RIXML.org
Research Information Exchange Markup Language

RIXML Implementation Guide

Background, first steps, and best practices for RIXML version 2.4

Published June 10, 2013

Table of Contents

Preface

Copyright and incorporation information

RIXML.org Limited ("RIXML") is a UK incorporated company.

Contact information

RIXML.org
5 Hanover Square
New York, NY 10004
website: RIXML.org
email: rixml@jandj.com
phone: 212-655-2945

Schema revision history

Date	Version	Comments
05 Feb 2013	2.4	Added terms and definitions to many enumeration lists; added major new element for describing organization expertise (with sub-structure); added additional person details; updated structure of Rating element; added mirror tags for certain tags to enable formatting; and added two new side-car schemas (Roster Updates and Coverage Updates). Does not break backward-compatibility with version 2.3.1.
11 Jan 2010	2.3.1	Patch release containing a spelling correction to one of the enumerations. Does not break backward-compatibility with version 2.3.
13 Apr 2009	2.3	Improved definitions of more than 120 enumerated values; changed some tag cardinalities from required to optional based on real-world use cases that demonstrated that some tags formerly marked as required did not always apply and should be optional. Does not break backward-compatibility with version 2.2.
26 Mar 2005	2.2	Updated specification to facilitate use of XBRL tagging; also imposes constraints on the tags covering language, currency, and country codes in order to enforce use of corresponding ISO standards. Does not break backward-compatibility with version 2.1.
25 Mar 2003	2.1	Major release incorporating updated information regarding tagging of ratings. **Does** break backward-compatibility with 1.0.
20 Jun 2002	2.0	Major release pulling together improvements along five vectors: globalization, events, distribution/entitlements, financial data, and fixed income. **Does** break backward-compatibility with 1.0.
20 Jun 2001	1.0	Initial production release of the RIXML specification, focusing on support for Equities.

Acknowledgements

RIXML.org would like to thank Alan Francis of Factset for contributing significant content, particularly regarding technical issues and best practices; Richard Brandt and Quark for contributing software and technical expertise; and Sal Restivo of JP Morgan for providing technical expertise.

Section 1: Introduction

About RIXML.org

RIXML.org is a consortium of buy-side financial services firms, sell-side financial services firms, and vendors who provide products and services for distributing investment research. The goal of RIXML.org is to define an open protocol that enhances electronic research distribution by improving the process of categorizing, aggregating, comparing, sorting, searching, and distributing global financial research.

The individuals who represent their firms include both IT experts and business-side project managers who represent the analysts, portfolio managers, and others who produce, distribute, and/or consume investment research.

About the RIXML specification

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

About this guide

This implementation guide is one component of the RIXML release packet, and is intended to be used in conjunction with the other components of the release. This document is designed to bridge the space between our *Getting Started Guide*, which gives an overview of RIXML and explains how it can help your organization, and the data dictionary and schema files, which provide the in-depth definitions and structure of RIXML and each tag in it.

This guide outlines the approach taken by RIXML.org in creating the RIXML standard, explains the fundamental concepts regarding tagging and XML, and provides guidance for implementing RIXML. It also explains some of the core principles of the RIXML specification and defines some best practices for implementing RIXML.

RIXML release packet

This implementation guide is one part of the RIXML release packet. Please check the rixml.org website to make sure you are using the most up-to-date release packet, which includes the following components:

- RIXML Implementation Guide (this document)
- RIXML Data Dictionary – human-readable catalog of attributes and elements, with descriptions
- RIXML schema files
- Level One addendum
- Release notes
- A set of sample instance document files

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

Where can I get more information?

A wealth of additional information can be found on the RIXML.org website, particularly in the Technical > Current 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:

RIXML Program Office
c/o Jordan & Jordan
5 Hanover Square
New York, NY 10004
Phone: 212-655-2945
Email: rixml@jandj.com
Fax: 212-422-8570

How can I provide feedback?

You can provide any feedback or suggestions to rixml@jandj.com. A summary of our process for incorporating changes can be found in the "Updates" portion of Section 2 of this document.

Section 2: Our Approach

Approach summary

RIXML.org 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 RIXML specification was developed to address the needs of:

- investment research end users such as analysts and portfolio managers
- the IT departments that support these end users
- vendors and others who create products used by these end users

When the RIXML consortium began, all of the participating firms 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 like to be 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
- distributing all research to research distribution vendors
- distributing 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 technical committee started by modeling the business requirements into an object model. It also re-organized the draft document to better address the requirements of and expectations for the RIXML specification. A parallel effort helped translate this object model into the actual RIXML schema. Some members of the technical committee took the schema back to their organizations and presented it to the potential end-users of the specification. The comments and concerns they elicited have been incorporated into the specification; others were added to an issues log for further analysis and potential inclusion in subsequent releases. This *Implementation Guide* provides details on our process, the specification, the object model, and the schema.

Rationale

As the RIXML specification was originally being created, and throughout every subsequent revision, there have been three guiding principles that the organization has kept in mind:

Guiding principle 1: The RIXML specification must provide an improvement for all parts of the research distribution process

The sheer magnitude of investment research available today, combined with an increasing number of distribution channels and an unprecedented need to navigate this content efficiently means that a sophisticated method of describing investment research information is essential.

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 increases the likelihood 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.

Guiding Principle 2: The specification must be able to expand and evolve to keep up with technology and business changes

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?

- **More research** – 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 and information that fit their needs.

Unfortunately, the tools used to create and distribute research had not evolved to handle these changes. The RIXML specification was designed to be able to keep up with changes in the marketplace, and was designed to be device-, product-, and distribution channel-independent.

Guiding Principle 3: End users must be able to benefit from RIXML without even needing to know it exists

RIXML is a technology solution to a business need – the need to better manage the huge amount of investment research, and to deliver it to the end user in a targeted, precise manner. However, this solution is implemented behind the scenes. End users will notice that their search results are getting better, or that content producers are delivering content formatted for optimal viewing on mobile

devices, or that vendors are creating products that deliver better integration between investment research and other investment data, **but they don't need to know about RIXML to do so.**

The Result

The RIXML specification 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 and a set of rules that guide the usage of the tags and enumeration lists.

RIXML IS:

- the tagging 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.
- an open standard, with no license fees

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

The rest of this implementation guide will provide the details of what XML is and how it works, how RIXML is structured, and how the tags, rules, and enumeration lists work together to improve the process of creating, aggregating, and distributing investment research.

