University of Reading

Communicating thematic data quality with web map services

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work.
Published version at: http://dx.doi.org/10.3390/ijgi4041965
To link to this article DOI: http://dx.doi.org/10.3390/ijgi4041965

Publisher: MDPI Publishing

www.reading.ac.uk/centaur

CentAUR

Article

Communicating Thematic Data Quality with Web Map Services

Jon D. Blower [1],*, Joan Masó [2], Daniel Díaz [2], Charles J. Roberts [1], Guy H. Griffiths [1], Jane P. Lewis [1], Xiaoyu Yang [3] and Xavier Pons [4]

[1] Department of Meteorology, Harry Pitt Building, University of Reading, Reading RG6 6BB, UK;
E-Mails: c.j.roberts@reading.ac.uk (C.R.); guy.griffiths@reading.ac.uk (G.G.);
j.p.lewis@reading.ac.uk (J.L.)

[2] Grumets Research Group, CREAF, Edifici C, Universitat Autònoma de Barcelona, 08193 Bellaterra,
Catalonia, Spain; E-Mails: Joan.Maso@uab.cat (J.M.); d.diaz@creaf.uab.cat (D.D.)

[3] Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China;
E-Mail: kxy@cnic.cn

[4] Grumets Research Group, Dep Geografia, Edifici B, Universitat Autònoma de Barcelona,
08193 Bellaterra, Catalonia, Spain; E-Mail: Xavier.Pons@uab.cat

* Author to whom correspondence should be addressed; E-Mail: j.d.blower@reading.ac.uk;
Tel.: +44-118-378-5213.

Academic Editor: Wolfgang Kainz

Abstract: Geospatial information of many kinds, from topographic maps to scientific data, is increasingly being made available through web mapping services. These allow georeferenced map images to be served from data stores and displayed in websites and geographic information systems, where they can be integrated with other geographic information. The Open Geospatial Consortium's Web Map Service (WMS) standard has been widely adopted in diverse communities for sharing data in this way. However, current services typically provide little or no information about the quality or accuracy of the data they serve. In this paper we will describe the design and implementation of a new "quality-enabled" profile of WMS, which we call "WMS-Q". This describes how information about data quality can be transmitted to the user through WMS. Such information can exist at many levels, from entire datasets to individual measurements, and includes the many different ways in which data uncertainty can be expressed. We also describe proposed extensions to the Symbology Encoding specification, which include provision for visualizing uncertainty in raster data in a number of different ways, including contours, shading and bivariate colour maps. We shall

also describe new open-source implementations of the new specifications, which include both clients and servers.

Keywords: GIS; data quality; uncertainty; WMS; OGC; INSPIRE; GEOSS; SDI

1. Introduction

The Web Map Service standard (WMS, [1]), published by the Open Geospatial Consortium and standardized by ISO as ISO19128:2005 [2], has been one of the most successful standards enabling the interoperability of geographic information. It allows clients to request georeferenced map images representing a wide variety of data sources, together with accompanying metadata. The standard is implemented in a wide range of open-source and commercial software, including GeoServer, MapServer, ArcGIS Server, Cadcorp SIS and ncWMS [3] (see http://www.opengeospatial.org/resource/ products/compliant for a list of certified compliant software). An overview of the standard can be found in Section 5.1 below. WMS is used as the visualization standard in many spatial data infrastructures such as INSPIRE [4] and the Global Earth Observation System of Systems (GEOSS) [5].

A common concern with Spatial Data Infrastructures (SDIs) in general, is that the user is frequently not provided with information regarding the quality of the underlying data. This is particularly important in the field of scientific data, in which the user often wishes to know about the uncertainty of the data. Although a number of methods are in use for visualizing uncertain data (such as those suggested by the INTAMAP project for visualizing uncertainty of interpolated maps [6], some examples for mapping uncertainty of weather ensemble data [7,8] and several others described in [9]) these are mostly not supported by WMS implementations, or are not supported in a standardized, consistent fashion; furthermore the relevant symbology standards (e.g., Symbology Encoding (SE), [10]) do not provide the required flexibility to accommodate these more complex visualization types, particularly for raster data. Some solutions for using WMS and SE in uncertainty visualization for vector data are given in [11], which focuses on use cases in spatial planning.

In this paper we describe a new profile of WMS for data quality (WMS-Q), which provides a number of mechanisms for using WMS to communicate data quality at a number of levels, from the dataset level to the level of individual samples (pixels). WMS-Q is entirely compatible with the current version of the WMS standard (1.3.0) and considers a number of underlying data types (including continuous and categorical data). Section 2 introduces the concept of "data quality", Section 3 clarifies the terminology we will use in the paper, Section 4 explores current means of representing data quality, Section 5 describes the design of WMS-Q, Section 6 briefly describes some current implementations of the profile, Section 7 describes how WMS-Q has been integrated into a "quality-aware" SDI, and finally Section 8 identifies areas of future work.

2. What is "Data Quality"?

Data quality is a concept that is difficult to define precisely. It means different things to different communities and it is sometimes confused with the amount of information. It is defined by the ISO

8402:1994 [12] standard as the "totality of characteristics of a product that bear on its ability to satisfy stated or implied needs". In the geospatial domain, it includes many measures including spatial accuracy, temporal accuracy, consistency, completeness, scope and attribute accuracy. There is a general consensus that quality is a *subjective* concept and each user defines a dataset as having good quality when it fulfils his expectations, that is, it *fits his purpose*. Unfortunately, data producers cannot anticipate the expectations of all future users. Thus, in practice they try to quantify some aspects by comparing the dataset with other reference sources that are accepted to be correct and provide a numeric quantification of the discrepancies in terms of uncertainty metrics and error statistics. Alternatively, they create a specification for their product, and for each dataset, they provide a confirmation of the product specifications conformity [13]. When these approaches are applied to the data, a quality indicator [14] is generated and attached to the metadata.

Data quality can be applied to different components of the data: spatial (e.g., positional accuracy), temporal (temporal accuracy), thematic (e.g., completeness, thematic accuracy, classification correctness, *etc.*) [14–16]. Some aspects covered by this paper apply to all of the above but the focus is on thematic accuracy.

