RIXML·org
Research Information Exchange Markup Language

Getting Started *Guide*

An overview of RIXML and how it can benefit sell-side firms, buy-side firms, research distribution vendors, and anyone else involved in the investment research distribution chain.

Executive Summary

A Shared Problem

RIXML was formed when a number of buy- and sell-side firms came together to discuss a common problem: the amount of investment research that was being distributed had outgrown the tools available to organize it.

The consumers of investment research needed better ways to access the precise research they needed, and the creators of investment research needed a better way to ensure that their research was seen by those who would benefit from it.

When investment research reports first began to be distributed electronically, adding a few simple tags like title, author, date, and ticker provided enough information for a user to determine whether the report would be useful to them. Now, however, the nature of electronic research report distribution has evolved. How?

- **Information overload** – millions of research reports are distributed each year.
- **More distribution channels** – originally, reports were distributed via email and through one or two major vendors. Now, there are many more vendors distributing research, and many buy-side firms have their own data repositories as well.
- **More sophisticated distribution channels** – consumers are demanding the ability to be alerted in a more sophisticated manner, to search in a more targeted way, and to receive just the reports that fit their needs.

Unfortunately, the tools used to create and distribute research had not evolved to handle these changes. Thus, buy- and sell-side firms, along with vendors who provide products and services for distributing investment research, created RIXML.org, a consortium whose goal is to define an open protocol that will improve the process of categorizing, aggregating, comparing, sorting, searching, and distributing global financial research.

The primary objective of the RIXML specification is to provide extensive capabilities for tagging any piece of financial research content, in any form or media, with sufficient metadata information to allow research users to search, sort, and filter the published research and to provide highly relevant information to decision-makers.

INSIDE

What is RIXML?

RIXML is a **set of tags** used to describe a research document or other research product, formed into an **XML specification,** along with a set of **enumeration lists** that provide the terms that are allowed to be used for certain tags.

RIXML IS:

✓ the tagging (metadata) that describes a document, not the document itself

✓ a set of components, rules, lists, and recommendations that enable clear description of investment research and other investment data

✓ a standardization of terms used to describe certain aspects of research, such as security type, region, etc.

RIXML IS NOT:

✗ a search engine

✗ a data repository

✗ a subject thesaurus

✗ a piece of software

✗ a definition of format or content of documents (resources) themselves

✗ a standardization of the types or formats of resources (documents, audio files, etc.) allowed

✗ a standard to define how data is to be transmitted

This document provides an overview of RIXML, and the rules, lists, and protocols that make it up. See the *Specification* section of the RIXML.org website for additional documentation, including the full user guide, the specification document, and the full schema.

About RIXML.org

RIXML.org is a consortium of buy- and sell-side financial services firms that are committed to improving the process of electronic research distribution by creating an open industry standard. The goal of RIXML.org is to define an open protocol that will improve the process of categorizing, aggregating, comparing, sorting, searching, and distributing global financial research.

Who can benefit?

The sheer magnitude of investment research available today means that a sophisticated method of describing this information is essential.

Regardless of where you are in the information chain, *you* need a better way!

Content Producers need a way to ensure that content is seen by their target audience

▸ RIXML's sophisticated structure enables content producers to provide exact tagging that better describes the content, which makes it more likely that it will be seen by those most likely to find it valuable.

Content Aggregators need to be able to deliver value-added products and services

▸ RIXML's standardization of tags enables development of sophisticated tools and services to address a variety of needs.

Content Consumers need a way to filter content so that only relevant content is seen

▸ RIXML's rich tagging expedites searching and accessing content, reduces information overload, increases efficiency, and improves access to research by standardizing sorting and filtering criteria.

How will I use RIXML?

A sell-side firm that adopts RIXML:

- Will use RIXML tagging in the metadata of the research they distribute to third-party data aggregators and directly to clients
- May use RIXML tagging in their internal research-creation process

A 3rd-party data aggregation firm that adopts RIXML:

- Will have a mechanism to accept research reports or other content with RIXML-tagged metadata
- Should provide an interface that allows end users to search using the enhanced metadata
- May have a mechanism to assist non-RIXML-compliant firms to add RIXML-based metadata to the content they contribute to the aggregator
- Is free to develop products and tools utilizing RIXML

A buy-side firm that adopts RIXML:

- Will benefit from enhanced search capabilities provided by 3rd-party data aggregators
- Will use RIXML-based feeds to receive metadata and research directly from sell-side firms
- May use RIXML tagging in their internal research-creation process

A More Efficient Model

RIXML can help you to stop reinventing the wheel. How? The value of research comes from the *information in the research report*, not from the tags used to describe that research. When a firm needs to use an entirely different methodology to submit research to the various vendors who distribute it, it wastes time and energy.