Updates

Because the investment research landscape and the technology landscape are always changing, the RIXML specification was designed to be adaptable. 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.

The RIXML organization is committed to keeping the specification up-to-date, but it is also aware that frequent revisions are difficult for contributors and aggregators to manage. We want to keep a release calendar that balances responsiveness to market changes against not overburdening the IT staff with having to revamp constantly. Generally, there is a new release of the RIXML specification approximately every 2-3 years, some of which are minor releases, and some of which are major releases. Minor version releases are numbered as point releases (i.e., 2.0 to 2.1, 2.1 to 2.2, etc.) and maintain backward compatibility. Major version releases are denoted as full-number increases (i.e., 1.0 to 2.0, 2.4 to 3.0, etc.); these generally require breaking backward compatibility.

Each time an updated version of the specification is available, the proposed specification will be posted on the RIXML.org website. All market participants are encouraged to respond with comments and suggestions within a 30-day period. At the end of the comment period, any necessary edits or

modifications will be made, then the new release packet will be announced and will be available on the RIXML.org website.

Contributors and vendors should add support for each new revision as soon as possible.

Section 3: Tagging and XML: What they are and why they are important

Before you can fully appreciate the power of RIXML, you must understand why tagging is important and why RIXML uses XML as its tagging framework.

Why do you need tags?

Imagine that you make and sell a large number of products, each of which has hundreds of pieces. To package these products, you have three choices:

1. you can put all the pieces for each product into an unlabeled cardboard box. From the outside, every box looks exactly the same.
2. you can put all the pieces for each product into a cardboard box that has a label on it giving a basic description of the what it contains.
3. you can put the pieces into multiple boxes that organize the various components, each with a very specific label describing the contents. For the more complicated sections, there are boxes within the boxes to keep everything well organized. All of these boxes are nested neatly in one big box, which has a very clear label describing the contents.

Which would you choose? Which one do you think your potential customers would prefer? In this example, the packaging makes a huge difference in the usability of the product itself. The product can exist without good packaging, but it becomes far more usable – and valuable – with it.

What is this product? Investment research. What are the pieces? All of the valuable information – company information, financial data, recommendation information, and analysis – that each report contains.

XML is sophisticated packaging for electronic information. Furthermore, RIXML is sophisticated packaging customized **specifically for** investment research.

The evolution of investment research distribution

In fact, distribution of investment research has evolved in a way much like the packaging described above. Understanding this evolution is key to understanding why RIXML is so important.

Investment research distributed by mail or as an email attachment is essentially packaged in an unlabeled box. Without opening the envelope or the PDF, you have no information about its contents. Once you do open it, you need to skim the report to determine whether it will be of use to you, then you need to manually re-key any information that you would like to include in a modeling spreadsheet.

The original electronic aggregation and distribution systems were a definite improvement. They added a small number of key tags to each research report. These tags were like a basic label, providing at least some information about the report - like ticker, publishing firm, author, and publication date. Definitely more useful than an unlabeled box, and fine if you if you are dealing with a fairly small number of research reports. When investment research was first being distributed electronically, most buy-side analysts and portfolio managers were happy to get every research report available on the companies they covered or owned in their portfolios, so the fact that the early electronic investment research aggregation systems only allowed for very basic searching and filtering was not a problem.

However, today's marketplace requires a more efficient way to manage the huge volume of investment research; moreover, investment research consumers expect to be able to receive information in various ways depending on the device or platform they are using at the moment.

What is a tag?

When you look at the front page of a research report, you can scan it quickly and 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.

This is fine as long as the number of reports you have to process is fairly small. However, the immense volume of investment research that is published today means that investment professionals rely on electronic systems to locate the research that is important to them.

<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

In order for electronic systems and products to make sense of reports, each of these fields must be tagged, so that the information can be placed in the appropriate fields of a searchable data table.

A tag is simply a way of indicating what information you have – for example: this is the title; this is the author; these are the tickers for the companies mentioned in this report.

What can you do with information once it is tagged?

Tagging investment research content provides benefit through all stages of the investment research creation, distribution, and consumption process:

- **Research creators** can apply a rich array of tags to their content. Doing so enables them to thoroughly describe their content, making it more likely that their research will find its way to precisely the clients who are looking for it via searching, real-time alerting, or automated data feeds. In fact, much of this rich, structured tagging information already resides in the internal databases that analysts use to create their research reports already, but without a standard like RIXML, this information becomes unstructured when the report is published, typically in as a PDF.

- Additionally, **research creators** can combine and re-use the discrete bits of information to create multiple products: the full research report could be published as a PDF and distributed to research aggregators; a research alert with just the title, author, rating upgrade, and brief description could be distributed using a smartphone app, and data feeds could be distributed to clients to allow them to automatically import content into their internal systems, etc.

- **Research aggregators** can create sophisticated searching and alerting tools that enable end users to pinpoint research reports, and can create tools to feed specific data fields into consumers' data tables and models.

- **Research consumers** can benefit not just from enhanced vendor offerings, but also because each data item is *componentized* – tagged as a standalone bit of information – and can be more easily downloaded directly into modeling spreadsheets or other applications.

The research creation and distribution systems used for investment research all rely on tagging to organize, sort, filter, and deliver investment research. However, they use tagging protocols that are not able to interact with one another. XML provides a way for the investment industry to change that.

What is XML?

XML is a highly flexible tagging system. It is a protocol that enables individuals or groups to define a custom set of tags that can be used to mark up content in the way that best describes the particular content that is being organized, and to define the set of rules that describe how these tags should be used.

The term XML stands for e**X**tensible **M**arkup **L**anguage.

> **eXtensible** refers to the fact that it is meant to be extended, or expanded. XML is made up of a core set of rules that describe how to create a tagging structure. Individuals or groups use these rules – *extend* them – to create the exact set of tags that they need to fit the content being described.
>
> **Markup** means that you are "marking up" the content – that is, adding tags to indicate that certain text is the title, other text is the author, other text is the publication date of the report, etc.
>
> **Language** refers to the fact that XML is a language, with a set of syntax rules. 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.