3. Terminology Used in This Paper

Different communities use different terms to represent similar concepts, or the same term for different concepts. The WMS-Q profile can be applied in many different communities, including Earth Observation, climate science, meteorology, oceanography, terrestrial studies and more. Therefore we must carefully define the terms we use in this paper, which may be different from common usage in particular communities:

Dataset: The term *dataset* is used in its most general sense of "a logical collection of data". The information in a dataset may be recorded physically in a single file or a set of files and may represent a single snapshot in time, or time-evolving information. The choice of how to group data into datasets is highly community-specific; typically, data providers will decide how to group information into logical units that make sense to them and their users. Discovery metadata [17,18] is typically generated at the dataset level.

Variable: A *variable* is a measured quantity, such as temperature, velocity or a land use classification. It can be observed or calculated. In this paper we use "variable" to represent the concept; the values and their uncertainty are recorded in *fields* (see below).

Sample: A *sample* is an individual measurement or observation of a single variable. A pixel in an Earth Observation image, or a grid cell in the results of a numerical simulation, may contain several samples, one for each variable that is recorded in that pixel or cell.

Component: If the values of a variable within a dataset are uncertain, then each sample of this variable comprises more than one piece of information (for example, components could be the "most likely" value of the variable and the upper and lower confidence range representing its uncertainty). Each piece of information is known as a *component*. Different ways of expressing uncertainty are described in Section 4.3 below.

Field: A *field* is a collection of values of a single component. Typically this will be recorded in a data file as a multidimensional array of scalar (numeric) values.

We illustrate these concepts through two examples.

Example 1: The Sea Surface Temperature *dataset* from the European Space Agency's Climate Change Initiative (CCI-SST, [19]) provides measurements of two *variables*: sea surface temperature and sea ice area fraction, recorded on a global grid at 1/20 degree (~5 km) resolution. The *sample* measurements of both variables are uncertain and therefore the values of each variable are expressed as two *components*, one representing the mean value and one representing its variance. The dataset therefore contains four scalar *fields* (two for each variable).

Example 2: The Landcover Classification of Landsat Barcelona-Girona scenes [20] *dataset* provides information on the land cover of that region of Catalonia, derived from satellite imagery. Each *sample* (*i.e.*, pixel) within the dataset contains the best estimate of the land cover type (expressed as a classification code), some alternative classification values plus several measures of confidence that the classification is correct. The dataset therefore contains one *variable* (land cover type) expressed as a number of scalar *fields*: some categorical (classifications) and some continuous (confidence measures).

4. Encodings of Data Quality

4.1. The ISO Suite of Standards

In current SDIs, data quality information can be included within structured metadata documents that describe geospatial datasets [21]. There are some metadata standards produced by International Standards Organization Technical Committee 211 (ISO-TC/211), the Federal Geographic Data Committee (FGDC), European Committee for Standardization Technical Committee 287 (CEN/TC-287), the International Cartographic Association (ICA) and others that include some or all of the quality components. The ISO19115:2003 [22] model (and its counterpart XML encoding, ISO19139:2007 [23]) is perhaps the most commonly-adopted standard for describing geospatial metadata. ISO19113:2002 [24] defined the quality principles that were included in the later standard ISO19157:2013 [14], which also provides details about quality measures and the metadata elements for quality documentation that can be incorporated to the general ISO19115 geospatial metadata model. The European SDI that results from the INSPIRE directive is one example of a multinational data catalogue that relies on these standards.

In addition to quality indications, lineage or provenance information is also considered a component that helps users to assess fitness-for-purpose. Provenance information includes information about the producer, which can influence the user trust, and also formal documentation about the processes, algorithms and data sources involved in the elaboration of a dataset [25]. ISO19115 incorporates provenance information through the LI_Lineage structures.

4.2. Building on the ISO Model

The GeoViQua project (http://www.geoviqua.org) built on these standards to provide new mechanisms to exchange quality information from both the data producer's and the data user's point of view. The GeoViQua Producer Quality Model (PQM) builds on ISO19157 and includes new support for discovered issues, workarounds, usage, citations, goodness-of-fit statistics and validation. The PQM is complemented by the User Quality Model (UQM), which captures user feedback on datasets, including numerical ratings with text justification; user comments and reports of usage and problems identified;

citation of publications; tags to facilitate data discovery; supplementary quality reports; and information on any spatial, temporal or thematic focus of the feedback. The schemas defining both the PQM and UQM can be found at http://schemas.geoviqua.org/GVQ/4.0/.

XML (e.g., ISO19139 documents) is a machine-readable format that is not well suited for direct presentation to the user. Tools are required to both present metadata and edit metadata [26]. These tools are able to produce human readable reports, e.g., in the form of HTML pages [27] using different strategies, XSLT being a popular choice for turning XML metadata document sources into human-readable HTML. Even though a human readable report is easily interpreted by an operator that is familiar with it, the work load to compare tens or hundreds of documents is impractical. More sophisticated visualization tools can be elaborated to represent the content of a metadata in a simplified and compact image that represents the existence of some information in the metadata by colouring facets of a GEO label [28] or by representing side by side some metadata values in a normalized scale such as a parallel coordinate plot [29].

4.3. Vocabularies for Describing Data Quality

The ISO standards provide a framework for structuring data quality information, but need to be used in combination with comprehensive controlled vocabularies that enable the precise description of quality information. As we shall describe in Section 5, these vocabularies can be used to convey the precise semantic meaning of WMS Layers.

UncertML (http://www.uncertml.org/) is a vocabulary of terms that can be used to describe the dispersion of values of uncertain variables and the components that comprise them. In UncertML, uncertainty can be represented in three main ways:

1. As individual **realisations** of an uncertain variable. For example, the uncertainty of a temperature field could be expressed by running a simulation ten times under different conditions, and recording the results of each simulation in a data file as a separate field.
2. As **summary statistics**. Instead of recording each realisation individually, the data provider may choose to calculate statistics describing the spread of values at each measurement location. For example, the spread may be expressed as a mean field and a variance field.
3. As **probability distribution functions** (PDFs). This is the most powerful representation of uncertainty and gives a functional form for the probability of a variable having a given value at a given location. In a data file, a PDF may be represented by specifying the functional form of the PDF as metadata (e.g., Gaussian or log-normal) and giving the values of each parameter of the distribution as a separate field.

