J1.2 USER PREFERENCES FOR WEATHER DATA DISSEMINATION STANDARDS ON THE WEB

John L. Schattel Jr.* and Paul R. Hershberg
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, Maryland 20910

Robert Bunge
Director, Internet Services
Office of Chief Information Officer
National Weather Service, NOAA
Silver Spring, Maryland 20910

1. INTRODUCTION

In 2003, the National Research Council (NRC) published a report entitled "Fair Weather: Effective Partnerships in Weather and Climate Services" (NRC 2003) recommending that "The National Weather Service (NWS) should make its data and products available in Internet-accessible digital form." The NRC report said that such a transformation would increase the availability and usefulness of these data. This seems like a reasonable conclusion given the Internet's 24/7 access and the Worldwide Web's reach directly into individual homes and offices around the globe. Once the data are delivered to where they are needed, their digital form enhances their usefulness by enabling users to apply them to a wide range of problems. Possible applications of the data, with their envisioned low cost-of-entry, include displaying them in personal web pages, redistributing them in a value added package, and integrating them in weather-sensitive decision support systems. Each of these examples suggests an era where machine-to-machine digital services allows the NWS to disseminate its data more quickly, efficiently, in forms suitable for additional processing, and to a larger user base.

To realize the benefits of transferring digital data to a user's computer, a data provider needs to deploy a system, in this case a web service, to handle the exchange of data. Fielding such a web service entails making a number of technical decisions. For example, a web service developer needs to choose a communication protocol, messaging format, and data encoding standard from several options in each component category. Web service users typically have preferences regarding these technical details, so picking a particular set of protocols and formats can affect service usage.

Over the past 4 years, the NWS has begun to develop an understanding of user preferences for digital services in general and service messaging formats and encoding standards in particular. This experience came as a result of deploying four web services designed to disseminate National Digital Forecast Database (NDFD) data. The four services are the 1) NDFD Simple Object Access Protocol (SOAP; W3C 2003) Service, 2) Web Feature Service (WFS; OGC 2005) Service, 3) Representational State Transfer (REST; Fielding 2000) Service, and 4) General Regularly-distributed Information in Binary Version 2 (GRIB2; WMO 2001) File Download Service. This paper explores the lessons learned from deploying and operating these services and the implications the new understanding has for other public web service developments.

2. DIGITAL DATA DESCRIPTION

The NDFD (Glahn and Ruth 2003) represents the official NWS 7-day forecast for over 40 sensible weather elements like maximum temperature, sky cover, probabilistic tropical cyclone wind speed, and probability of hail. The forecasts are prepared in a distributed fashion at NWS Weather Forecast Offices (WFO) and Centers throughout the United States and its territories. The portion of the database created at each office/center is transmitted to a central server where the digital data are merged into a single database. The aggregated forecast data are then encoded in GRIB2 and made available for dissemination via web services (See Fig. 1).

* *Corresponding author address*: John L. Schattel, W/OST24, 1325 East-West Highway, Silver Spring, MD, 20910; e-mail: John.Schattel@noaa.gov.

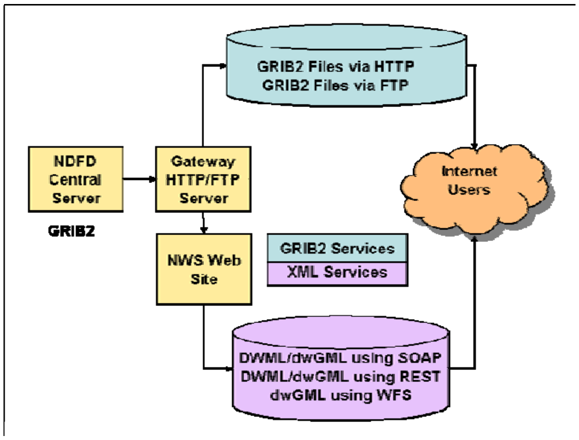

Figure 1. NDFD web services.

The forecast for a given weather element can be thought of as a grid of values with each value corresponding to a point on the Earth's surface. There are grids for separate geographical areas (e.g. Coterminous United States (CONUS), Alaska, Hawaii, Puerto Rico, and Guam). Values in a grid are valid, depending on the weather element, for either a specific time instance or span of time during the forecast period. The number of forecast projections in the forecast period also varies by weather element. A complete list of weather elements and their projections/valid times is available in the NDFD Technical Information Page (NWS 2008e).