The core of XML is the set of syntax rules about how tags are to be structured and how content must be placed in order to be covered by the tags. For example:

- there is a specific way to indicate that the file is an XML file
- the tags are enclosed in angle brackets (that is "<" and ">") whereas the actual content is not
- tags must be nested properly
- tags are case sensitive

However, this is just the framework. Using this framework, an individual or group can determine what tags are required to describe their content, and can create tags using XML's syntax rules. As long as the XML file conforms to the XML syntax rules, it is said to be "well-formed" XML. That means that, like a correctly formed English-language sentence, the file conforms to the rules required for XML and for that particular XML specification.

An XML file can be a stand-alone file, or it can act as a tagging file that travels along with a document.

XML is a way of structuring data so that it can be exchanged. Many firms use XML to manage and structure content so that it can move seamlessly through their own internal systems. However, XML becomes even more powerful when a schema is embraced by an industry as the standard mechanism for communicating or exchanging information. When this occurs, the process is streamlined from beginning to end. These tags can be leveraged to manipulate the data depending on the end user's requests or requirements. Data tagged this way can be presented through different platforms (desktop, mobile device, etc.) and different media types (browser, PDF, mobile device app, spreadsheet, etc.), with different displays (as a standalone report, a chart, a grid, or a data feed), by changing the presentation style, without making any changes to the actual data.

XML and investment research

Investment research contains much information that must be exchanged – there is financial data that must be pulled from data repositories into research reports, the research information itself must often be compiled and re-compiled into multiple research products, and the resulting reports must flow from the research producers to the research consumers, often through content aggregators. This highly complex, highly inter-related, highly time-sensitive system is one that would benefit from a standardized information exchange protocol – and that protocol is RIXML.

Section 4: RIXML: a custom XML specification for investment research

What is the RIXML specification?

RIXML uses the XML protocol to structure a set of tags and rules that govern the way that investment research content is described. Investment research content includes not only research notes and reports, but also other investment research, recommendation, and analysis content such as conference call webcasts, morning meeting compilations, etc. These tags and the rules that govern them are called the *RIXML specification*.

The RIXML specification contains the list of tags identified by RIXML-member firms as being important to describe a research document. There are hundreds of tags.

- Some of these tags are required, while others are optional.
- Some tags are required for certain types of content, 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; conversely, for an industry overview report, you must provide the industry, but you may or may not choose to provide tickers for the companies mentioned in the report.
- Some tags contain sub-tags, creating a nested structure that helps hold, organize, and group other tags.

The specification also contains the rules that must be followed in using these tags. For example:

- Every report must have a title.
- A company report does not need to provide an industry tag, but an industry overview report does.
- The valid terms for a rating action must be selected from a particular list of valid terms for that field (this list is called an enumeration list).
- Any date field that is used must be structured in a particular way, following the ISO standard for date.

These tags and rules are collected and described in the *schema*. The RIXML schema files are the actual set of .XSD files that represents the relationships and components defined within the object model. The "XSD" extension indicates that these files are created following the rules defined by the XML Schema Definition standard from W3C.

The schema files contain the detailed list of tags used in the RIXML specification, the rules that govern them, and the list of terms that are valid for enumerated fields. This schema files can also be used to validate actual RIXML instance documents to ensure that the firm implementing the specification is creating RIXML-compliant content.

In addition to the schema files, each RIXML release packet includes a *Data Dictionary*. This document contains all of the tags, attributes, enumeration lists, and information contained in the schema files, but formatted in a more human-readable way. The *Data Dictionary* enables the members of a RIXML implementation team to determine which tags they will need in order to describe their content, the best practice guidelines for using those tags, and the way to organize those tags.

How is the RIXML specification applied to research content?

To apply RIXML to a particular piece of research content (such as a research report or an audio file of a morning call), the relevant tags along with the information that is described by these tags are formed

into a tagging file. This tagging file is called an *instance document*. This instance document does not replace the research report or audio file. Instead, it travels along with the research item, providing the necessary tagging to the desired data repositories. Within the instance document are the tags and content that describes the content in the research item (for example, <Title>Initiating Coverage of XYZ Corporation</Title>) as well as the tags and content being used to uniquely identify the research item (for example, <Name>Strategy_Daily_20120512.pdf</Name>).

By standardizing a common set of tags for the information that is generally included in research content, such as security identifiers, industries and sectors, ratings and recommendations, asset types, etc., the creators, aggregators, and consumers of investment research are able to vastly improve the process for getting the right research into the hands of the consumers who want and need it. 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. The remainder of this *Implementation Guide* contains details on implementing RIXML, core concepts behind RIXML, and details on the schema itself. These sections provide information that is critical to understand in order to begin to leverage the strengths of RIXML.

Section 5: Implementing RIXML

Hopefully, by now you have a sense of how RIXML can help your organization. The next step is determining exactly what implementing RIXML means, and what steps are needed in order to do so at your firm.

Implementation options

Accurate tagging provides a way for research creators to communicate with research consumers about the contents of every research document they publish. The RIXML standard's goal is to streamline the research distribution process for content producers, consumers, and aggregators.

The RIXML specification offers the ability to enhance investment research with hundreds of tags, allowing the producers to describe each piece of investment research in great detail, and even to describe *parts* of their research reports in great detail. However, the RIXML organization also recognizes that this can seem like an overwhelming task, and so has defined two levels of implementation:

Level One

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 research report (PDF, audio file, etc.)
- issuers discussed, with cash flow, revenue
- securities discussed, with rating, target price, estimates
- sector/industry classifications
- country or region
- intended audience

Supporting 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 are used by back-end systems to manage entitlements instead.

Note that any valid Level One RIXML instance document is a fully valid RIXML document. Level One includes all required tags, along with the optional tags that RIXML.org member firms defined as tags that are critical in describing investment research.

Full implementation

While Level One 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.