In this way, a set of scalar fields can be suitably "tagged" with UncertML terms to allow the user to understand that they represent the components of an uncertain variable and, when taken together, provide a measure of the variable's uncertainty.

The QualityML vocabulary [30] builds on UncertML [31] and ISO19157 to provide a more comprehensive suite of terms that describe a broader range of quality measures. These terms include statistics that summarise the uncertainty of an entire field (e.g., the root-mean-square error) and terms that describe uncertainties in categorical data (e.g., the commission error or the Kappa coefficient).

The NetCDF-U specification (currently under discussion within the Open Geospatial Consortium, [32]) provides a mechanism for using the richness of the UncertML vocabulary within NetCDF files. QualityML could also be used in NetCDF for the same purpose.

5. Design of WMS-Q

This section describes the main features of the WMS-Q profile. The current version of the full WMS-Q profile document can be found at http://www.geoviqua.org/Docs/WMS-Q_v2.d_final.pdf.

5.1. Overview of WMS

We briefly describe here the main relevant features of the WMS standard (version 1.3.0), in order to aid understanding of the WMS-Q extensions we describe in the sections below.

The data provided by a WMS server are divided into *Layers*, which represent the basic units of information. The server provides metadata about these Layers through a machine-readable service metadata document (also called the "Capabilities" document) that can be requested by clients through the *GetCapabilities* operation. All Layers have a human-readable Title field. Further metadata can be provided through external documents (both human-readable and machine-readable), which are linked to Layers in the service metadata document through the *MetadataURL* tag.

Georeferenced images (maps) of the Layers are requested through the *GetMap* operation. There are two ways of controlling the appearance of a map image. The simplest method is for the server to advertise fixed *Styles*, each of which is identified by a name that can be requested in the *GetMap* request. A more sophisticated method can be provided using a WMS server that is enabled with the *Styled Layer Descriptor (SLD)* [33] extension. In this case, the client has much closer control over the appearance of the requested map image by sending the server an XML document encoded using the *Symbology Encoding (SE)* standard. Such a server is often referred to as an *SLD/SE-enabled* WMS server.

The WMS standard permits Layers to be arbitrarily nested within a service metadata document. Layers that have a <Name> property are *displayable* (which means that they can be requested in a GetMap operation). WMS permits any Layer to be displayable, even if that Layer contains child Layers. Layers that are not displayable do not have a <Name> and are usually used as a grouping mechanism for child Layers. A WMS service metadata document therefore contains a hierarchy of Layers. The WMS standard specifies rules (Table 7 in [1]) that govern the inheritance of properties of parent Layers by child Layers; these rules are important to the design of WMS-Q: see Section 5.6.

The *GetFeatureInfo* operation allows a WMS client to retrieve more information about a specified pixel in a map image. The WMS standard does not specify a particular format for the information that is returned by the server and so implementations can differ widely in their behaviour.

5.2. Design Goals for WMS-Q

The design goals of the WMS-Q profile were:

1. To maintain compatibility with version 1.3.0 of the WMS standard. In other words, WMS-Q aims to be a *profile* of WMS, not an *extension*. Therefore, instead of adding new functionalities, it provides a set of rules on how to use the existing mechanisms permitted by the WMS standard

to convey dataset, sample and variable level quality. In this way, standard WMS clients will be able to read information from a WMS-Q.

2. To re-use existing general methods for expressing data quality, where appropriate (e.g., UncertML, see Section 4.3 above), but to avoid methods that are highly specific to particular communities.

3. To be independent of any particular format or convention for data or metadata storage.

4. To focus on conveying the *thematic accuracy* of both categorical and continuous raster data (including its uncertainty), but allow techniques to be more widely applicable in future (e.g., to vector data).

5.3. Identification of Conformance to WMS-Q

In WMS 1.3 and previous versions, there is no standard mechanism at present for advertising to clients that a WMS service instance conforms to a particular profile. In WMS-Q we use a Keyword at the top level of the service metadata document, with the value "WMS-Q" in the vocabulary "http://www.geoviqua.org/def/doc/conventions/vocabulary". Sample Capabilities documents illustrating this are referenced in Section 5.6.

5.4. Dataset-Level Quality

In WMS-Q, datasets (see "Terminology", above) are represented by Layers that are not displayable, but act as containers for child layers representing variables. These Layers can be arbitrarily nested within other non-displayable ones, as chosen by the data provider, to impose a logical structure. (In this case the non-displayable Layers act as "folders" and typically appear as such in the legend or the layer list in WMS client software.)

Data quality is expressed at the Dataset level by associating the relevant Layer with a MetadataURL that points to a suitable, standardized metadata document (e.g., an ISO19115 dataset descriptor). However, we recommend that data quality information be associated more precisely at the Variable level where possible.

5.5. Variable-Level Quality

In WMS-Q, Variables are represented by Layers that are immediate children of Layers that represent Datasets. Quality information at this level is provided, as with Dataset-level information, by external documents linked through MetadataURL tags. The linked ISO19115 data metadata documents will have one or more DQ_DataQuality elements reporting overall quality indicators represented by estimators or statistical summaries (e.g. root-mean-square errors or confusion matrices). Since a ISO19115 document can be used to convey several DQ_DataQuality elements, it is important to keep a link of each DQ_DataQuality to the original WMS-Q layer. This can be achieved using the sub-elements of the MD_Scope, and document level = "layer" and levelDescription = *layer name*. Then a linkage of a MD_Distribution can be used to describe the link to the WMS service as a whole.

If the Variable does not contain sample-level quality information, then the Layer representing the Variable will be displayable and therefore the client can request map images of the Variable. If the Variable does contain sample-level quality information, its corresponding Layer must be

non-displayable, acting only as a container for child Layers that represent the *components* of the samples, as will explained in the next section. This Layer contains the Keyword "qualityCollection" from the "http://qualityml.geoviqua.org/1.0/" vocabulary. It may also contain Keywords from other vocabularies—for example, if the uncertainty of the Variable at the sample level is expressed using a set of summary statistics (see Section 4.3 above), the Layer may be tagged with the Keyword "statisticsCollection" from the UncertML vocabulary.