Because this specification was written by buy-side and sell-side firms, along with vendors who distribute investment research, it addresses many of the problems that we each face.

Let's say you are a large research firm, and you want to begin distributing your research through a new research aggregator. Your IT group will need to get the specifications from this research aggregator, then they will need to develop a mechanism to transmit the required information from your infrastructure to the research aggregator.


Research firm → Research vendor
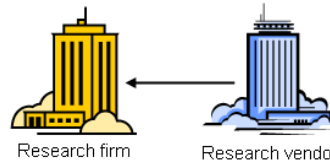
Then, one of your clients, a large mutual fund company, decides that they would like to begin receiving a direct feed of your research models, so that they can make them available on their internal intranet. Again, your IT group will need to work with their IT group to develop the tags and the protocols required to accomplish this.

Mutual fund firm ↔ Research firm
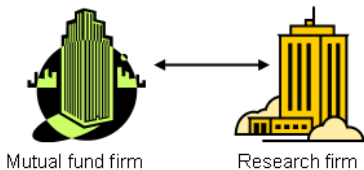
Now, let's assume that your firm eventually has to go through this process with a number of research aggregators, and several individual clients.

Every time one of these firms introduces an enhancement or wants to add a new tag, your IT group must devote time and resources to implement these changes.

Mutual fund firm 1, Mutual fund firm 2, Hedge fund → Research firm → Research vendor 1, Research vendor 2, Research vendor 3

Finally, extrapolate this out to all of the research firms, all of the research consumers, and all of the research distribution firms involved in the investment research universe. As you can see, a huge amount of time, effort, and money is being spent on the routine task of describing research content.
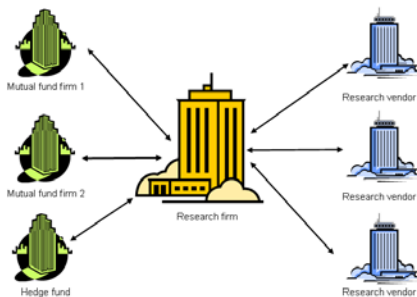
RIXML.org was founded to develop a way to make this process more efficient.

5 ways RIXML helps the research distribution process

1 Standardized specification reduces the cost of distributing research

2 Better tagging helps consumers find the research they need

3 Sophisticated tagging options help ensure research gets to the right consumers

4 Enumeration lists group similar content together for efficient browsing and searching

5 Clearly-defined rules enable consistent description of investment research and other investment data

Our Approach

As a first step, all of the participants of the consortium provided a set of real-life scenarios — use cases — in which a standard like RIXML would help them. Buy-side participants said that they would be like to be to able to search, sort, and filter information published by the research providers, to provide relevant research to their decision-makers in an easy-to-use format, and to create tools to help them commingle research from different sources.

Sell-side firms stated that they would like to be able to focus on developing and producing content, and to dedicate fewer resources to the technical issues around delivering it.

The resulting list of requirements included a listing of the various types of research, such as:

- ‣ Single-company reports
- ‣ Morning notes
- ‣ IPO reports
- ‣ Macroeconomic research
- ‣ Sector, industry, and country overviews
- ‣ Compilations of several company reports
- ‣ Weekly, monthly, and quarterly publications

It also included the various tasks that need to be accomplished, such as:

- ‣ Searching
- ‣ Browsing
- ‣ Alerting
- ‣ Distribution of all research to research distribution vendors
- ‣ Distribution of subsets (such as models) to individual buy-side firms
- ‣ Email distribution to targeted recipients

Based on these use cases, the group determined the initial list of tags that would be needed. From there, a technical committee was given the task of translating this working draft into the RIXML schema.

The final specification is made up of the following components:

- ‣ Object model
- ‣ User guide
- ‣ Schema
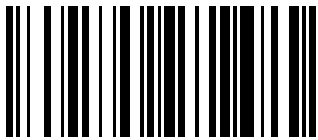- ‣ An issue log that is used to capture ideas for further development

Each of these documents is available in the *Specification* section of the RIXML.org website, and contains valuable additional information.