First steps

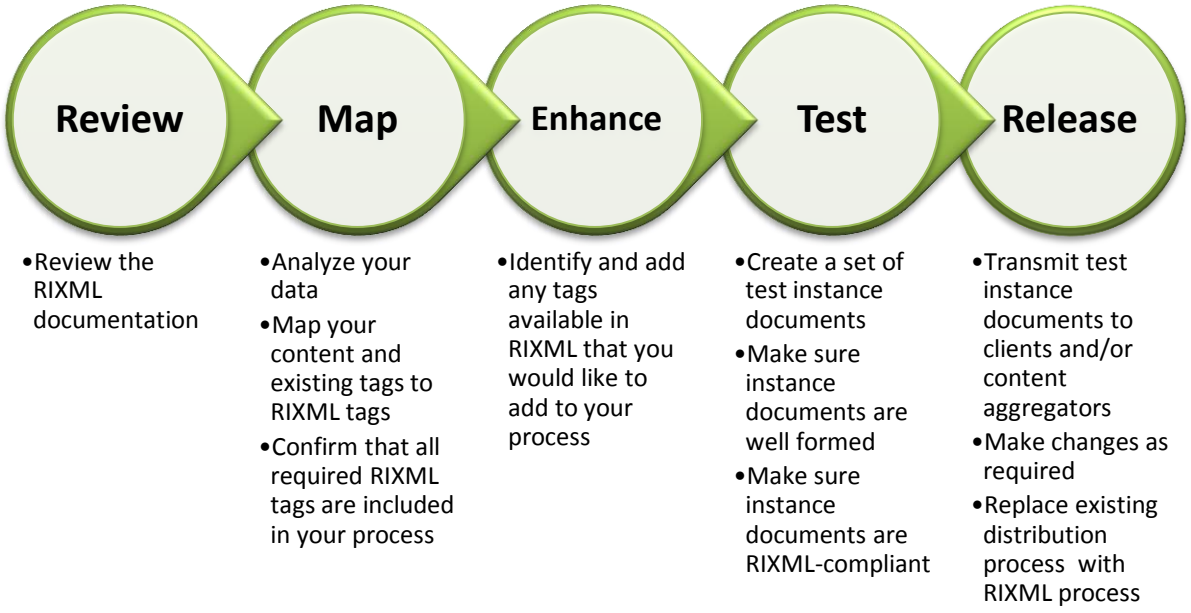
The steps for implementing RIXML vary depending on where you are in the investment research distribution chain. Below are steps that content producers, content aggregators, and content consumers would take. This section will outline what implementing RIXML means and the steps involved in doing so.

For content producers

Sell-side firms and other producers of investment research 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

Their first steps would look something like this:

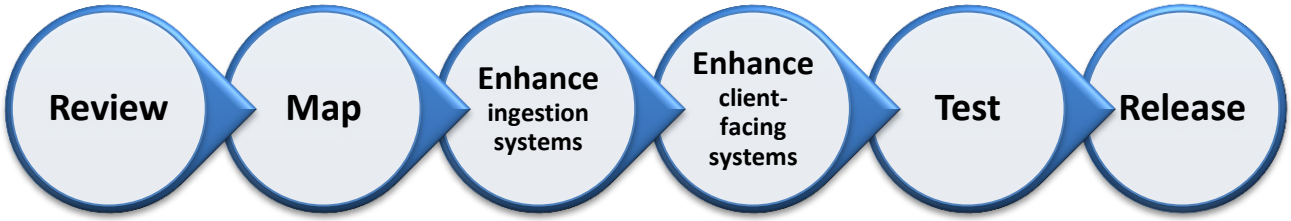


For content aggregators

Research aggregation vendors will 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

Their first steps would look something like this:



- Review the RIXML documentation

- Analyze your content aggregation system
- Map your tags to RIXML tags
- Confirm that all required RIXML tags are included in your process; correct your system as needed

- Identify and add any tags available in RIXML that you would like to add to your process
- Update databases, content ingestion systems, contribution tools, etc.
- Work with content producers to update content submission process

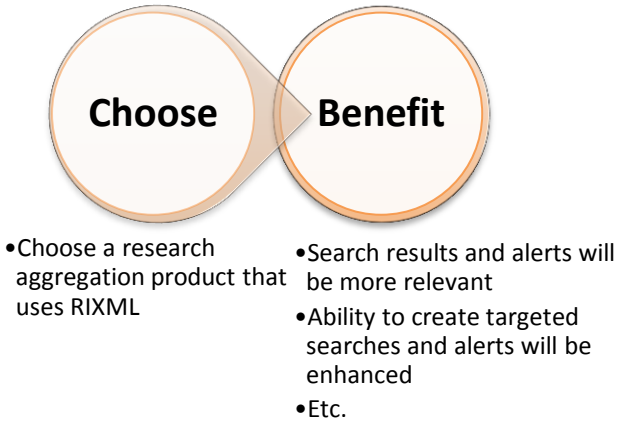
- Update client-facing systems (research search pages, alert setup pages, etc.)

- Create a test database
- Import a set of test RIXML documents
- Validate content in internal databases
- Validate that client-facing systems properly use this content
- Make corrections as needed

- Release into production

For content consumers

For content consumers, adopting RIXML is simple:



Yes, it's that simple. RIXML was designed such that end users do not even need to know about RIXML to benefit from it.

Of course, many firms that are content consumers have custom research solutions; in that case, your custom products would follow the steps outlined for content producers and/or content aggregators, as needed. By having your internal research creation tools, the data feeds you receive from sell-side firms, and the research you obtain from content aggregators all using RIXML, you are able to integrate all of this content in a way that improves efficiency for both your investment professionals and your IT department.

For others

Vendors involved in other parts of the research creation and distribution chain will need to identify which parts of the process described above apply to them. We encourage you to contact RIXML for assistance.

Section 6: Core principles and concepts

Regardless of where you are in the investment research creation and distribution process, and regardless of whether you decide to begin with RIXML Level One or the full implementation, there are a number of core principles and concepts that are critical to understand. This section covers these core principles and concepts.

Basics

Content producers must be able to write once, send everywhere

Without RIXML, content producers need to create and maintain entirely separate mechanisms to send their research content and its associated tagging information to each research aggregation vendor and client that receives their research. However, using RIXML enables firms to create and maintain a single process for transmitting their research content to all of their data aggregation vendors and clients, greatly improving efficiency and reducing maintenance costs.

There may be cases where a vendor may require a small amount of additional tagging; for example, to add a company identifier, author identifier, entitlement group information, etc. using that vendor's proprietary methodology. Note that this information should be added *in addition* to the report's core tagging. For example, if one vendor has a proprietary method for uniquely identifying authors, that information can be added to the report author information *in addition* to the standard name, phone number, email address, etc. This way, the XML file is still valid and understandable if sent to a different client or vendor who does not use that identification system. The content of the remaining tags should remain consistent and should not be changed when being sent to various vendors or clients.