5.6. Sample-Level Quality

If the Variable contains sample-level quality information all sample components of a Variable are represented as direct children of the Variable's Layer (at the bottom level of the hierarchy, *i.e.*, each component is a leaf node in the tree of Layers). These child Layers are displayable and are given Keyword tags that describe the semantics of the component, using terms from vocabularies such as QualityML or UncertML where possible.

Although sometimes it is convenient to visualize these components (child Layers) individually, it is often highly desirable for a WMS client to be able to display a single map image that represents the variable as whole, perhaps by overlaying contours of some uncertainty metric (e.g., variance) over a raster image representing the mean field (see Figure 1). In some cases, this effect can be achieved by requesting separate images from the server (one for each component) and composing them on the client. However, this method does not work for some methods of visualizing data quality (for example, bivariate colour maps, see below). For this reason, WMS-Q servers provide an extra displayable Layer as a sibling of the other child Layers that represent components. This "special" child Layer represents the Variable as a whole and is given the Keyword "qualityComposition". It is the first Layer to be listed as a direct child of the Layer representing the Variable (in this case it can also have a metadata document describing its quality attached in the MetadataURL as above). It is intended that clients regard this "special" child Layer as a sensible default portrayal of the uncertain variable.
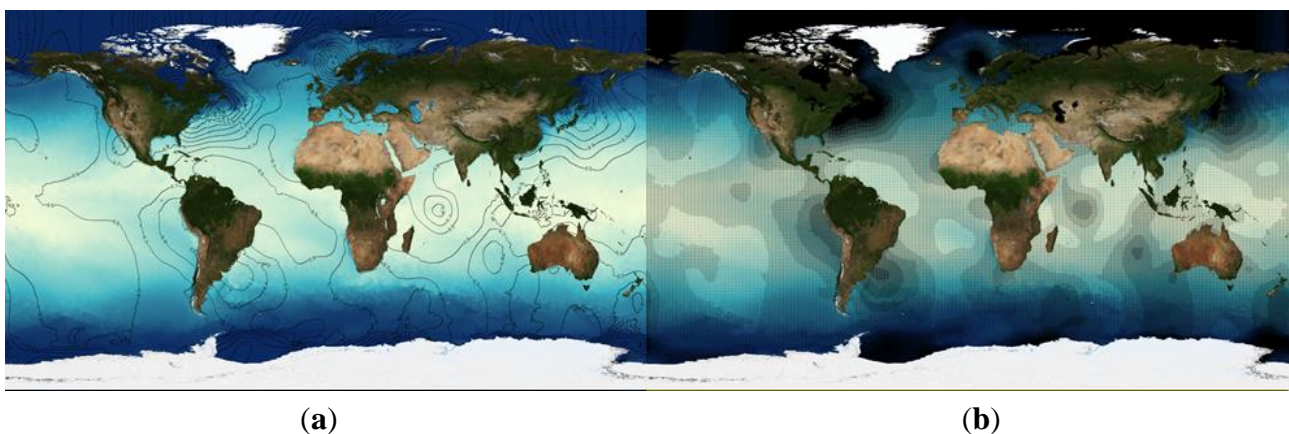

(**a**)   (**b**)

Figure 1. *Cont*.

(**c**)   (**d**)
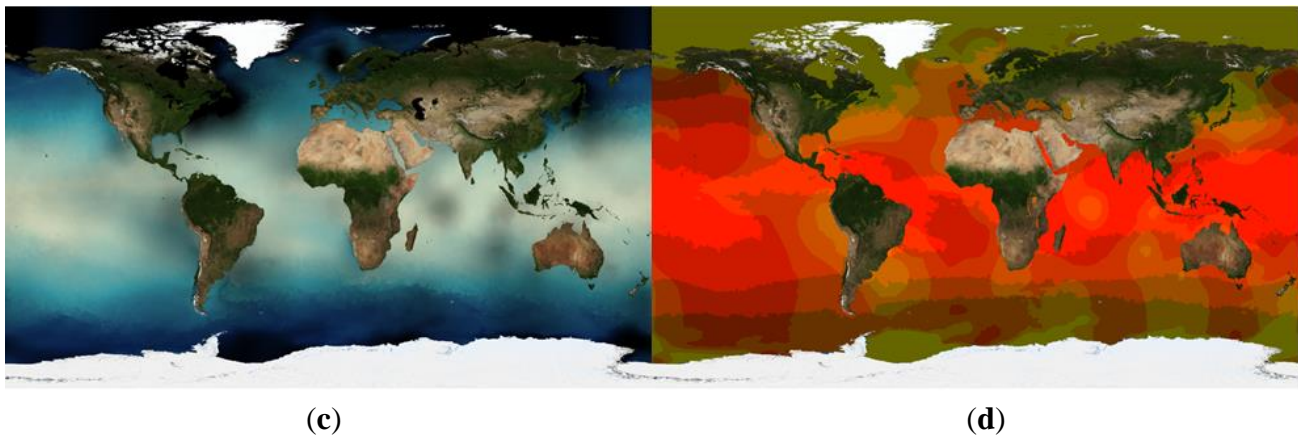
Figure 1. Sample visualisations of sea surface temperature and its uncertainty from the Sea Surface Temperature dataset from the European Space Agency's Climate Change Initiative (CCI-SST) dataset [19], generated by the ncWMS-Q software (see Section 6.1) using the proposed extensions to the Symbology Encoding (SE) specification (see Section 5.8). From top left: (**a**) temperature encoded as lightness of colour, with overlain contours of its uncertainty (variance); (**b**) uncertainty represented through successive levels of stippling, with denser stippling representing high uncertainty; (**c**) uncertainty represented as black shading; (**d**) use of a bivariate colour map, with temperature encoded as brightness and uncertainty encoded as colour saturation.

Note that early drafts of WMS-Q attempted to achieve the same goal using a displayable Layer at the variable level, which was allowed to contain children representing its components. However, the WMS standard states that all Styles are inherited by child Layers. It would be inappropriate for Styles that are suitable for whole variables (probably defining compositions of components) to be inherited by Layers representing individual components (for example, a style that plots contours of uncertainty on top of a colour-mapped field would not be suitable for portraying a single component). Because of this, the Layer representing the variable cannot be displayable, therefore we make the Layer representing the whole Variable a displayable child, allowing it to contain a different set of Styles from its individual components.