The Benefits of Automation

Think of RIXML as being like the UPC bar code that appears on many items that you buy. Like a barcode, RIXML:

RIXML

▸ identifies the product uniquely

▸ is applied by the creator of the product following a standard set of rules and protocols

▸ can be used in many different ways throughout the distribution/consumption chain (inventory, check-out, etc.)

How does this help? Well, just as you can make sense of a research report by looking at the cover, a shopkeeper can figure out how many cans of peas are in the shop by looking at the shelf and counting the remaining cans. However, what about when the shop grows to have thousands of products, or expands to a chain of grocery stores? Just as it made sense to automate the inventory process, we in the investment research world realize that it now makes sense to standardize our protocols to enhance automation of the routine task of describing research content.

Another similarity? Once the barcode became a standard on packaging, creative vendors came up with new uses for it far beyond the ones initially envisioned, including:

▸ more sophisticated inventory systems

▸ barcode systems to track package shipments

▸ handheld scanners to automate the wedding registry process

Likewise, research distribution vendors have already begun expanding product offerings in ways that take advantage of the rich tag set to create solutions to their clients' needs.

The best news is that with RIXML, as with UPC codes, the end users are able to benefit without needing to be at all familiar with the technology behind it.
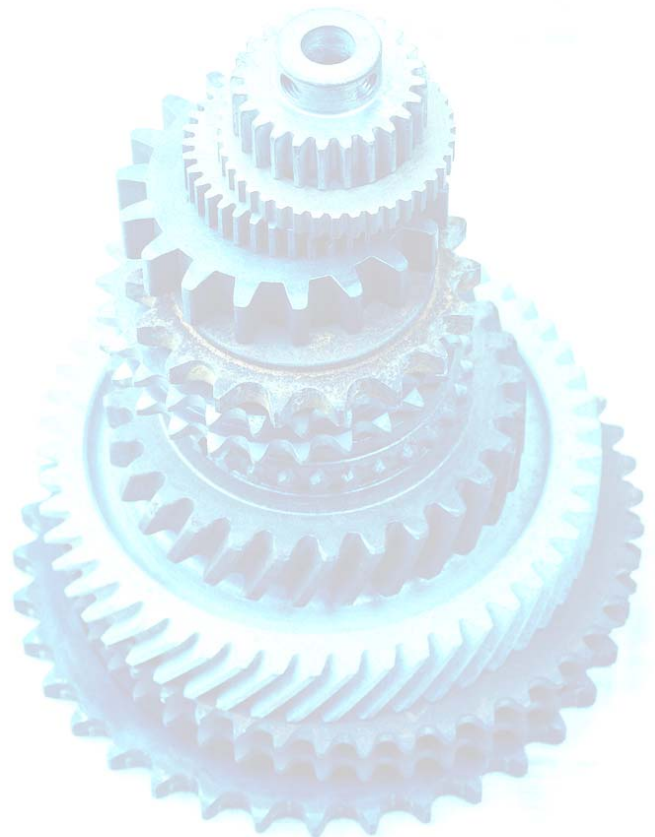
Why do you need tags?

When you look at the front page of a research report, your eye can make sense of it. You can find the title, date, publishing firm and author; you can figure out what the report is about by looking for company names or tickers, macroeconomic concepts, etc. However, in order for electronic systems and products to make sense of them, each of these fields must be tagged, so that they can be placed in the appropriate fields of a searchable data table.


```
<Publisher> JPMorgan
<Publication date> 06/10/2008
<Author> Moscowitz
<Title> 3G iPhone Arrives with New Price
        Point, Broadening Global Reach,
        and Enterprise Focus
<Ticker> AAPL
```

Publisher	Pub date	Author	Title	Ticker
BrokerFirm	01/08/2009	Smith, John Jones, Mary	Apple announces iTunes price cuts	AAPL
JPMorgan	06/10/2008	Moscowitz, Mark	3G iPhone Arrives with New Price Point, Broadening Global Reach, and Enterprise Focus	AAPL
FirmTwo	02/14/2008	Brown, A.J. Green, G.	Quarterly Earnings Call Summary	IBM
JPMorgan	01/19/2008	Moscowitz, Mark Mills, Suzanne	Summary of JPMorgan Tech Conference	AAPL DELL IBM

To do this, a tagging file is created that travels along with the document, providing the necessary tagging to the desired data repository:

By standardizing a common set of tags for asset types, international identifier mappings, industries and sectors, ratings and recommendations, etc., customized filters of various kinds can be created, allowing for better management of the large amount of information currently being distributed.