3. SERVICE DESCRIPTION

Drawing from NRC (2003), the NWS has incorporated into its strategic plan (NWS 2005) a goal to make its forecasts available in digital form employing existing and new technologies to facilitate the data's use. The NDFD and its suite of web services represent an attempt to achieve that goal. The four web services available to NWS users possess a combination of communication protocol, messaging format, and data encoding standard designed to support the preferences of different user communities. Table 1 summarizes the components used by each service. The following sections describe the four web services, the technology components they use, and why those technologies might be a good fit to user preferences.

3.1 NDFD GRIB2 Download Service

In December 2002, the NWS deployed the NDFD GRIB2 Download Service (GS; NWS 2008e) which made human-created digital forecast data available to the public via the Internet. GS allows a web service client to retrieve GRIB2-encoded files using File Transfer Protocol (FTP) and Hypertext Transfer Protocol (HTTP). GRIB2 represents a well defined World Meteorological Organization (WMO) standard using binary encoding and data compression to reduce file size. As a result, GS users require a specialized decoder to unpack the compressed data. The NWS provides service users with a free application called degrib (NWS 2008b) that enables users to extract the meteorological data from its numerical weather forecasts as well as NDFD GRIB2 files.

GS exposes 135 megabytes (MB) of GRIB2-encoded NDFD data to users. Since users may find it impractical and unnecessary to download a file containing the entire database, GS partitions NDFD data into a number of directories containing files that subset the database based on data location and valid time. The location dimension of the directory structure allows users to obtain data for a specific geographic area (e.g. CONUS, Hawaii, and Guam) called a sector. The CONUS sector is further broken into 16 subsectors (e.g. Northeastern United States). The data are also stratified by valid time so that for each sector there is a subdirectory containing data for days 1-3 of the forecast period, days 4-7, and day 8 and beyond. Within a given valid time subdirectory, a user finds a GRIB2-encoded file for each element appropriate for that sector and valid period.

The target audience of GS includes users requiring a significant geographical portion of the NDFD database. These users include customers needing input for regional and national products and services and customers having data access

Table 1. NDFD service protocols and formats.

Service	**Comm. Protocol**	**Messaging Format**	**Data Encoding Standard**
NDFD SOAP	HTTP	SOAP	DWML dwGML
NDFD WFS	HTTP	WFS	dwGML
NDFD REST	HTTP	----	DWML dwGML
NDFD GRIB2	HTTP FTP	---- ----	GRIB2

requirements that would be inefficient to satisfy if using other NDFD services. GS users often need the data compression offered by GRIB2 and are comfortable decoding that weather community standard.

3.2 NDFD SOAP Service

A year and a half after deploying the GS, the NWS implemented a web service using SOAP messages sent over an HTTP connection. The SOAP-based service allows a user's client application to retrieve data for 1) one or more gridpoints, 2) a specific time or range of times, and 3) for a list of particular forecast elements. The ability to access NDFD values at this high granularity fills a niche left open by GS. For users wanting to display their local forecast on a personal web page, the NDFD SOAP Service (SS; NWS 2008d) appears as an efficient alternative to downloading multi-MB sized files.

In addition to exposing small subsets of data to users, the SS also exploits standards familiar to the web application development community. SOAP is a widely used Worldwide Web Consortium (W3C 2003) specification for Internet messaging. With built-in support in many popular programming languages, SOAP and its companion Web Services Description Language document make it easy for users to integrate the service into their web applications. In some cases, integrating the service is automated by their development tools.

The NDFD data contained in the SOAP response is returned to the user encoded in Extensible Markup Language (XML). XML, another W3C specification, is the de facto standard for data exchange over the Internet and uses a schema to define a list of tags and their order. The SOAP service returns NDFD data encoded according to the Digital Weather Markup Language (DWML; NWS 2008a) and Digital Weather Geography Markup Language (dwGML) schemas. DWML is an NWS developed schema while dwGML is an application schema of Geography Markup Language (GML; OGC 2004). DwGML was added to the SOAP service in September 2007 to provide users access to NDFD data encoded in a standards-based schema. Using these schemas helps lower the cost of entry by allowing web developers to use familiar decoding tools (parsers) to extract NWS data from the service's response.