Example 1: In the CCI-SST dataset described in Section 3 above, there is a "sea surface temperature" variable. The uncertainty of each sample is expressed using two components representing respectively the mean and variance of the variable. This is expressed in WMS-Q using a non-displayable Layer representing the variable, which is tagged with the "qualityCollection" and "statisticsCollection" Keywords. This Layer has three children: the first child represents the variable as a whole, and provides styles that enable the client to visualize the mean and variance together in a single image (e.g., using contours of variance overlain on a colour-mapped image). The remaining two children represent the mean and variance components respectively, and are tagged with Keywords containing the terms "mean" and "variance" from the UncertML vocabulary. This is represented in graphical form in Figure 2. The complete WMS-Q service metadata document can be retrieved using this link: http://ncwms.geoviqua.org/wms?SERVICE=WMS&REQUEST=GetCapabilities&VERSION=1.3.0&DATASET=cci-sst. A demo client showing the data can be found here: http://ncwms.geoviqua.org/godiva2.html.
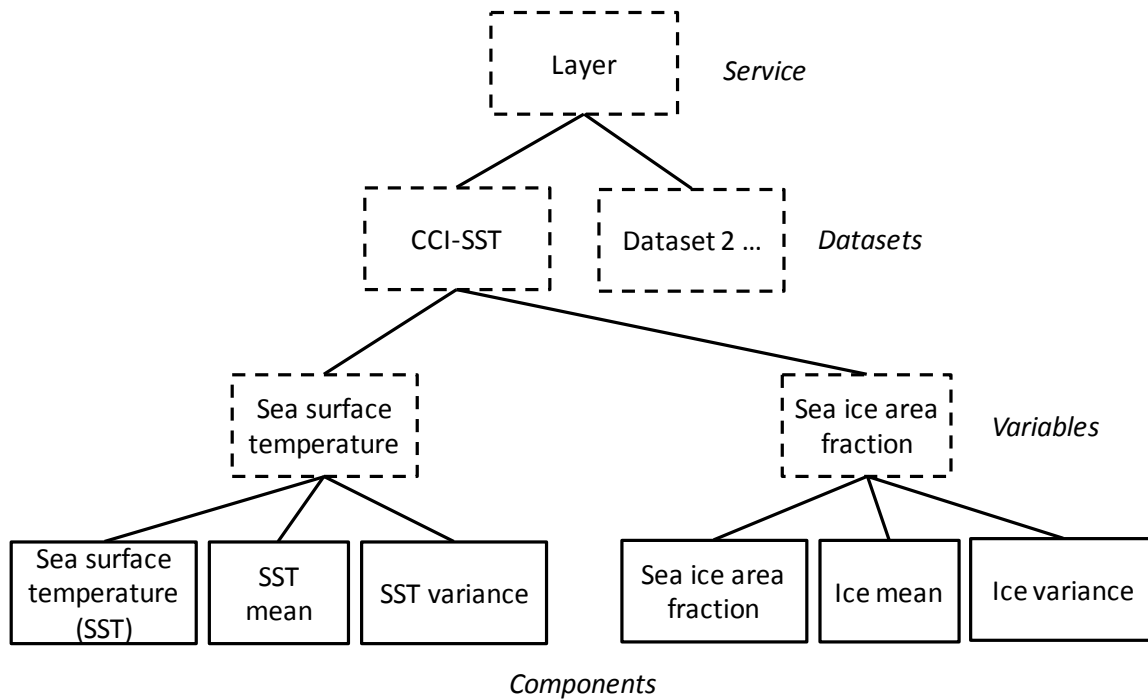
Figure 2. Schematic representation of the structure of Layers in a particular "quality-enabled" profile of Web Map Service (WMS-Q) service instance, illustrating the Service-Dataset-Variable-Component hierarchy. Each box is a Layer in the tree: blue boxes represent non-displayable Layers, whereas orange boxes represent displayable Layers. The derivation of this hierarchy is given in Section 5.5.

Example 2: We consider the "Landcover Classification" variable derived from Landsat in the Spanish National research project DinaCliVe. The classification method applied here combines common remote sensing classification techniques with the cadastre parcels and forces that each complete parcel is considered as a unity allowing statistical treatment of the pixels inside [20]. The variable and its uncertainty are expressed at the sample level using several components. Some of them are categorical components such as the class with the most presence in the parcel and the second and third class present in that parcel. Some other components are continuous such as fidelity, representativity, promiscuity, entropy, *etc.* (see http://qualityml.geoviqua.org for their definition). The variable layer has several children: the first child represents the component that most users will want to see: the class with the most presence in the parcel. This is tagged with the Keyword http://qualityml.geoviqua.org/1.0/values. The remaining classes help the user to understand the variability of the classification in the parcel and the confidence of the classification. All of them are tagged with the Keywords containing the right terms coming from the QualityML vocabulary. The complete WMS-Q service metadata document can be retrieved using this link: http://www.ogc.uab.cat/cgi-bin/GeoViQUA/WMSQ/MiraMon.cgi?SERVICE =WMS&VERSION=1.3.0&REQUEST=GetCapabilities. A demo client showing the data can be found here: http://wms-q-demo.geoviqua.org/geoviqua/wmsq/.

5.7. Behaviour of GetFeatureInfo

As mentioned in Section 5.1 above, the WMS standard does not specify the data returned by the server in response to a GetFeatureInfo request. In WMS-Q, if GetFeatureInfo is requested for a layer that represents a Variable as a whole, we recommend that the server responds with a graphical or structured textual representation of the probability distribution function of the value of the variable at the requested point, if possible. If this is not possible, the server should respond with a document containing the values of each of the sample components at that point. This aspect of WMS-Q has not so far completely defined and will be refined in future versions.