In addition, this advanced level of tagging and access to individual contextual elements allows for a new generation of alerting mechanisms, navigational frameworks and enhanced user interfaces, and delivery to mobile devices.

Once the data is structured with tags, a number of types of applications can be developed:

‣ Robust search page with high quality results

‣ Alerting functionality

‣ Customized interfaces

How does RIXML help?

RIXML is an XML specification that contains the list of tags identified by RIXML-member firms as being important to describe a research document.

Some of these tags are required, while others are optional. There are some tags that are required for some types of reports, but not for others. For example, if you publish a company report, you *must* provide a company name and ticker, and you *may* choose to include the industry that the company is in; an industry overview report requires you to provide the industry, but does not require you to provide tickers.

When we refer to **tags**, we are referring to the labels that have descriptions of the content. There are actually two different types of tags: **elements** and **attributes.**

An **element** represents a piece of data. That data has a name. It may have attributes, and it may have specific relationships with other elements.

Example:
<**Title**> 3G iPhone Arrives with New Price Point, Broadening Global Reach, and Enterprise Focus </**Title**>

An **attribute** is a property associated with an element. It provides additional descriptive information about the element.

Example:
<Title **language="eng"**> 3G iPhone Arrives with New Price Point, Broadening Global Reach, and Enterprise Focus </Title>

Some elements have no attributes, some have several.

In order to organize the tags, there is a type of element, called a **container element**, that does not have any real data value associated with it. It serves only as a container for grouping other elements and/or attributes.

Example:
<Authors>
<Author> Moscowitz, Mark</Author>
<Author> Smith, Jane </Author>
</Authors>

All other elements and attributes are referred to as **data items**. Data items *do* have real data associated with them.

At a high level, *Product* is the top-level tag in a RIXML document. The term *Product* is used in lieu of *Document* because RIXML is not limited to just PDF- and text-based documents; it allows for a wide range of content, including audio files, video files, webcasts, and meeting announcements. The RIXML standard tags cover all content regardless of the media type. File format is just one of the tags that describes the item.

Additionally, the concept of *Product* refers to a unique research idea, as opposed to an actual research publication. It is possible to publish multiple documents or other files (called "resources") with the same productID, provided they all refer to the same discrete research idea. An example of this would be an audio file of a conference call, along with a PDF of the transcript of the call, and another transcript with a French translation of the conference call.

Within the *Product* container element, there are four major groups of information. In each, there are tags that are used to capture information about the product. There are hundreds of tags in the full specification, although not all tags are used for every product. Here are the four major groups, along with a sample of the types of tags contained in each:

▸ **Source** This is the information about the publishing firm, as well as the individual authors or team that created that document.
> Organization & organization contact information
> Group & group contact information
> Person & person contact information

▸ **Content** This section contains the title, subtitle, synopsis, and other high-level types of description of the document, as well as containing the filename of the file or files that constitute the content.
> Resource ID   Title
> File name   Subtitle
> File type   Synopsis
> Document title   Description
> Language

▸ **Context** This describes the content and how it is intended to be used: what the main topic is, the ticker(s), sector(s), issuer(s), etc.
> Issuer details   Subject
> Security name   Index
> Security ID (ticker, etc.) Asset class
> Publication date   Asset type
> Event information   Sector/industry
> Product category   Country
> Product focus   Region
> Product classification   Rating
> Subject   Issuer financials
> Index
> Series information (frequency, issue number, etc.)
> Coverage, weighting, and rating action

▸ **Legal** This contains the copyright information, the legal disclaimer information, and any other legal information that needs to accompany that particular research product.
> Copyright
> Disclaimer
> Disclosure information

So for example, a research product would first be tagged with a unique identifier. The product would then be further tagged for its source (the publisher and analyst who produced the note), its content (the title of the document, the type of file, and so forth), the context (industry, sector, economic, company), and legal information, if any (copyright information, disclaimers, disclosures etc.).

There are various other sub-tags within each category that make up this overall framework. The full list can be found in the *Specification* section of the RIXML.org website.

Implementation Options

With RIXML, there are 2 levels of implementation:

Level 1

The basic implementation of RIXML includes the tags that were identified by RIXML-member firms as the ones most critical to research. Level One is a subset of the full schema that represents a common starting point for introducing RIXML into a research platform. It includes the most important and most widely implemented tags. At a high level, the tags include:

- Time and date of publication
- Publication status (released, revised, deleted)
- Product type (report, comment, model)
- Language
- Name of the publishing firm
- Name of the author(s)
- Title, subtitle, abstract, synopsis
- Subject
- Reference to the PDF file
- Issuers discussed, with cash flow, revenue
- Securities discussed, with rating, target price, estimates
- Sector/industry classifications
- Country or Region
- Intended audience

> Level One is a subset of the full schema that represents a common starting point for introducing RIXML into a research platform.