No recipient may demand or expect customization on the part of contributors that would render the RIXML file unfit for some recipients or render the RIXML file non-compliant with the RIXML schema.

Identifying report focus is critical

One of the key goals of the RIXML specification is to make it easier to define 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, municipality 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.

Defining the focus of the report gives the other tags a context. It enables the research creator to describe the report in such a way that a research consumer can find it or be alerted to it when it matches the criteria they have established. This increases the value of research by ensuring that it is seen by the right consumers at the right time.

The focus of the report can also define whether certain tags are required, highly recommended, or optional.

Tagging

The RIXML specification has hundreds of tags, but not all of them will be used to describe any one report. The specification was designed to describe many types of research – morning call summaries, in-depth company reports, macroeconomic overviews, and more. It also provides the framework for handling estimate changes, event notifications, etc.

Using tags accurately and thoroughly, and utilizing RIXML's enumeration lists, ensures that end users can be confident that their search results include all of the research that they do need, while minimizing irrelevant results.

Because the core concept of RIXML is to structure content by the use of tags, there are many aspects of tagging that are critical to understand. These include:

Tagging should always be done in the best interest of consumers

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. Any doubt about how to populate a tag should be resolved in the way that best serves consumers.

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).

Tagging should clearly identify what a report is about

Tag values should identify what a report is mainly about, not merely what it mentions. The key question to ask before adding a tag is, "Will a person who searches for content using this tag be interested in this report?" Although there are instances in which the answer to that question is a matter of opinion, the answer to this question is fairly straightforward in most cases.

The RIXML specification provides a number of tags that work together to help research creators accurately tag content. This enables content aggregators to distribute this content to the right consumers, and enables content consumers to search for precisely the research they want, without getting bogged down in irrelevant research.

This *Implementation Guide* and the *Data Dictionary* provide specific information about the tags that are used to identify what a report is about. As you read these documents, you will learn the details about how to put this concept into practice.

The main tag used to describe what a report is about is the ProductFocus tag. (See the "Identifying report focus is critical" section, above, for an introduction to this tag.)This tag indicates whether a report is mainly about a security, about a country, an asset class, etc. Once that has been defined, the RIXML specification enables a report to be further tagged to indicate even more precisely what it is mainly about.

EXAMPLE 1:

Two reports might be tagged as covering a particular industry. However, one report is an industry overview; the other one provides short updates on several companies in that industry.

1. The first report would be tagged with a ProductFocus of SectorIndustry and no company identifiers would be added.
2. On the second report, the ProductFocus would be Issuer, because it is actually about the individual companies, not about the industry as a whole. The identifiers of the various companies discussed would be added, all with their PrimaryIndicator set to "No" (because none of them is in fact the primary focus of the report.

EXAMPLE 2:
A company report is mainly about XYZ Auto Corporation; in part of the report, it discusses XYZ's sales figures relative to those of its competitor Acme Corporation. This report would be tagged with the ProductFocus of *Issuer*, because unlike the example above, even though the two companies (issuers) in this example are in the same industry, the report is about one particular company and discusses company-specific topics, not industry-wide ones. In this case, XYZ Auto Corporation would be tagged with a PrimaryIndicator of "Yes," indicating that it is the primary focus of the report. The identifier (ticker) for ACME Corporation would also be included, but with a PrimaryIndicator of "No" since it is not the primary focus of the report.

EXAMPLE 3:
A report that covers a merger should be tagged with the Subject tag that indicates this (MergerAcquisitionDivestiture). It would likely have a ProductFocus of Issuer, as in the example above; depending on the nature of the report, both companies might be tagged with a PrimaryIndicator of "Yes," indicating that they are both the primary topics of the report. Another report that merely mentions this merger (such as an industry overview) would generally *not* be tagged with the Subject tag of MergerAcquisitionDivestiture, because it doesn't cover the topic in sufficient detail to be useful for someone searching for information on this topic.

As you can see, the Product Focus tag works in conjunction with the tags related to the focus that has been chosen. Additional information about these tags can be found in this *Implementation Guide* and the *Data Dictionary*.

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 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.

Tag types and relationships

One of the key strengths of an XML specification is that it provides the ability to define *relationships* between and among the tags within each instance document. Although these relationships, and the way they are implemented with the RIXML specification, can get very technical, it is extremely helpful to understand at a high level the main types of relationships that exist. Below is a basic description of the key types of relationships that are used in RIXML.

As described in section 3 above, the structure of XML specifications, including RIXML, can be thought of as a system of boxes within boxes and descriptive labels on those boxes. In looking at the RIXML Data Dictionary (the human-readable version of the specification), you will notice that the following terms are used to describe the boxes and labels:

- elements
- attributes

An **element** is a tag that describes a certain kind of information. These are the key building blocks.

Many elements are **aggregations**. An aggregation is an element that has other elements in it. Aggregations enable related content to be organized and kept together. Because an XML specification structure is organized like nested boxes (in a tree structure), many elements are aggregations that have one or more elements within it.

Additionally, many elements have **attributes** associated with them. An attribute is a property associated with an element that provides additional descriptive information about the element. Attributes can be thought of as information on labels attached to a box. Some elements have no attributes, some have several.

EXAMPLE:
To demonstrate how the types of tags work with each other, it would be helpful to see some real RIXML tags in use. Here we will build a section of RIXML tagging that does just that. Let's suppose that we want to build the part of a RIXML tagging file that describes the firm that is publishing a research piece. To do that, we will use the <Organization> element. This is the set of tags used to indicate the name of an organization that is related to the research item. It will always have a closing tag of </Organization>. We begin with that tag:

<Organization>
</Organization>

To indicate that the publishing firm is a sell-side firm, we add an attribute called *type*. This attribute's values must come from the OrganizationTypeEnum enumeration list. The relevant value is "SellSideFirm" – let's add that now:

<Organization *type="SellSideFirm"*>
</Organization>

The next tag we'll add indicates the actual name of the publishing organization. This tag is called <OrganizationName>:

```
<Organization type="SellSideFirm">
   <OrganizationName>Smith & Co.
    </OrganizationName>
</Organization>
```

These are the required tags. In addition, the publisher will most likely add the tags for the contact information for the firm – address, main phone number, etc.