3.3 NDFD WFS Service

In an effort to replace NWS developed DWML with an industry standard schema and to better serve the Geographical Information System (GIS) user community, the NWS deployed the NDFD WFS Service (WS; NWS 2008f) in May 2007. The WS uses the Open Geospatial Consortium (OGC) WFS specification for its messaging protocol to improve interoperability with GIS clients. The data returned by the WS is encoded in dwGML which is an application schema of GML. Like WFS, GML is also an OGC specification and was developed for the GIS community. The NWS application schema provides XML tag names that fit the NDFD data but retains the GML types so a GML-aware client can easily process the encoded data.

Both the WFS and GML specifications have multiple versions that user clients have implemented. To accept requests for multiple WFS versions and provide the appropriate GML-encoded response, the WS supports WFS 1.0, WFS 1.1, GML 2.1.2, and GML 3.1.1 schemas. The WS also attempts to address the complication involving the partial implementation of the GML specification. Because GML is a large and complex schema, a given client may only implement support for certain GML types and may make processing assumptions that limit the client's ability to understand a generic WFS response. To accommodate these varying implementations, the WS, encodes its data in five features structures with different GML data types. Two of the feature structures are compatible with the GML2.1.2 schema while three feature structures are compatible with the GML3.1.1 schema.

When a user requests one of the five dwGML-encoded feature types, he/she must supply filter information to achieve the granularity of the SS (returns data for a single gridpoint, at a single time, and for a single element). The filters are encoded in the request using the OGC Filter Encoding Implementation Specification. If the user does not supply a filter, the response includes the most recent forecast data for points representing 193 cities across the United States.

3.4 NDFD REST Service

Some NDFD users prefer to avoid the messaging layer that comes with the SS and WS. These users requested an interface that used a REST approach that omits the standardized messaging protocol in favor of a simpler resourcebased approach. For these users, the NWS established the NDFD REST Service (RS; NWS 2008c) which allows users to submit requests via the HTTP Get method. RS requests take the form of a query string that identifies the service and provides the same inputs expected by the SS. RS users do not rely on standard tools to assist with integrating the service into their application. Rather, they must review service documentation to understand how to submit a valid request. Once the request is submitted, the RS returns DWML- or dwGML-encoded data just like the SS.

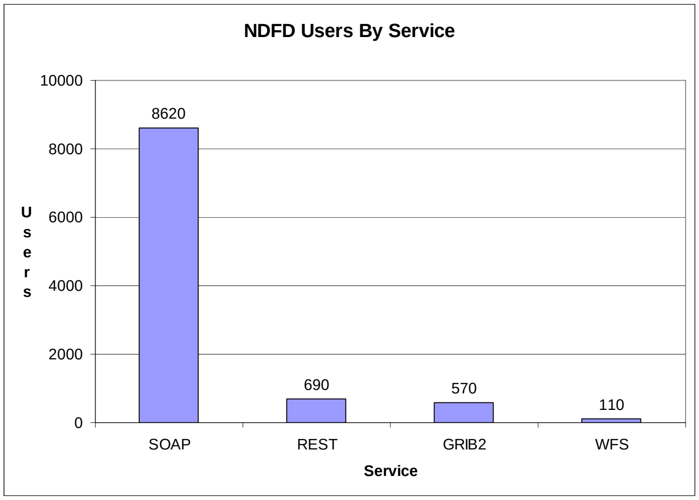

Figure 2. Number of users by NDFD service.

4. SERVICE USAGE

The number of users accessing an NDFD service is one of the principle metrics for measuring user preference. To count the number of users, each service logs the Internet Protocol (IP) address of the computer making the service request. The challenge is that a user may access the service with more than one IP address. To avoid over counting users, each IP address is resolved to a domain, or machine, name using an online domain name lookup service. The resulting user names are compared and similar names are combined to represent one user. The limitation of this approach is that multiple users who access the service through the same internet service provider or from behind a shared firewall can appear as the same user. This fact results in an underestimation of the number of users. As a result, user numbers in this paper should be considered a minimum value for the actual number of users.

The estimated number of users for each service is shown in Fig. 2. The number of users varies considerably across the four NDFD services with the SS having at least 8,620 users. This value compares to 690 RS users, 570 GS users, and 110 WFS users. With an order of magnitude more users, the SS is judged to be the preferred means of accessing NDFD data.

The popularity of the SS also shows up in the number of "downloads" from the service. As can be seen in Fig. 3, the SS, with its 1.7 million downloads per day, satisfies an order of magnitude more requests than the other services. Currently, the RS serves 106,000 downloads per day followed closely by the GS at 94,000 downloads per day. The NDFD WFS service trails the other services with 130 downloads per day.

