NRSS: A Protocol for Syndicating Numeric Data

Jerry Liu, Glen Purdy, Jay Warrior, Glenn Engel
Communications Solutions Department
Agilent Laboratories
Palo Alto, CA 94304 USA
{jerry_liu, glen_purdy, jay_warrior, glenn_engel}@agilent.com
April 2005

Abstract

This paper proposes NRSS (Numeric Really Simple Syndication), a protocol for syndicating numeric data over the web. It builds upon RSS (Really Simple Syndication) version 2.0, a popular protocol used to syndicate headlines for news stories around the web, and the data model from IEEE 1451.1, a standard for representing and describing numeric measurement data. This note provides an overview to NRSS and outlines some possible usage scenarios. It also describes how NRSS extends regular RSS and illustrates how a NRSS numeric summary feed can be constructed via some examples.

1. Introduction

NRSS is a protocol for syndicating numerical data over the web. These numeric contents include examples such as the temperature of a pond, the price of Agilent Technologies stock, or the spin-up time of a disk drive. It is an extension of the popular protocol Really Simple Syndication (RSS) Version 2.0 specifications [1].

RSS is used for news web sites like CNN.com to syndicate the headlines of textual files, such as news stories, to other web sites. The process works as follows. An RSS client gets a news feed from CNN.com based on RSS, meaning that it queries and receives from the CNN.com server a summary of the news articles available on CNN.com, along with links to the actual news story. If a reader is interested in one or more of the headlines, he can then follow the links in the news summary to retrieve the entire news story file via traditional means. The intended use of the RSS protocol is for a news reader to retrieve news summaries from various news sites without needing to browse through the sites to retrieve entire files just to get a summary. Many RSS clients exist today, and most popular news sites, as well as a growing number of smaller sites, provide RSS news feeds for the contents that exist on their site.

In this paper, we propose NRSS, an extension of numeric, rather than textual, data model onto the basic RSS protocol and describe how the different attributes needed to transfer numeric data transfer can be mapped onto RSS. In the same way that RSS is used to retrieve headlines of news stories, NRSS can be used to retrieve the highlights of a particular numerical data source, as well as the associated information to make sense of the numerical data and links to retrieve all or portions of the original data set.

The version of the RSS specification used in this note is RSS 2.0 available at

http://blogs.law.harvard.edu/tech/rss

With this mapping based on RSS, when a numeric data consumer retrieves a NRSS Numeric Summary from a Data Set provider, it can:

❑ Discover the identities of the numeric data sets available on the server.

❑ Retrieve a description and latest values from the numeric data set.

❑ Retrieve metadata about the data set such as the number of points, last updates, etc.

❑ Determine when the measurements were last updated.

❑ Determine when to check back for new changes.

❑ Subscribe to be notified when new measurements of interest appear.

2. Definitions

This section defines the terms used in this document.

❑ Numeric Data Source - The entity which provides the actual Data Sets. This entity can be a sensor making measurements and generating numeric data on the fly, or a database archive which has archived the data and is serving them on request.

❑ Data Item - A single numeric value. An example is the price of Agilent technologies at 2:00 pm.

❑ Data Set - A collection of Data Items grouped together because they describe the same subject.. An example is a collection of Agilent Technologies stock prices throughout an entire day.

❑ Numeric Newsfeed - A file containing summary information about a Data Set.

❑ Newsfeed Client - A program which retrieves Numeric Newsfeed, parses the contents, renders the information to the user, and allows the user to retrieve the entire Data Set.

❑ Newsfeed Provider - A program which provides Numeric Newsfeed to the Newsfeed Client.