Tag data type

Elements and attributes are the tags that identify the information being tagged. The main types of tags used in the RIXML schema are:

- **Alphanumeric tags:** the content in many tags is not restricted, and can contain any text, of any length. For example, the title, synopsis, and author fields are free text tags. In the schema files and the Data Dictionary, free text tags are identified by the term *String*.
- **Numeric tags:** the content in some tags must be a number. In the schema files and the Data Dictionary, numeric tags are identified by the term *Integer*.
- **Date/time tags:** there are many date fields used in the RIXML specification, including publication date, date of status change, the date that a set of entitlement parameters is valid, etc. All date fields are formatted following ISO 8601. Date/time tags are identified by the term *TimeInstant*; the term *TimePeriod* is used for fields that indicate duration (such as the duration of a meeting). For more information about the formatting of date tags, and the management of timezones, please see the Date/Time section, below.
- **Enumerated tags:** some tags utilize an enumeration list. For these tags, *only* the values included in the enumeration list are valid values for that field. For additional details, see the Enumeration Lists section, below. In the schema files and the Data Dictionary, enumerated

tags indicate the specific enumeration list that applies to that tag.

One vs. many

Some tags will logically be used only once in any research item, but others might be used more than once. For example, a research report will only have one title, but might have more than one author. The official name for this in the specification is *Cardinality*. Each tag's cardinality is defined in the schema.

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

One	1	the tag must be used, and can only be used once; for example, title
Up to one	..1	the tag may be used once, or may not be used at all, but it cannot be used multiple times for that research item; for example, subtitle
Zero or more	0..*	the tag may be used once, multiple times, or not at all; for example, issuer
One or more	1..*	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. If a report uses the issuer tag, then it must use at least one identifier for that issuer, but it can use multiple identifiers if desired.

Required vs. optional tags

In the RIXML specification, each tag is defined as being either required, optional, or highly recommended. 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, a company identifier (a ticker, CUSIP, 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 mention any individual company enough to warrant including it as a tag.

Technically speaking, a tag that is highly recommended is actually optional; however, the member firms of RIXML.org have determined that it is best practice to use it.

The recommended practice is to include all tags that are relevant, whether they are required or optional.

Tag re-use

When a tag or a set of tags can be used in multiple places, it is generally defined one time, and is used and re-used as needed throughout the specification. For example, contact information (address, website, email information, etc.) can exist for an organization, for a group within an organization, or for an individual. The RIXML specification has an 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.

Primary vs. non-primary

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

Note that it is possible to have a field where multiple terms are listed, but all are set with the primaryIndicator of *No*. For example, in an industry overview, a number of tickers are referenced. Since none of them is the central topic of the report, each has its primaryIndicator set to No. This enables research aggregation vendors to create search pages that allow end users to search for research primarily about a company, or for any research that mentions it.

Enumeration lists

Many RIXML tags utilize enumeration lists. An enumeration list is the list of all terms that are allowed to be used in that field. It is a controlled list that is designed to ensure that like information is grouped together.

For example, for rating actions, we have standardized a list that includes the following:

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

These lists were developed with the input of RIXML 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 more easily. For example, these may become the items in a dropdown list on a search page.

The "PublisherDefined" option

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, or when the publisher wants to add one or more values in addition to the ones in the enumeration lists. The PublisherDefined option enables RIXML users to provide their own custom ("publisher-defined") values in a way that remains RIXML-compliant.

It is critical that the PublisherDefined option be used sparingly and judiciously. While it is tempting to use this tag to essentially bypass the enumeration list and use the terms that you use internally, it is important to realize that using it can actually hide your research. The only reason to use this is if the information is important AND no existing tag comes close. Remember that tags allow similar information to be grouped together. Thus, if you tag your research as "PublisherDefined" and use "Initiating," all of your research will be missing for someone who searches using the term "Initiate" that

already exists in the enumeration list. If you contribute your research to a research aggregation vendor, bear in mind that the vendor will probably use the enumerated values in drop-down lists on search pages, etc. "Initiate" will in the dropdown list that end users use to create searches; "Initiating" will not, so your research will not show up in search results.

The PublisherDefined value is used when a publisher does not find the desired value, or a close substitute, in the enumerated list. An attribute is set to PublisherDefined, and a companion tag is used to hold the publisher's own value.

EXAMPLE 1:
A publisher uses the term "Initiating" in their internal systems. The RIXML enumeration list for the Coverage Action field uses "Initiate" but also contains a "PublisherDefined" option.

PREFERRED:	Publisher would map their term "Initiating" to "Initiate"
ACCEPTABLE:	Publisher would map their term "Initiating" to "Initiate" but would also use the "PublisherDefined" field, adding "Initiating" in addition to "Initiate."
UNACCEPTABLE:	Publisher would use the "PublisherDefined" field to bypass the enumeration list and use "Initiating" instead.

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. This is particularly useful in instances when using RIXML for both internal processes and publication to external sources.

EXAMPLE 2:
In example 1, "Initiate" and "Initiating" are essentially synonyms, and there would be little reason to add "Initiating" in addition to "Initiate." However, the PublisherDefined value exists because there are instances when a valid term exists that has no close match in the enumeration list. For example, the IssuerSecurityIDType tag is used to identify what type of security is being used. There are over a dozen options, including CUSIP, SEDOL, Bloomberg code, Reuters Identification Code, etc. However, other valid codes could exist that would uniquely identify an issuer.

PREFERRED:	Publisher would select one or more of the existing security ID types and map their tag to them.
ACCEPTABLE:	Publisher would provide at least one identifier from the existing list, but would also use the "PublisherDefined" field, adding additional identifiers as needed.
NOT RECOMMENDED:	Publisher would only use the "PublisherDefined" field, adding desired identifier.
UNACCEPTABLE:	Publisher would use the "PublisherDefined" field to bypass the enumeration list and using the term "ReutersCode" instead of "RIC."

We recommend that publishers provide at least one value from the existing enumeration list whenever possible, because great care has gone into creating enumeration lists that provide the broadest list of terms commonly used in the industry. As other terms come into use, it is fine to add to the list, but generally, it is better to map your terms to the closest existing tag rather than creating a new one. It is NOT acceptable to bypass the enumeration list solely to simplify processing on your side.

IMPORTANT NOTES:

1. 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. If a publisher does not support the non-standard value that you contribute, the field will be blank in their database and will not show up in searches, etc.