5.8. Extensions to the Symbology Encoding Standard

There are a number of ways of portraying uncertainties in raster data, either as continuous or categorical fields. A full description of all options is beyond the scope of this paper, but common strategies include:

- Portraying the "best estimate" as a colour-mapped image, overlain with contours showing a measure of data uncertainty (top left of Figure 1).
- As above, but uncertainty is represented through varying levels of stippling or texture (top right of Figure 1), as used in the Assessment Reports of the Intergovernmental Panel on Climate Change (e.g., [34]).
- As above, with uncertainty represented using black shading, the opacity of which increases with data uncertainty (bottom left of Figure 1).
- Using a bivariate colour map (e.g., [35]) in which the colour of a pixel is a function of two variables, the "best estimate" and the uncertainty (bottom right of Figure 1).
- Using glyphs (*i.e.*, small icons), the shape, size or colour of which can be mapped to different components of an uncertain variable. A special case of this is the use of "confidence triangles" (e.g., [36]), which visualize the estimated spread of data by dividing the image into squares, each of which is divided into two triangles. The lower triangle is assigned a colour representing the lower bound of the variable, and the upper triangle is coloured according to the upper bound of the variable. The contrast in colours between the two triangles gives a visual estimate of the uncertainty.

The current version of the SE standard allows for only a limited number of means of visualizing raster data using the RasterSymbolizer. Data values can be mapped to pixel colours and the opacity of the image as a whole can be altered. Other functions in SE are oriented around optical remote sensing imagery (with assumed red, green and blue channels) or digital elevation models. Contours, stippling, bivariate colour maps or glyphs are not supported in SE. We therefore propose extensions to SE to support these portrayal types. For reasons of space, a full description of the implementation of these new styles is beyond the scope of this paper, but are described in a draft document at http://www.geoviqua.org/Docs/03_ncWMS_Styling_Specification_1.0.pdf and are implemented in the ncWMS-Q software (see Section 6.1 below and a test client at http://ncwms.geoviqua.org/sldtest.html).

Clients can request that map images be generated using these new styles by crafting a SLD document using these new styles and passing it to the server. Alternatively, a server may offer named styles that implement these new styles, for convenience and simplicity (at the expense of some flexibility).

5.9. Mixing "Quality-Enabled" Data with "Non-Quality-Enabled" Data

In WMS-Q, not all Layers will necessarily have quality information (at the variable or sample level or either) attached. It is permissible to mix "quality-enabled" and "regular" layers in the same WMS service instance. The presence of MetadataURL links to metadata documents with quality indicators and the special keywords "qualityCollection", "qualityComposition" or other keywords in the UncertML or QualityML vocabulary allow the client to detect the quality-enabled layers in a WMS service.

6. Implementations

6.1. Server Implementations

The WMS-Q profile has been implemented in two systems in the context of the GeoViQua project. The ncWMS software [3] has been adapted to be compliant with WMS-Q, providing visualisations of raster data (e.g., numerical simulations and satellite imagery). Output from ncWMS-Q can be seen in Figure 1, showing data from the CCI-SST dataset of sea surface temperature and sea ice [19]. ncWMS is designed to require minimal configuration through automatic detection of metadata from source files in NetCDF format; this automatic configuration has been extended to the recognition of NetCDF-U data files, which contain information about data uncertainty. Therefore system administrators need only to format their data in NetCDF-U and the ncWMS-Q software will automatically provide a WMS-Q compatible endpoint. Manual configuration is possible in cases in which NetCDF-U files cannot be provided.

The Centre for Ecological Research and Forestry Applications (CREAF) has developed a WMS/WMTS server supporting WMS-Q based on MiraMon Map Service technology [37]. This allowed us to test the implementation for both continuous and categorical variables. For continuous variables, the Climatic Atlas of the Iberian Peninsula provides the following quality indicators regarding temperature: mean annual values, extrapolated areas, residual regression, and stability test. For categorical variables, the Landsat classified atlas was complemented by a set of quality indicators regarding land cover: first, second and third most present classes, fidelity, representativity, promiscuity, majority categories, entropy, and uncertainty.

The CREAF implementation of WMS in MiraMon Server is based on a prerendering process that serves both WMS and WMTS endpoints. In practice, this means that symbolization is associated to the layer by the operator publishing the layer. This has the advantage that rich and recommended visualizations can be offered to standard clients not fully aware of the particularities of the WMS extension. Multiple visualizations can be provided but users cannot control the symbolization parameters. The ncWMS-Q software supports the above proposed extensions to the SE standard (see Section 5.1 above) and therefore in this case the client has close control over the result of the image rendering process.

6.2. Client Implementations

ncWMS is packaged with an interactive map-based visualisation client called Godiva2 [38]. Godiva2 has been adapted for WMS-Q compliance, giving users the ability to interactively explore uncertain data on the web.

The MiraMon Map Browser has been enhanced making it aware of WMS-Q. It groups layers to this end following the service metadata document as defined by the layer hierarchy explained within this article. Only modifications on layer grouping and manual reordering by the user were required, proving how easily existing standard WMS clients are to adapt to WMS-Q. In addition, two different tools were added in order to allow users to easily compare parameters associated with a certain layer. The user is able to dynamically modify transparency of layers on runtime to see the layers below; and shift, allowing hiding dynamically one area of the top layer, so the ones below can be seen.

7. Integration of WMS-Q in a Quality Enabled Spatial Data Infrastructure

WMS-Q was developed in the context of the GeoViQua project (http://www.geoviqua.org), which also developed several other components of a "quality-enabled" SDI. Although these components were originally designed and developed as enhancements to the Global Earth Observation System of Systems (GEOSS), they are more widely applicable.

Data cataloguing and discovery was implemented by enhancing the Discover and Access Broker (DAB) to produce a quality-enabled version (DAB-Q, http://essi-lab.eu/do/view/GIcat). This harvests the WMS-Q service metadata ("Capabilities") documents and maps the different Layers there present into ISO 19115 documents, which form the basis of the data catalogue. Those metadata ISO documents can be then used in the GEO Portal (http://www.geoportal.org) to find datasets which are quality enabled. Thus, when a WMS-Q enabled dataset is found in a search, in addition to all the current links presented, links to the different variables and their quality components are given to the user.

WMS-Q services are found through the "WMS-Q" keyword within the "http://www.geoviqua.org/ def/doc/conventions/vocabulary" vocabulary (see Section 5.3 above). Variables are detected as having components at pixel-level or not, depending on the presence of the keyword "http://qualityml.geoviqua. org/1.0/qualityCollection", and if they are defined are displayable or not displayable (having the <Name> tag or not, see Section 5.5 above).