The fact that the SS supports more users and downloads than the other services likely derives from the service's use of standards like XML and SOAP. XML and SOAP are both taught in universities around the world. As a result, there are a large number of developers familiar with these technologies and the tools that make their use easy. Conversely, there are relatively few developers familiar with community standards like GRIB2 and WFS, and newcomers like REST.

Another factor contributing to the relative differences in usage between the WS and other

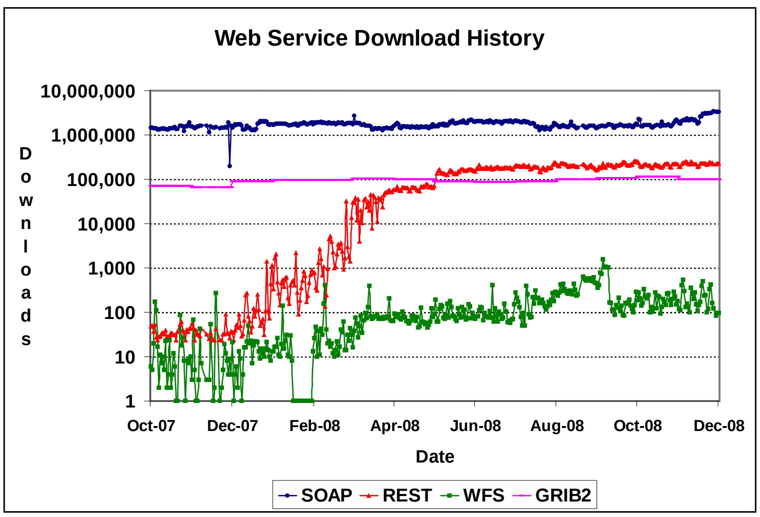

Figure 3. NDFD download history.

NDFD services is maturity of the service's standards. For mass appeal, a service needs to use a stable set of standards. XML, SOAP, and GRIB2 have been stable since 2000. During the same period (See Fig. 4), significant changes have been made to both the GML and WFS specifications. This lack of maturity in the OGC standards has been a challenge for client developers and has kept the population of potential WS users small.

5. USER ACCESS PATTERNS

NDFD Internet web services are available to users continuously. But users will likely only access the service when their particular application requires data. In Figs. 5–7, the percent of the days that a user accesses the service during a sliding 90-day window is presented. Both the RS and SS have sizeable percentages of daily users, 27% and 22% respectively, while the WS has a much smaller 6%. At the other end of the frequency-of-use spectrum, the WS has nearly 75% of its users as either one time or infrequent users. For the SS and RS, the infrequent user population is smaller at 42% and 26% respectively.

User counts presented in Section 4 include users that accessed the service just once. Given the ease with which clients can access a web service, it is not surprising that some users will access the server to experiment with the service but not have a use for the weather forecast data the service disseminates. Combining the “Daily” and “Regular” user categories in Figs. 5–7 indicates that the number of users leveraging NDFD data may be closer to 2451, 220, and 6 for the SS, RS, and WS respectively.

In addition to the usage pattern describing the percentage of days a user accesses NDFD services, a user may also retrieve data multiple times during a given day. Figs. 8–10 illustrate how often users prefer to download data. Given that users are free to download data as frequently as they like, it is somewhat surprising that over 85% of users retrieve data less than or equal to once per hour. One access each hour would allow an individual user to retrieve all the locations and weather elements they need and get any of the hourly updates to the service database. But some users, with customer facing services, may respond to user requests of their own. These users seem to make up some of the 7%-14% of users who download data more frequently. Some of the more-frequent-than-hourly requests also include users who do not know they can combine multiple gridpoints and multiple elements in a single request. Figs. 8–10 also show a small number of users who exceed the NWS guideline that a user submits no more than 720 requests per hour to gridpoint-based services like SS, RS, and WS. These users should probably be using the GS.

6. SUMMARY

Between late 2002 and early 2008, the NWS deployed four web services that allow users to retrieve NDFD data. Each service uses a different combination of communication protocol, messaging format, and data encoding standard. Based on the number of users of each service and the number of downloads those users make, NDFD users prefer to use a service that employs HTTP, SOAP, and data encoded in a simple XML schema tailored to the data. This preference is probably due to the shear number of users familiar with SOAP and XML and the widely available tools that make using them easy. It also helps that these standards are relatively stable.