3. Block Diagram

The NRSS system consists of three primary components: a Numeric Data Source, a Newsfeed Provider, and a Newsfeed Client.
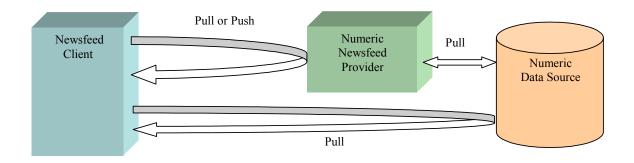

Figure 1. The Data Producers (Newsfeed Provider and Data Source) and the Consumer (Newsfeed Client) in an NRSS System.
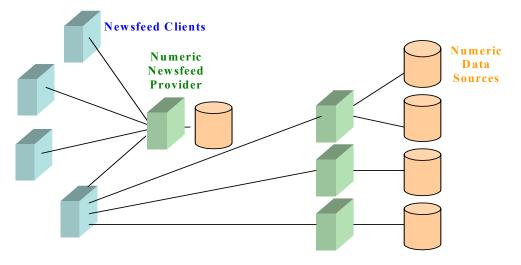
Figure 2. One Newsfeed Provider can provide Numeric Newsfeeds for multiple Data Sources. A Newsfeed Client can obtain Numeric Newsfeeds from multiple Newsfeed Providers at once, like in RSS.

The Newsfeed Client retrieves either a Numeric Newsfeed from the Newsfeed Provider or the actual Data Set contents themselves from the Numeric Data Source. The Newsfeed Provider is responsible for providing the NRSS based Numeric Newsfeed as well as other related information about the Data Sets. Numeric Data Source is responsible for providing the actual Data Items organized into Data Sets.

A single Newsfeed client can retrieve newsfeeds from several newsfeed providers, and a single newsfeed provider can supply feeds to multiple newsfeed clients.

4. Operational Description

In normal operation, a Newsfeed Client queries the Newsfeed Provider periodically to retrieve a Numerical Newsfeed. The Numerical Newsfeed provides information such as identity (i.e. what the data is) and metadata (i.e. units, time range) of the numeric data source, as well as a digest of the Data Set (i.e. latest value, maximum value, and/or average value), pointers to the actual Data Set, and description of how to access and interpret the actual Data Set.

In most cases, the Numerical Newsfeed describing the Data Set should provide sufficient information for the Newsfeed Client. However, if the Newsfeed Client wishes to retrieve Data Items from the actual Data Set, it can use the information in the Numerical Newsfeed to interface with the Data Source to retrieve the Data Sets.

The Newsfeed Client can also register to be notified if certain conditions are met by the numeric data source, such as when a numeric value has exceeded a preset threshold. This is another mode of operation which can be used in place of or in addition to the polling method. Details on how this is accomplished is given in Section 7.

5. Data Model

Data consists of numbers, such as measurements, which describe some entity in the world. Metadata is data about data. It provides the additional information necessary to interpret the data. Data Representation describes the mechanism that is used to store and manipulate data.

For example, a piece of data may be the number 78. Metadata may provide information such as the dimension (Temperature), units (Degrees), when this data was measured (yesterday), or where it was measured (Palo Alto). A representation of this data may be manifested in a XML document, such as <nrss:arg n="Value">78</nrss:arg>.

A Numeric Newsfeed contains elements of these three aspects of a data set (the actual data, description of the representation, and the metadata). Since these elements may be quite large, particularly the data, the newsfeed would contain only a summary of each of these elements, along with instructions to the client as to how to retrieve the rest of the information.

6. Newsfeed Format

This section describes how the numeric data set summary can be represented in XML, using the RSS. 2.0 tags for compatibility. A schema must be made available for all NRSS feeds.

NRSS introduces a new set of tags to be included within the <description> field of a regular RSS newsfeed. With this mapping, regular RSS news readers are able to parse and display a NRSS numeric summary. If an XSL stylesheet is included with the numeric summary, and the reader is XSL stylesheet aware, the reader can even render the numeric information as human readable HTML.

The NRSS tags describe various attributes of the data and metadata of the data set. The attributes are drawn from the list described in the IEEE 1451.1 standard [2].

The following list shows how the tags in RSS are mapped into the NRSS numeric data model.