A prototype to discover datasets with certain statistical properties was implemented, allowing searches that are aware of the dataset quality (for example, standard deviation less than 1 meter). The final port of this quality enhanced implementation to the GEO Portal search engine is to be done.

8. Discussion and Future Work

This paper introduces WMS-Q, a profile of the WMS standard that provides a mechanism for communicating data quality. The essential features of the profile are:

- Communication of data quality at the level of datasets, variables and individual samples
- Re-use of concepts from related standards and vocabularies, including UncertML and QualityML, using the WMS Keyword tag to communicate the quality and uncertainty of a measured variable.
- Full compatibility with version 1.3.0 of the WMS standard.

- Extensions to the SE specification to give greater control over visualizations of uncertain components.

This profile was developed in the first version as an OGC public engineering report [39]. However, during further test in the GeoViQua project, some problems in the document were detected and solved. The results of these modifications were presented here and also in the WMS Standards Working Group that is now considering it as a possible OGC standard profile.

We have presented examples of use of the profile for two raster datasets: one representing continuous data (sea surface temperature) and one representing categorical data (land use classification). Further extensions to the SE specification could be devised in order to support more ways of visualizing uncertainty, such as the "uncertainty ribbons" and "graduated glyphs" described in [7]. Some extensions to vector data was also tested in the GeoViQua project and will be investigated in future work, building on approaches such as those described in [11].

Parallel work has extended these concepts to the Web Map Tile Service (WMTS) standard and will be described in future publications. WMTS was developed reusing many concepts of WMS but with scalability and performance in mind. Both have the concept of Layers and both deliver maps as a result. An important difference that is important for applying this profile is that layer nesting cannot be done in WMTS. Instead, the concept of *themes* was introduced, decoupling the layer definition from the layer interdependencies. A theme is defined as a tree structure that can contain other themes that can represent variables and layer references that can represent uncertainty components. This way, a quality theme can be defined to describe the sample quality components relation for a variable.

This paper focuses in discovery standards but many concepts could be reused in data access standards and services, such as the Web Coverage Service (WCS) standard. The version 2 [40] of this standard has adopted a modular approach that allows for better extensibility. In WCS, GMLCOV [41] (a GML application schema for coverages) is used to describe coverage offerings and it could be profiled to accurately identify uncertain variables and the components that convey them at the sample level. In addition, WCS operations could be extended to request the necessary quality and uncertainty components.

Recently, we have started to collaborate with weather forecasting agencies including MeteoFrance, the UK MetOffice, KNMI, DWD, ECMWF and AFWA, who are working together on a WMS best practice way to better communicate probability of forecasts resulting model ensembles, through WMS. These organizations are already providing WMS services and they have implemented their own solutions. By considering each ensemble result as a statistical sample and derived products as statistical probability distributions, we recognized the relevance of WMS-Q as a solution for ensemble standardization. We expect that the results of these discussions can converge in the use of an improved version of WMS-Q that fulfils the requirements of the weather and ocean forecasting community.

version of the WMS-Q profile was developed in the OGC interoperability experiments series, ninth edition (OWS9). Xavier Pons is recipient of an ICREA Academia Excellence in Research grant (2011–2015).

The authors thank the editors and three anonymous reviewers for constructive comments on earlier versions of this paper.

Author Contributions

Jon D. Blower, Joan Masó, Daniel Díaz and Xiaoyu Yang. designed the WMS-Q profile and wrote the paper. Guy H. Griffiths, Charles J. Roberts and Xiaoyu Yang implemented the profile in the ncWMS software. Daniel Díaz implemented the MiraMon WMS-Q client extensions and Joan Masó and Xavier Pons contributed to the server extensions. Charles J. Roberts, Guy H. Griffiths and Jane P. Lewis designed the extensions to the Symbology Encoding specification.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. De la Beaujardiere, J. *OpenGIS Web Map Service (WMS) Implementation Specification Version 1.3.0*; OGC 06-042; Open Geospatial Consortium: Wayland, MA, USA, 2006.
2. Technical Committee ISO/TC 211, Geographic Information/Geomatics. *ISO 19128:2005 Geographic Information—Web Map Server Interface*; ISO: Geneva, Switzerland, 2005; p. 76.
3. Blower, J.; Gemmell, A.; Griffiths, G.; Haines, K.; Santokhee, A.; Yang, X.A. Web Map Service implementation for the visualization of multidimensional gridded environmental data. *Environ. Model. Softw.* **2013**, *47*, 218–224.
4. Masser, I. *Building European Spatial Data Infrastructures*; Esri Press: Redlands, CA, USA, 2007.
5. Khalsa, S.J.S.; Nativi, S.; Geller, G.N. The GEOSS interoperability process pilot project (IP3). *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 80–91.
6. Pebesma, E.; Cornford, D.; Dubois, G.; Heuvelink, G.B.M.; Hristopulos, D.; Pilz, J.; Stohlker, H.; Morin, G.; Skøien, J.O. INTAMAP: The design and implementation of an interoperable automated interpolation web service. *Comput. Geosci.* **2011**, *37*, 343–352.
7. Sanyal, J.; Zhang, S.; Dyer, J.; Mercer, A.; Amburn, P.; Moorhead, R.J. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Trans. Vis. Comput. Graph.* **2010**, *16*, 1421–1430.
8. Potter, K.; Wilson, A.; Bremer, P.T.; Williams, D.; Doutriaux, C.; Pascucci, V.; Johhson, C. Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated ViSUS-CDAT systems. *J. Phys. Conf. Ser.* **2009**, *180*, doi:10.1088/1742-6596/180/1/012089.
9. MacEachren, A.M.; Robinson, A.; Hopper, S.; Gardner, S.; Murray, R.; Gahegan, M.; Hetzle, E. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartogr. Geogr. Inf. Sci.* **2005**, *32*, 139–160.

10. Müller, M. (Ed.) *OpenGIS Symbology Encoding Implementation Specification Version 1.1.0*; OGC 05-077r4; Open Geospatial Consortium: Wayland, MA, USA, 2006.

11. Keijzer, S.C. Visualizing Scale Related Uncertainty in Web Maps for Spatial Planning. Master's Thesis, Faculty of Geosciences Theses, Utrecht University, Utrecht, The Netherlands, 2011.