Full implementation

While level 1 provides the ability to tag the information about what the report is about, the full implementation enables a larger degree of *componentization* – that is, the ability to tag discrete bits of information. When a research document is created, it is just a "brick" of information. Some publishers will find it helpful to componentize the information in the research report, pulling out information such as:

- Rating action
- Event details (event date, venue, etc.)
- Author's contact information

There are hundreds of tags in the RIXML specification. The more tags a publisher adds, the more sophisticated a vendor can get, and the more targeted a user's search results can be.

> The full implementation enables a larger degree of componentization – that is, the ability to tag discrete bits of information.
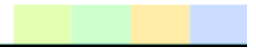
What does it mean to implement RIXML?

Sell-side firms interested in implementing RIXML will need to:
- Update their publishing tools so that they produce RIXML tagging data as reports are published
- Ensure that their tagging policies include the RIXML "required" and "highly recommended" tags, minimizing use of the "PublisherDefined" option
- Validate their content against the standard
- Apply the RIXML standard to their distribution models

Vendors need to:
- Update their content ingestion protocols
- Update their applications to take advantage of RIXML tagging and enumeration lists
- Require content providers to adhere to the standard

Buy-side firms need to:
- Work with technology vendors and/or internal systems developers to adopt the standard
- Require content providers to adhere to the standard
- Integrate RIXML into their investment process

What does it mean to support RIXML Level One?

Supporting of RIXML Level One involves providing tagging for the subset of tags identified in the Level One specification. This ensures that documents have the minimum amount of information identified by RIXML-member firms as essential. Not all tags will be used for all publications; for example, the RIXML file for an economic overview with no reference to any specific company will not include a SecurityID tag, even though SecurityID is a Level One tag – it just doesn't apply to that publication. However, in order to be RIXML Level One compliant, the RIXML file must include all level one tags that *do* apply for that publication. The absence of a Level One tag from a particular RIXML file will mean that although the publisher supports the tag, it determined that the most accurate, meaningful action for that particular RIXML file was to omit it.

Vendor products will be required to support all Level One tags whenever populated by a publisher. Level One tags will be part of the data model behind each vendor platform, and each tag will be available to end users in a meaningful way via the end user interface. Substantive participation will typically mean the ability of platform users to include the tag in searching and/or filtering criteria. Some tags may not be useful or meaningful in search actions, but participate in the context of entitlements instead.

Key concepts

There are a number of key concepts built into the RIXML specification. Below is a summary of these concepts. Additional details can be found in the specification documents.

Report focus

One of the key goals of the RIXML specification is to make it easier to tell more precisely what a report is actually about. Is it an in-depth company report, or a sector overview? Is it a macroeconomic research piece about a certain country, or a multi-company overview of a number of companies in that country?

This is accomplished by the use of a tag that enables the publisher to specify what the report's focus is. A report can be tagged as focusing on a particular issuer (company or other entity), sector, industry, region, country, index, or exchange; or it may more broadly focus on a particular asset class, asset type, or security type.

In defining the focus of the report, you give the other tags a context; the focus of the report can also define whether certain tags are required, highly recommended, or optional.

Required vs. optional tags

In the RIXML specification, each tag can be considered to be required, optional, or highly recommended. A tag that is highly recommended is officially optional; however, the member firms of RIXML.org have determined that it is best practice to use it.

Whether a tag is considered required, optional, or highly recommended often depends on the type of research being described. For example, in a company report, the identifier (a ticker, CUSUP, or other unique identifier) is required. However, in an industry overview, company identifiers are optional – if the report mentions specific companies by name, the publisher may wish to include them; a more general industry overview that mentions concepts but not specific companies may not require them.

The recommended practice is to include all tags that are relevant.

Tag re-use

When a tag or a set of tags can be used in multiple places, we define it one time, and re-use it throughout the specification. For example, contact information (address, website, email information, etc.) can exist for an organization, for a group within an organization, and for an individual. The RIXML specification has a container element called *ContactInfo* that gets re-used whenever there is a place in the specification that allows contact information to be provided. Re-using tags makes the RIXML specification easier to implement and easier to keep up-to-date.