In addition to seeking a service that provides a low cost of entry for them, NDFD web service users also factor in their data volume requirements. Users needing data for only a few gridpoints will choose one of the point-based services which include SS, RS, and WS. Users requiring large portions of the database will find it efficient to use GS. NWS web service experience suggests that deploying only one service to disseminate environmental data to a diverse user base may not provide

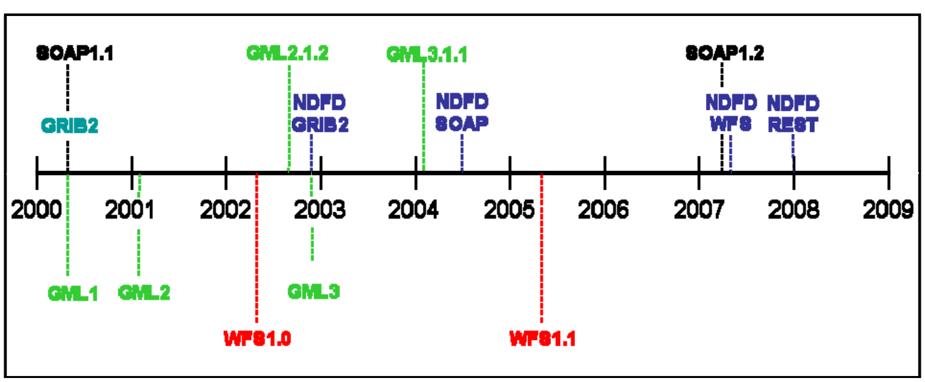

Figure 4. NDFD service and standards deployment chronology.

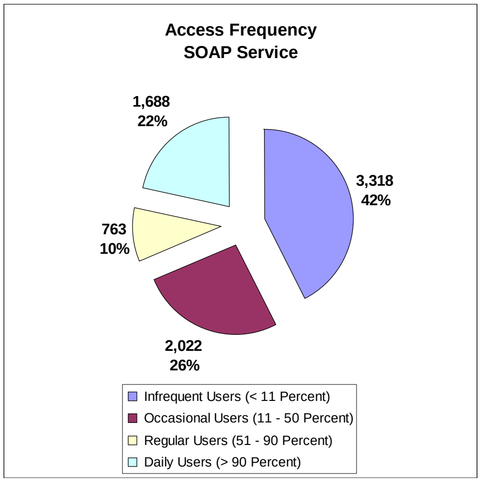

Figure 5. SOAP Service access frequency.

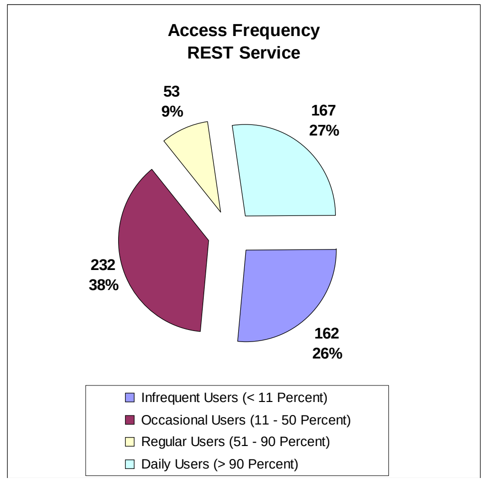

Figure 6. REST Service access frequency.

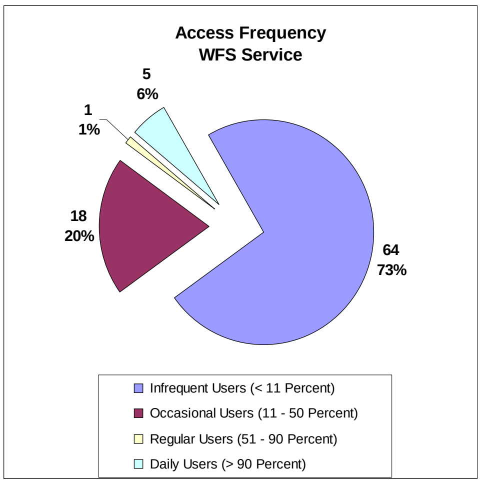

Figure 7. WFS Service access frequency.

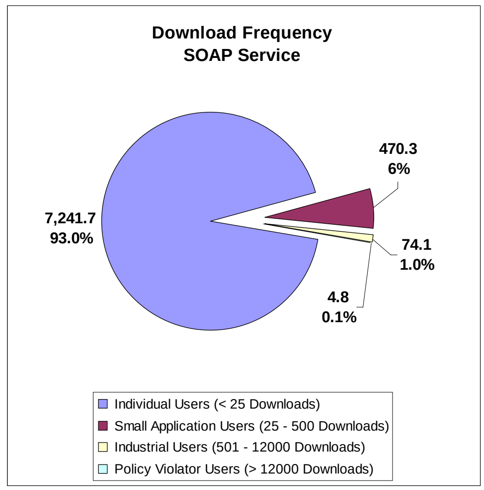

Figure 8. SOAP Service download frequency.