- ❑ **rss** - Required. Top level tag. Should be version 2.0
  - o **channel** - Required. There may be multiple tags. Represent a collection of numeric data sources. A channel may contain multiple numeric data sets.
    - ▪ **title** - Required. Name of this collection of data sources.
    - ▪ **link** - Required. URL of the page providing the feed.
    - ▪ **description** - Required. Text description of this collection of data sources. The content will be displayed to the user.
    - ▪ **category** - Optional. One or more strings that denotes the category of measurement collection.
    - ▪ **generator** - Optional. Name of the program that generated this file.
    - ▪ **cloud** - Optional. A URL where a client can register to receive notifications of new measurements being available. Details elaborated in following sections.
    - ▪ **ttl** - Optional. The remaining time that this document is valid, in seconds.
    - ▪ **image** - Optional.. A user viewable image that contains a synopsis of the data.

- **item** - Required. There may be multiple tags. Represents a numeric data set. There may be multiple numeric data sets aggregated into a channel.

  - **author** - Optional. URI for the data source that generated the data.

  - **category** - Optional. String that denotes the category of numeric data source. There may be multiple category tags if the content belongs in multiple categories.

  - **pubdate** - Optional. Timestamp of when this numeric data set was last updated.

  - **source** - Optional. URI of the Record (channel) that this measurement (item) is a part of.

  - **description** - Required. Contains the measurement data and metadata payload.

    - **nrss:metadata** - Required. Information that provides metadata about this measurement set, such as the number of measurement points, maximum/minimum values, time range that the measurements span, units, uncertainty, etc.

      This section contains name/value pairs which provide the metadata for the Data Set. The representation is based on an XML schema derived from the Argument data model as described in the IEEE 1451.1 standard. In this schema, there are only two tags: *arg*, and *argArray*, corresponding to Argument and Argument Array types. An ArgumentArray is a collection of name/value pairs of Arguments, and all data types are considered to be an Argument, including an Argument Array.

      The name of an Argument is indicated by an attribute named *n* for the *arg* tag. The type of an Argument is indicated by an attribute called *t*. As an example, an Argument which is named 'size' and of type Integer would have a tag which looks like

      <arg n="size" t="int">

      The possible set of attribute name (n="x") values for the arg tag is unbounded, though the well-known and/or required ones are listed below. One or all of these may be included in a metadata description. The inclusion of some of the names may exclude others. The Numerical Newsfeed may contain a only portion of the total metadata, in which case the remaining metadata is provided at the location specified by the **url** parameter.

      - **abscissaIncrement -** Increment along the independent variable per IEEE 1451.1 definitions.

      - **description** - a human readable string which describes the data set.

- **dimension -** Dimension of the measurement. Describes what kind of physical quantity this measurement represents.

- **interpretation -** Measurement interpretations, per IEEE 1451.1 definitions. Possible values: actuator, sensor, computation, setpoint, etc**.**

- **name -** Name of data set

- **numberOfOctets -** For digital parameters only

- **numberOfSigDigits -** For analog parameters only

- **parameterType -** Measurement type, per IEEE 1451.1 definitions. Possible values: analog, discrete, digital, etc.

- **scaleType** - Possible values: nominal, ordinal, absolute, ratio, logarithmic.

- **sensorUrl** - URL of the SensorML document which characterizes the sensor used to generate this data set.

- **size** - Number of data items in this data set.

- **startTime -** Timestamp of first value in data set. Expressed in ISO 8601 notation.

- **stopTime -** Timestamp of last value in data set. Expressed in ISO 8601 notation.

- **uncertainty -** Measurement uncertainty

- **unit -** Unit of measurement

- **url -** URL for retriving the rest of the metadata.

- **nrss:data** - Required. Information that provides the actual data content and describe how this numeric data set is represented. There may be multiple instances of this tag if the data in the datasets are divided into multiple data partitions. The data may reside in multiple partitions if the data is represented in multiple formats. Though the majority of the datasets should only have one partition.

  There are two children tags under this tag, one provides the data while the other provides representation information.

  - **nrss:content -** Required. Information that describes the data contents and/or how to retrieve the actual data contents. This section contains name/value pairs in the same format as nrss:metadata section. The possible attribute names are listed below.

    - **average** - Average value of the data partition.