Related tags

In the RIXML specification, some tags are grouped together, and are used and re-used as a set. For example, people information. Information about a person can be used in many ways in the RIXML specification (as the author, a member of a group, or an event host, speaker, or attendee); but any time a person is mentioned, there is a set of related tags that provide related information – contact information, address, etc. The official term for this is *inheritance* – when a person is listed anywhere in the specification, they inherit all of the attributes of the Person class.

One vs. many

Some tags are allowed to be used multiple times, others cannot. The options include:

- **Up to one** - the tag may be used once, or may not be used at all, but it cannot be used multiple times for that file; for example, subtitle
- **One** – the tag must be used, and can only be used once; for example, title
- **Zero or more** – the tag may be used once, multiple times, or not at all; for example, issuer
- **One or more** – the tag must be used at least once, but can be used multiple times; for example, a security can be associated with one or more identifiers – ticker, Bloomberg code, RIC, etc.

The official name for this in the specification is *Cardinality.*

Primary vs. secondary

In cases where there are multiple data items are associated with a tag (see one vs. many, above), it is often desirable to mark some of the terms as being primary. For example,

- a report may mention several companies, but one of them may be the primary focus of the report
- a report may have several authors, but one of them is the main author

Many elements in RIXML can optionally include a primaryIndicator attribute, which can be set to *Yes* or *No*.

Use of XBRL for financial data

Because the XBRL protocol is designed to address the specific requirements of describing financial information, the RIXML specification directs users to insert the appropriate XBRL tags and content when they are tagging financial information. See the RIXML.org website for more information about the relationship between RIXML and XBRL.

Enumeration lists

In addition to standardizing the tags that are used, the RIXML specification also standardizes some of the terms that the publisher can use to populate these tags. For example, for rating actions, we have standardized a list that includes the following:

Initiate	Revise	ReviewForUpgrade
Upgrade	NewRating	ReviewForDowngrade
Downgrade	Affirmed	ReviewDirectionUncertain
Reiterate	PositiveOutlook	RatingWithdrawn
Drop	NegativeOutlook	

These lists were developed with the input of member firms, and are updated and revised as needed. Standardizing terms facilitates bringing similar content together and removes the ambiguity between the different tagging methods and systems used by different publishers and vendors. This will enable consumers to quickly find the research that they are interested in and compare the offerings from various firms much easier. For example, these may become the items in a dropdown list on a search page.

Publisher defined terms

In many enumeration lists, there is an option called *PublisherDefined*. This is available for those instances where none of the available terms in the list applies.

Some data items in the RIXML schema take values only from an enumerated list. For example, the classificationType attribute of a SectorIndustry can take any of the following values: *GICS*, *ICB*, or *PublisherDefined*. GICS and ICB are well-known industry classification systems. However, a particular publisher may wish to specify an industry associated with a research piece based on the publisher's own proprietary classification system or based on some new system not yet included in the RIXML list. The *PublisherDefined* value is used when a publisher does not find the desired value in the enumerated list. An attribute is set to *PublisherDefined*, and a companion tag is used to hold the publisher's own value.

Note that technically, a publisher could use *PublisherDefined* wherever it appears, populating the data with whatever terms they use in their in-house systems. However, this would essentially make this content invisible to consumers who are using tools that use the RIXML-defined enumeration lists in dropdown menus, etc. Thus, it is strongly recommended that *PublisherDefined* be used sparingly. Note that since many tags allow multiple values, a publisher could use a term from the RIXML enumeration list *and* add a PublisherDefined value for use in internal systems, etc.

How does the specification evolve and change?

Changes will be made to the specification as needed to address business and technology issues and additional versions of the specification will be released accordingly.

Each time an updated version of the specification is available, we will post access to the specification on the RIXML.org website and encourage all market participants to respond with comments and suggestions within a 30-day period. We are very interested to hear from both buy- and sell-side users regarding what additional content they would like to see tagged as well as what hurdles they encountered for implementing RIXML or any other concerns they had regarding the specification. The purpose of industry feedback is to find out how we can enhance the RIXML protocol to further improve the process of categorizing, aggregating, comparing, and distributing global financial research.

Where can I get more information?

A wealth of additional information can be found on the RIXML.org website, particularly in the *Specification* section.

Additionally, firms who are working to adopt RIXML will find the support of the RIXML.org organization in assisting to answer any questions:

<div align="center">

RIXML.org
email: rixml@jandj.com
phone: 212-655-2945

</div>