12. Technical Committee ISO/TC 176/SC 1, Concepts and Terminology. *ISO 8402:1994 Quality Management and Quality Assurance—Vocabulary*; ISO: Geneva, Switzerland, 1994.

13. Devillers, R.; Gervais, M.; Bédard, Y.; Jeansoulin, R. Spatial data quality: From metadata to quality indicators and contextual end-user manual. In Proceedings of the 2002 OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, Istambul, Turquie, 21–22 March 2002.

14. Technical Committee ISO/TC 211, Geographic Information/Geomatics. *ISO 19157:2013 Geographic Information—Data Quality*; ISO: Geneva, Switzerland, 2013; p. 146.

15. Kresse, W.; Danko, D.M. *Springer Handbook of Geographic Information*; Springer-Verlag: Berlin, Germany, 2012; p. 1120.

16. Veregin, H. Data quality parameters. In *Geographical Information Systems*; Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W., Eds.; John Wiley and Sons: New York, NY, USA, 1999; pp. 177–189.

17. Developing Spatial Data Infrastructures: The SDI Cookbook. Available online: http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf (accessed on 28 September 2015).

18. Lawrence, B.; Lowry, R.; Miller, P.; Snaith, H.; Woolf, A. Information in environmental data grids. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **2009**, *367*, 1003–1014.

19. Merchant, C.J.; Embury, O.; Roberts-Jones, J.; Fiedler, E.K.; Bulgin, C.E.; Corlett, G.K.; Good, S.; McLaren, A.; Rayner, N.A.; Donlon, C. *ESA Sea Surface Temperature Climate Change Initiative (ESA SST CCI): Analysis Long Term Product Version 1.0*; NERC Earth Observation Data Centre: Oxford, UK, 2014.

20. Pons, X.; Sevillano, E.; Moré, G.; Serra, P.; Cornford, D.; Ninyerola, M. Distribución espacial de la incertidumbre en mapas de cubiertas obtenidos mediante teledetección. *Rev. Teledetec.* **2014**, doi:10.4995/raet.2014.3059.

21. Devillers, R.; Bédard, Y.; Jeansoulin, R. Multidimensional management of geospatial data quality information for its dynamic use within GIS. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 205–215.

22. Technical Committee ISO/TC 211, Geographic Information/Geomatics. *ISO 19115:2003 Geographic Information—Metadata*; ISO: Geneva, Switzerland, 2003.

23. Technical Committee ISO/TC 211, Geographic Information/Geomatics. *ISO/TS 19139:2007 Geographic Information—Metadata—XML Schema Implementation*; ISO: Geneva, Switzerland, 2007; p. 111.

24. Technical Committee ISO/TC 211, Geographic Information/Geomatics. *ISO 19113:2002 Geographic Information—Quality Principles*; ISO: Geneva, Switzerland, 2002.

25. Di, L.; Yue, P.; Ramapriyan, H.K.; King, R.L. Geoscience data provenance: An overview. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 5065–5072.

26. Aguilar, R.; Pan, J.; Gries, C.; San Gil, I.; Palanisamy, G.A. Fexible online metadata editing and management system. *Ecol. Inform.* **2010**, *5*, 26–31.

27. Aditya, T.; Kraak, M.J. A search interface for an SDI: Implementation and evaluation of metadata visualization strategies. *Trans. GIS* **2007**, *11*, 413–435.

28. Lush, V.; Bastin, L.; Lumsden, J. Developing a GEO label: Providing the GIS community with quality metadata visualisation tools. In Proceedings of the 21st GIS Research, Liverpool, UK, 3–5 April 2013.

29. Ahonen-Rainio, P. Visualization of Geospatial Metadata for Selecting Geographic Datasets. Ph.D. Thesis, Helsinki University of Technology (TKK), Espoo, Finland, 2005.

30. Sevillano, E.; Ninyerola, M.; Zabala, A.; Bastin, L.; Masó, J. *QualityML: A Dictionary for Quality Metadata Encoding*; EGU: Munich, Germany, 2014.

31. Williams, M.; Cornford, D.; Bastin, L. Describing and communicating uncertainty within the semantic web. In Proceedings of the 7th International Semantic Web Conference Uncertainty Reasoning for the Semantic Web Workshop, Karlsruhe, Germany, 26 October 2008.

32. Bigagli, L.; Nativi, S. (Eds.) *NetCDF Uncertainty Conventions (NetCDF-U) OGC Discussion Paper*; OGC 11-163; Open Geospatial Consortium: Wayland, MA, USA, 2013.

33. Lupp M. *Styled Layer Descriptor Profile of the Web Map Service Implementation Specification*; OGC 05-078r4; Open Geospatial Consortium: Wayland, MA, USA, 2007.

34. IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK, 2013.

35. Teuling, A.J.; Stöckli, R.; Seneviratne, S.I. Bivariate colour maps for visualizing climate data. *Int. J. Climatol.* **2011**, *31*, 1408–1412.

36. Pebesma, E.J.; de Kwaadsteniet, J.W. Mapping Groundwater Quality in the Netherlands. *J. Hydrol.* **1997**, *200*, 364–386.

37. Pons, X. *MiraMon.Sistema d'Informació Geogràfica i Software de Teledetecció Centre de Recerca Ecològica i Aplicacions Forestals*; CREAF: Bellaterra, Spain, 2000.

38. Blower, J.D.; Haines, K.; Santokhee, A.; Liu, C.L. GODIVA2: Interactive visualization of environmental data on the Web. *Philos. Trans. R. Soc. A* **2009**, *367*, doi:10.1098/rsta.2008.0180.

39. Blower, J.; Yang, X.; Masó, J.; Thum, S. *OWS 9 Data Quality and Web Mapping Engineering Report*, OGC 12-160r1; Open Geospatial Consortium: Wayland, MA, USA, 2013.

40. Baumann, P. (Ed.) *Web Coverage Service Interface Standard-Core, Version 2.0.1*; OGC 09-110r4; Open Geospatial Consortium: Wayland, MA, USA, 2012.

41. Baumann, P. (Ed.) *GML Application Schema—Coverages, Version 1.0.1*; OGC 09-146r2; Open Geospatial Consortium: Wayland, MA, USA, 2012.