- **current** - Current data item in the data partition.
  - **timestamp** - Timestamp of data item
  - **value** - Value of current data item.
- **max** - Highest value in the data partition.
  - **timestamp** - Timestamp of data item. Expressed in ISO 8601 notation.
  - **value** - Value of current data item.
- **min** - Lowest value in the data partition.
  - **timestamp** - Timestamp of data item. Expressed in ISO 8601 notation.
  - **value** - Value of current data item.
- **url** - URL where the rest of the data can be retrieved.

- **nrss:representation** - Optional. Information that describes how this numeric data set is represented. This section contains name/value pairs in the same format as NRSS:metadata section. The possible attribute names are listed below.
  - **format** - Denotes the format of the representation description. Currently the only ones supported are *dap* (for Data Access Protocol (DAP) [3] format) and *csv* (two different comma separated file types).
  - **url** - URL for retrieving the rest of the representation description.

7. Data Formats

Currently, three different representations are supported for data representation. This is reflected in the *format* attribute within the <nrss:representation> tag. The possible choices are:

- ❑ dap - Data Access Protocol version 2.0 data format [3] from OpenDAP[4]. The representation description is fetched from the supplied *url* parameter. The representation is in the format as described by DAP version 2.0 specifications.

- ❑ csv1 - Columns of numbers, separated by commas. The representation description is fetched from the supplied *url* parameter. The representation is in a file with as many rows as there are columns in the data set. The name of each column is in its own row.

  A typical csv1 representation may look like:

```
<nrss:representation>
```

```
                    <nrss:arg n = "format">csv1</nrss:arg>
                    <nrss:arg n = "url">http://localhost/desc.txt</nrss:arg>
            </nrss:representation>
```

With the representation file may look like:

```
    value

    timestamp
```

And the actual data file:

```
    24.3    2005-01-23 23:34:00Z

        34.6    2005-01-23 23:35:00Z
```

❑ csv2 - Columns of numbers, separated by white space characters or commas. There is no separate representation description. Instead, the contents of the first row in the dataset is used as the description for the rest of the columns.

A typical csv2 representation may look like:

```
    <nrss:representation>
            <nrss:arg n = "format">csv2</nrss:arg>
    </nrss:representation>
```

with the actual data file:

```
    value    timestamp

    24.3    2005-01-23 23:34:00Z

        34.6    2005-01-23 23:35:00Z
```

8. Notification

In typical RSS mode of operations, the News Client queries the Newsfeed Provider for Newsfeeds. However, it is also possible for a client to register an interest in some aspects of the data set, such as the data set exceeding a certain value or when the current data meets some some predefined criteria.

The cloud field contains a URL where a client can register to be notified of the events of interest. The parameters are passed in as HTTP parameters.

These are the parameters:

❑ action - actions to take (i.e. add or delete a notification)

❑ channel - the names of the channels to monitor

❑ max - the maximum rate (notifications per year) that the client should receive a notification.

❑ min - the minimum rate (notifications per year) that the client should receive a notification.

❑ method - the method that should be used to notify the client (possible values: SMTP or HTTP).

❑ method-addr - the address used to notify the client (either a URL for HTTP or an e-mail address for SMTP).

❑ threshold - specifies the type of alarm condition to monitor

❑ threshold-parm - content dependent on type of threshold

These are the name value pairs which will be included in a notification from the provider back to the client

❑ status - status code

❑ value - the value of the data item of interest.

9. Examples

Newsfeed Example

Here is an example Numeric Newsfeed about the condition of the hard drive on a PC. These are measurements that one may read from the S.M.A.R.T. disk drive monitoring system in modern ATA drives.

This feed contains two data items. The first data item is the spin-up time of the drive while the second is the number of errors that has been recorded. The first data item is an example where the available data is provided within the feed, while the second data item is an example of where some of the available information has been provided, as well as the URLs where the Newsfeed Client can retrieve the rest of the information.