2. If you find that there is a term new to the industry (for example, a new issuer identification term, a new type of firm publishing investment research, etc.), RIXML.org recommends that you submit the new term to RIXML.org for consideration to be added to the appropriate existing enumeration list.

Use of other standards

One of the key goals of RIXML is to enhance integration. Therefore, if there are ISO or other standards that are already in common use in the investment industry for particular tags, the RIXML organization will generally choose to use those existing tags or standards rather than creating separate ones. The following other standards are used for the relevant tags in RIXML. Links to each standard's official website, which contain the information needed to implement it, can be found in the Technical>Reference Links section of the RIXML website.

Country Codes

The RIXML specification uses the ISO 3166-1 alpha-2 standard for identifying countries. This standard defines 2-letter codes for each country. For example, the identifier for the United States is 'US', not 'USA' or 'U.S.' or 'United States'.

Currency Codes

All currencies used within the RIXML specification adhere to the ISO 4217 standard's three-character alphabetic codes.

Date/time

RIXML uses ISO 8601 for all date/time tags. See the Date/time section below for more information.

GICS codes

The Global Industry Classification Standard (GICS) was developed by Morgan Stanley Capital International (MSCI) and Standard & Poor's (S&P). The GICS classifications are designed to provide the global financial community with accurate, complete and standard industry definitions.

ICB codes

Dow Jones Indexes and FTSE have created a classification system called the Industry Classification Benchmark (ICB). The system is supported by the ICB Universe Database, which contains over 40,000 companies and 45,000 securities worldwide from the FTSE and Dow Jones universes.

Language Codes

Both the Product and Resource objects use the ISO 639-2/T specification to identify language. For example, the identifier for English is 'eng', not 'en' or 'English').

Market Identifier Code

RIXML uses ISO 10383 Market Identifier Code to identify securities trading exchanges.

MIME Types

The RIXML Resource object has a tag for specifying the MIME type of the attached or linked media. MIME types are explained and defined in RFC 2046.

XBRL for financial data

Because the XBRL standard 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.

Date/time

Date/time tags used in RIXML, such as createDateTime and publicationDateTime, have very specific formatting requirements because they must conform to the ISO 8601 standard for dateTime values. Specifically,

- the year must 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 3rd day of the month is '03', not '3'.
- date/time values should include a time zone indicator

Although the ISO standard for date/time does not require a time, RIXML strongly advises against omitting the time OR using a default value of 00:00:00 (that is, midnight). Because many data aggregators display reports in reverse chronological order, submitting research with the time of 00:00:00 will make them appear in results further back in the results than they should, instantly "hiding" them behind the research that was correctly tagged as it was published.

Likewise, using a time of 23:59:59 (that is, one second before midnight) is also discouraged, as some data aggregators do not display research with a publication date (or time) in the future. Thus, reports published with a time of 23:59:59 may not show up until just before midnight, hours *after* they should have shown up.

If a report is published at 10:15 AM GMT, correctly applying the time value in the publication date field will ensure that it appears at the top of a vendor's "recent research" screen shortly thereafter, and will appear in a portfolio manager's alert list in a timely manner. If that report is published with 00:00:00 as the time, it will appear with the information that was published at midnight, which has long since scrolled off the screen; if it is published with 23:59:59, it may not be displayed until just before midnight, and will have long since scrolled off the screen by the next morning.

Although the RIXML recommendation is to convert your publication time to Zulu (UTC) time, only do this if you actually convert the time; do not simply use the "Z" time zone indicator while using your local time. This will cause research to be mis-tagged and hidden for the same reasons as described above.

Additionally, the ISO standard does not require a time zone indicator. However, 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. Therefore, RIXML strongly recommends that you convert your time to UTC; if this is not feasible, the recommendation is to include the time in local time WITH the appropriate time zone indication. Vendors may or may not convert it, depending on their internal processing protocol.

Key do's and don'ts regarding date/time fields:

DO:	use time as well as date
DO:	use Zulu (UTC) time OR accurately indicate which time zone was used
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 default time of 00:00:00 or 23:59:59 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.

Phone numbers

Unfortunately, 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

This is one of the few fields that you may need to customize depending on the requirements of each research aggregation vendor.

Because many vendors combine the country code and the phone number tags, you should not put the country code into both the CountryCode and Number tags.

Issuer/security identification

There is no single, universal standard for identifying securities.

- Different types of securities have different systems – equities are assigned ticker symbols by the exchanges that list them; fixed income securities are assigned CUSIPs, SEDOLs, etc.
- different countries and regions have different issuing agencies
- equity securities are assigned tickers by the exchange that issues them, but the same ticker is often used to identify different companies on different exchanges. For example, the ticker "F" is used by the NYSE to identify Ford, but Borsa Italiana uses it to identify Fiat.
- Different research aggregation vendors have created identification systems with unique identifiers, usually by combining the ticker and an indication of exchange or country, but these are not recognized by other vendors.

In order to enable investment research publishers to create a unique identifier that can be recognized by all vendors, RIXML enables publishers to indicate the security type (CUSIP, exchange ticker, RIC or Bloomberg code, etc.) and the trading exchange for each security identifier. These tags work together to ensure that the publisher can clearly indicate which security is being referred to in a way that enables all research aggregators and other recipients to understand.

The trading exchange tag is governed by ISO 10383, which provides a standard for Market Identifier Codes.

RIXML recommends that publishers include multiple identifiers from the most universally accepted identifiers – CUSIP, SEDOL, ISIN, Bloomberg, and RIC (Reuters) codes, and *also* include the vendor-neutral "Security + security type + Market identifier code" combination.

Section 7: The RIXML Schema

Each section of the RIXML schema is made up of a package name (a name for the collection of objects that were logically grouped together). Within each package there is a description of elements and attributes that make up that package. The schema files and the *Data Dictionary* contain the same information. An electronic system that produces RIXML tagging files will use the schema document to "tell" it how to format the output for any given piece of research content. This output file has a file extension of .XML and is referred to as an "instance document." It will contain all of the tags that are relevant to describing the underlying research piece (a PDF, an audio file, etc.), as well as the tags that are required by RIXML validation.

Product is a high-level tag in a RIXML instance document, meaning that it acts as a parent to many sub-elements – an outer box into which all of the other boxes fit. The term *Product* is used rather than the term *Document* for two key reasons:

First, RIXML is not limited to just PDF- and text-based documents; it allows for a wide range of content formats, including audio files, video files, webcasts, and meeting announcements. Each of these is a product, but you probably wouldn't use the term "document" to describe all of them. The RIXML standard tags cover all content regardless of the media type. Of course, one of the tags that describes each item identifies the media format of that item.

