paying a fixed rate of interest or dividends per period) financial instruments (e.g., bonds, notes, CDs, treasury securities, etc.), or the fixed income market in general, then classify the report with asset class Fixed Income.

If a research report discusses commodity securities (e.g., gold futures, oil futures, steel forward contracts, options on silver, etc.) or a commodity market in general, then classify the report with asset class Commodity.

Note: You should be careful not to confuse a commodity stock report, such as report analyzing the common stock of a gold mining company, with a commodity report discussing gold futures. If the report discusses the stock of a gold mining company, it should be classified with asset class Equity. Classify a report with asset class Commodity only if it discusses the commodity market in general (e.g., the report discusses factors that affect the market price of gold) or if it discusses/analyzes commodity securities (i.e., financial instruments which have a commodity as an underlying asset, e.g., gold futures).

If a research report discusses currency securities (e.g., USD futures, currency swaps, EUR options, etc.) or the currency market in general, then classify the report with asset class Currency.

Note: Not every research report should have an Asset Class classification (e.g., economic reports are generally not about any particular asset class).

### **Product Classifications – Discipline**

There are three RIXML discipline types: Investment, Strategy, and Economics. The Discipline tag value should identify the stage of top-down investment analysis reflected in the report.

A top-down analysis begins with analyzing the investment climate and economic conditions from a global, regional, or country perspective. This includes looking into general environmental factors and macroeconomic trends that impact the market, such as the general health of the economy as characterized by GDP (gross domestic product), inflation rates, interest rates, unemployment, wars, geopolitical tensions, etc. Research reports that relate to this first stage of top-down analysis are classified with Discipline type Economics.

The next step in top-down analysis is to compare different groups of investment vehicles with respect to risk and return to determine relative attractiveness. Groups that may be compared include markets (e.g., U.S. versus European versus Asian), industries (e.g., Semiconductors versus Information Technology Services), asset classes (e.g., equity versus fixed income), asset types (e.g., derivatives versus emerging markets credit instruments), security types (e.g., common stocks versus bonds), capitalization (small-cap stocks versus large-cap stocks), and maturities (short versus long maturities). Reports that focus on comparing groups of investment vehicles are classified with Discipline type Strategy.

The last step in top-down analysis is to select specific stocks, bonds, or other investment vehicles from the groups identified in the previous stage that will maximize investors' risk-adjusted return. Research reports that provide recommendations (i.e., buy, sell, etc.) for specific investment vehicles (e.g., sell Amazon's 6.25 % bonds maturing in 2007; buy IBM's common stock, etc.) are classified with Discipline type Investment.

### **Product Classifications – Discipline – Research Approach**

Research Approach answers the questions: what type of research technique did the analyst use in the research report; did the analyst use fundamental analysis, quantitative analysis, or technical analysis?

Fundamental analysis is a valuation method that uses financial analysis to determine the investment potential of a security. Fundamental analysis focuses on factors that relate to a

company's business, products, management, financial condition, etc. A research report classified with research approach Fundamental will usually analyze a company's financial statements, profits, sales, as well as non-financial information such as demand for products, new product releases, the quality of the company's management, how well the company is positioned in its industry, etc.

Note: Most single-company and sector/industry reports employ fundamental analysis.

Quantitative analysis is a valuation method that purely uses computational and statistical techniques to evaluate the attractiveness of financial instruments. It is an approach that ignores non-quantified information such as qualitative factors (e.g., management, products, business environment), events or market movements.

Technical analysis is a valuation technique used to predict securities' future performance by looking solely at raw historical market metrics (e.g., security price, trading volume) and derivative values (e.g., moving average, relative strength, on-balance volume). Technical analysts examine peaks, bottoms, moving averages, trends, patterns and other technical indicators that help predict future price movements.

### **Weighting Action**

FactSet views "weighting" as an asset allocation term which answers the question "Should a portfolio have more/less/the same of this industry/country/asset class relative to its weight in the market?"

Analysts publish weightings on industries, countries, regions, and asset classes (usually stocks versus bonds) to provide portfolio managers with guidance on asset allocation (e.g., which industry should be over-weighted in the portfolio, which country should be under-weighted in the portfolio, what percent of the portfolio should be in bonds versus stocks, etc.).

Analysts most often weight industries. Common weightings include Overweight, Underweight, Neutral/Equal/Market Weight, etc. Any action with respect to an analyst's weighting on an industry, sector, country, region, or asset class represents a Weighting Action.

Note: Although analysts frequently use terms like overweight, underweight, and market-weight with respect to individual stocks, these terms in that context are more akin to buy/neutral/sell ratings and are not true weightings in the asset allocation sense. Changes to company "weights" should be considered rating actions and not weighting actions even though they are expressed using terms like "Overweight" or "Underweight".

## How to Code Vendor-Specific Entitlement Codes

Most vendors have a way of directing a report to a specific audience (which may be referred to as a user group or distribution group). The audience is assigned a text or numeric code (e.g., NON\_US or 12). Each such code should appear within its own distinct AudienceTypeEntitlement element within its own distinct Entitlement element within its own distinct EntitlementGroup element as follows:

```
<EntitlementGroup>
  <Entitlement includeExcludeIndicator="Include" primaryIndicator="Yes">
    <AudienceTypeEntitlement
audienceType="PublisherDefined">BBGCODE1</AudienceTypeEntitlement>
  </Entitlement>
</EntitlementGroup>
<EntitlementGroup>
  <Entitlement includeExcludeIndicator="Include" primaryIndicator="Yes">
    <AudienceTypeEntitlement
audienceType="PublisherDefined">REUTERS5</AudienceTypeEntitlement>
  </Entitlement>
</EntitlementGroup>
```

Note that RIXML 2.3 enables you to associate entitlement codes with a particular vendor.

## Enumerated Tag Value Definitions

The previous version of this document included comprehensive documentation on what the enumerated tag values mean. This documentation has now been incorporated into the RIXML 2.3 Users Guide which can be found at [www.rixml.org](http://www.rixml.org).