Note that the first data item does not have a data content URL provided, meaning that there is no additional data to be fetched at the data source. Thus, the data representation field has not been provided as well.

```
<rss version="2.0" xmlns:nrss="http://www.nrss.org">
      <channel>
            <title>Jerry's Disk Drive!</title>
            <link>http://www.example.com/nrss/drive</link>
            <description>Status of Jerry's Disk Drive</description>
            <language>en-us</language>
            <copyright>Copyright 2005 Jerry Liu</copyright>
            <generator>NRSS v0.0.5</generator>
            <webMaster>email@example.com</webMaster>
            <category domain="http://www.example.com/nrss">
                  diskDrive
            </category>
            <cloud domain="radio.xmlstoragesystem.com"
                  port="80" path="/RPC2"
                  registerProcedure="xmlStorageSystem.rssPleaseNotify" protocol="xml-
rpc"/>
            <ttl>60</ttl>
            <item>
                  <link>http://www.example.com/nrss/drive/spin</link>
                  <guid>http://www.example.com/nrss/a1866</guid>
                  <pubdate>Mon, 13 Mar 2005 18:54:09 GMT</pubdate>
                  <description>
                        <nrss:metadata>
                              <nrss:arg n="size">1230</nrss:arg>
                              <nrss:arg n="unit">mS</nrss:arg>
                              <nrss:arg n="uncertainty">10</nrss:arg>
                              <nrss:arg n="description">
                                    Spin-up time of drive
                              </nrss:arg>
                        </nrss:metadata>
                        <nrss:data>
                              <nrss:content>
                                    <nrss:argArray n="current">
                                          <nrss:arg n="value">78</nrss:arg>
                                          <nrss:arg n="timestamp">
                                                2005-02-03 12:52:09.000Z
                                          </nrss:arg>
                                    </nrss:argArray>
                                    <nrss:argArray n="average">
```

```
                                        <nrss:arg n="value">56</nrss:arg>
                                </nrss:argArray>
                                <nrss:argArray n="low">
                                        <nrss:arg n="value">17</nrss:arg>
                                </nrss:argArray>
                                <nrss:argArray n="high">
                                        <nrss:arg n="value">122</nrss:arg>
                                </nrss:argArray>
                        </nrss:content>
                </nrss:data>
        </description>
</item>
<item>
        <link>http://www.example.com/nrss/drive/readerror</link>
        <guid>http://www.example.com/nrss/a1868</guid>
        <pubdate>Mon, 13 Oct 2004 18:54:09 GMT</pubdate>
        <description>
                <nrss:metadata>
                        <nrss:arg n="description">
                                Number of read errors recorded
                        </nrss:arg>
                        <nrss:arg n="size">230</nrss:arg>
                        <nrss:arg n="unit">errors</nrss:arg>
                        <nrss:arg n="uncertainty">0</nrss:arg>
                        <nrss:arg n="url">
                                http://www.example.com/nrss/a1868.xml
                        </nrss:arg>
                </nrss:metadata>
                <nrss:data>
                        <nrss:representation>
                                <nrss:arg n="format">DAP</nrss:arg>
                                <nrss:arg n="url">
                                        http://www.example.com/dods/a1868.dds
                                </nrss:arg>
                        </nrss:representation>
                        <nrss:content>
                                <nrss:arg n="url">
                                        http://www.example.com/dods/a1868.dods
                                </nrss:arg>
                                <nrss:argArray n="current">
                                        <nrss:arg n="value">14</nrss:arg>
                                        <nrss:arg n="timestamp">
                                                2004-09-13 09:54:09.000Z
                                        </nrss:arg>
                                </nrss:argArray>
                                <nrss:argArray n="min">
                                        <nrss:arg n="value">0</nrss:arg>
                                </nrss:argArray>
                        </nrss:content>
                </nrss:data>
        </description>
</item>
        </channel>
</rss>
```

10. References

[1]   Dave Winer. RSS 2.0 Specification. http://blogs.law.harvard.edu/tech/rss. 2002.
[2]   Institute of Electrical and Electronics Engineers. IEEE Std 1451.1-1999, IEEE Standard for a Smart
       Transducer Interface for Sensors and Actuators - Network Capable Application Processor (NCAP)
       Information Model. 1999.
[3]   James Gallagher, et al. The Data Access Protocol - DAP 2.0.
       http://www.opendap.org/pdf/dap_2_data_model.pdf. 2004.
[4]   Open-source Project for a Network Data Access Protocol. http://www.opendap.org.