Second, the concept of Product refers to a unique piece of research content, as opposed to a particular publication. In most cases, these are one and the same. In some cases, though, there can be multiple documents or other files (called "resources") associated with one product. An example of this would be an audio file of a conference call, a PDF of the transcript of that call, and another transcript with a French translation of the conference call. Each of these would have a unique resourceID, but would all have the same productID. Note that this is different from a report that refers to other reports within it; RIXML allows a publisher to provide a reference to a particular other report. The method for doing so is covered in the *Data Dictionary.*

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

- **Source:** This is the element that contains information about the publishing firm, as well as the individual authors and/or teams that created that document.
  - Organization & organization contact information
  - Group & group contact information
  - Person & person contact information
- **Content:** This is the element that contains the title, subtitle, synopsis, and other high-level types of description of the research item, 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 is the element that describes the content and how it is intended to be used: what the main topic is, the ticker(s), sector(s), issuer(s), research discipline, related content, product focus, etc.

Asset class	Issuer financials	Region
Asset type	Product category	Sector/industry
Country	Product classification	Security ID (ticker, etc.)
Event information	Product focus	Security name
Index	Publication date	Subject
Issuer details	Rating	

Series information (frequency, issue number, etc.)
Coverage, weighting, and rating action

- **Legal:** This is the element that 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

EXAMPLE:
A research report is published.
It is first tagged with a unique product identifier.
It is also tagged with a unique resource identifier.
(As described above, the research identifier technically refers to the research report; whereas the product identifier refers to the "unique research content" contained within it.

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)
its context (industry, sector, economic, company)
its 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 of tags, along with definitions and proper usage, can be found in the *Data Dictionary.*

The Object Model

To organize and visualize the tags that are used in an XML schema such as RIXML, the designers use something called an object model. This model shows the different objects that are used by the schema. For example, in RIXML, you could think of "author" as an object – as a box within the specification. This object is described by the <Person> tag; when used in the PersonGroupMember with a Role attribute set to "Author" within the Source element, it indicates that the person was an author of the research report being tagged. Within the <Person> tag, there are a number of subtags that provide further description, such as the person's name. The specification also provides the ability to define whether a person was the primary author or not. Another object might be Rating. Within that object would be the tag that indicates what the rating is, along with other tags indicating the date the rating was published, etc. The Object Model provides a helpful way of seeing the tags and the relationships between and among the tags.

The *Data Dictionary* provides object model diagrams throughout the document that show all of the elements and attributes and the relationships between them.

XML Schema Files

The RIXML documentation includes, among other files, a set of schema documents and a data dictionary.

- The schema documents (in .XSD format) contain the detailed list of tags used in the RIXML specification, the rules that govern them, and the list of terms that are valid for enumerated fields. These files are sometimes called "XSD files" because the file extension for an XML schema document is .XSD.
- The *Data Dictionary* provides this same information in a more human-readable format. Both contain all of the tags in the specification, along with the relationships between them.

The RIXML schema is the actual set of XSD files that represents the relationships and components defined within the object model. These files can also be used to validate actual RIXML instance documents to ensure that the firm implementing the specification is creating RIXML-compliant content.

Section 4 of the RIXML *Data Dictionary* contains a section for each of the packages used to organize the tags within the specification. Within each package there is a description of objects that make up that package. Following the definition of any object is a listing and definition of the elements and attributes associated with that object (indicated by the use of an upper- or lower-case first letter, respectively). If the class is an aggregator of other classes, the aggregations are listed after the listing of elements and attributes. Where an enumeration modifies an attribute (denoted by the suffix "Enum," e.g., "SubjectEnum"), refer to the list of Enumerations at the end of the *Data Dictionary*.

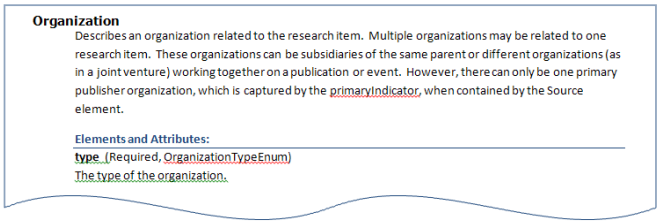
An electronic system that produces RIXML tagging files will use the schema document to "tell" it how to format the output for any given piece of research content. This output file has a file extension of .XML and is called an "instance document." It will contain all of the tags that are relevant to describing the content of the related piece of research (a PDF, an audio file, etc.), as well as the tags that are required in order for the RIXML file to be valid.

EXAMPLE:

The schema file has a section that looks like this:

```
<xs:element name="Organization" type="OrganizationType">
   <xs:annotation>
      <xs:documentation>Describes an organization related to the research
      item.  Multiple organizations may be related to one research item.
      These organizations can be subsidiaries of the same parent or different
      organizations (as in a joint venture) working together on a publication
      or event.  However, there can only be one primary publisher
      organization, which is captured by the primaryIndicator, when contained
      by the Source element. </xs:documentation>
   </xs:annotation>
</xs:element>
```

The corresponding information in the *Data Dictionary* looks like this:



Organization

Describes an organization related to the research item. Multiple organizations may be related to one research item. These organizations can be subsidiaries of the same parent or different organizations (as in a joint venture) working together on a publication or event. However, there can only be one primary publisher organization, which is captured by the primaryIndicator, when contained by the Source element.

Elements and Attributes:

type (Required, OrganizationTypeEnum)
The type of the organization.

The corresponding section of an instance document (the output file) looks like this:

```
<Organization type="SellSideFirm">
     <OrganizationName>Smith & Co.
     </OrganizationName>
</Organization>
```

Section 8: Conclusion

This guide was designed to give you a firm grasp of the core concepts behind RIXML, details on the RIXML framework, and a sense of the steps required to implement RIXML at your firm. This guide is designed to work with the *Data Dictionary*, which gives in-depth information about every tag used in the RIXML specification.

If you have any additional questions, please do not hesitate to contact the RIXML organization.

RIXML.org
5 Hanover Square
New York, NY 10004
website: RIXML.org
email: rixml@jandj.com
phone: 212-655-2945