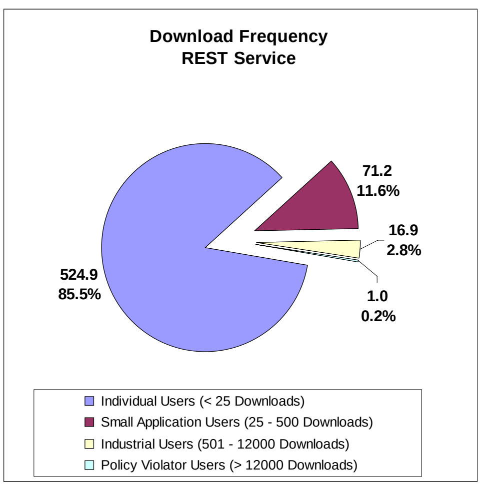

Figure 9. REST Service download frequency.

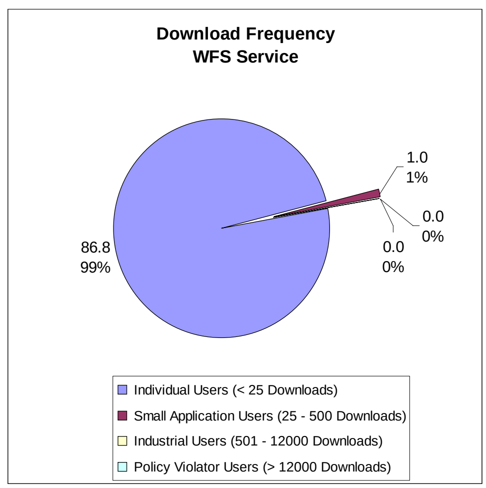

Figure 10. WFS Service download frequency.

the optimal solution. Rather, two or more services addressing the preference of different user communities will likely serve users best. With over 10,000 users and 2 million downloads per day among them, NDFD web services demonstrate that the NWS has addressed user preferences for accessing digital forecast data. Together these services have moved the NWS closer to meeting its goal to improve the value of its data by making them available in digital forms.

7. REFERENCES

Fielding, R. T., 2000: Architectural Styles and the Design of Network-based Software Architectures. Dissertation, University of California, Irvine, 162 pp.

Glahn, H. R., and D. P. Ruth, 2003: The new digital forecast database at the National Weather Service. Bull. Amer. Meteor. Soc., **84**, 195–201.

NRC, 2003: *Fair weather, effective partnership in weather and climate services*. National Academies Press, 220 pp.

NWS, 2008a: Digital Weather Markup Language Specification. [Available online at http://weather.gov/mdl/XML/Design/MDL_XML_Design.htm]

____, 2008b: NDFD GRIB2 Decoder Page. [Available online at http://www.weather.gov/mdl/NDFD_GRIB2Decoder/]

____, 2008c: NDFD REST Service. [Available online at http://weather.gov/forecasts/xml/#nonSOAP]

____, 2008d: NDFD SOAP Service. [Available online at http://weather.gov/forecasts/xml/]

____, 2008e: NDFD Technical Information. [Available online at http://weather.gov/ndfd/technical.htm]

____, 2008f: NDFD WFS Service. [Available online at http://weather.gov/forecasts/xml/OGC_services/]

____, 2005: Working Together to Save Lives, National Weather Service Strategic Plan for 2005 - 2010, 36 pp. [Available online at http://weather.gov/sp/NWS_strategic_plan_01-03-05.pdf]

OGC, 2004: OpenGIS® Geography Markup Language (GML) Implementation Specification. [Available online at http://opengeospatial.org/standards/gml]

____, 2005: Web Feature Service Implementation Specification. [Available online at http://opengeospatial.org/standards/wfs]

W3C, 2003: SOAP Version 1.2 Part 1: Messaging Framework. [Available online at http://w3.org/TR/2003/REC-soap12-part1-20030624/]

WMO, 2001: Manual on Codes, FM 92-XII Part II GRIB - General Regularly-Distributed Information in Binary Form, Vol. 1.2, 306.