International Journal of Electrical and Computer Engineering (IJECE)
Vol. 7, No. 6, December 2017, pp. 3164~3175
ISSN: 2088-8708, DOI: 10.11591/ijece.v7i6.pp3164-3175

Towards a Consistent Measurement Stream Processing from Heterogeneous Data Sources

Mario Diván[1], María de los Ángeles Martín[2]
[1]Economic and Law School, National University of La Pampa, Argentina
[1,2]Engineering School, National University of La Pampa, Argentina

Article Info

Article history:

Received Jun 16, 2017
Revised Sep 25, 2017
Accepted Oct 10, 2017

Keyword:

Evaluation
Heterogeneous data sources
Interchange schema
Measurement
Stream processing

ABSTRACT

In this work an updating of the C-INCAMI (Context-Information Need, Concept model, Attribute, Metric and Indicator) conceptual framework for Measurement and Evaluation projects was proposed. The updating incorporated better supporting for the measures stream processing. Therefore, a new version of the measurement interchange schema based on the updated C-INCAMI framework was introduced. This new schema incorporated the concept of "complementary data" linking them with geographic information. The complementary data could be associated with the measures and allowed us incorporating video, geographic information, text plain, audio or pictures with the quantitative measures (deterministic or estimated) jointly. A practical case associated with the Weather Radar of the Experimental Agricultural Station (EAS) INTA Anguil (Province of La Pampa, Argentina) was shown, indicating the advantages of the new schema.

Corresponding Author:

Mario Diván,
Economic and Law School | Engineering School,
National University of La Pampa,
Coronel Gil 353, 1st Floor, Santa Rosa (CP 6300), La Pampa, Argentina.
Email: mjdivan@eco.unlpam.edu.ar

1. INTRODUCTION

Nowadays, there are processing architectures, which allows the real-time data processing through configurable topologies such as Apache Storm and Spark [1], [2]. In this type of architectures, you can dynamically define the processing topology over the data streams. This allows adjusting to different computation necessities, being possible delegating the data structural definition and its meaning inside of the application. In this context, we include to the Processing Architecture based on Measurement Metadata (PAbMM) [3], [4]. PAbMM is supported by the framework for Measuring and Evaluating (M&E) called C-INCAMI [5]. CINCAMI incorporates metadata to the M&E process, promoting repeatability, comparability and consistency.

In PAbMM, the data streams are organized in terms of C-INCAMI/MIS (Measurement Interchange Schema) [6]. CINCAMI/MIS allows homogenize and gather data and metadata jointly inside the same stream from heterogeneous data sources. Through the metadata, we can describe the measurement context additionally to the entity under measurement, which permits avoid analyzing the measure in isolation way.

In our first approximation of the PAbMM to the weather radar (WR) of the EAS INTA Anguil (Province of La Pampa, Argentina) [4], we saw that the WR could provide us not just raw data but also pictures. Therefore, we take into account that C-INCAMI/MIS was incomplete because it just gave support to contextualized measurements under the idea of numeric values but not pictures, geographic information, sound or video. Moreover, the C-INCAMI conceptual framework just supports deterministic values, but not estimated values coming from a likelihood distribution.

However, the proposed changes in the measurement interchange schema could presents a disjunctive situation. It is that to say, even when the incorporation of descriptive information could clarify the measurement context (or the entity under analysis), it is highly possible that we incorporate an overhead in the data stream processing. In this sense, our decision is aligned with a future idea related with a data monetization strategy in which a better description of the measured environment prevails on the performance.

Thus, this work has two main contributions. The first contribution is the extension of the C-INCAMI conceptual framework for incorporating the complementary data, the relationships between them and the linking to the measure concept. This is a key asset because now we can consider the measures not just from the point of view quantitative but also spatial and temporal too. The quantitative point of view is identified trough the deterministic values or the likelihood distribution that allow us monitoring the entity under analysis. The spatial point of view is associated with the geographic information and/or their related data (such as pictures, audios or videos) that describe the contextual situation for the entity under analysis and complementing the quantitative point of view. Finally, the temporal point of view is possible because we have the timestamp associated with the quantitative and spatial point of view. Follows, we can analyze jointly numeric values (deterministic or estimated) and the complementary data (pictures, audio, etc) along the time.The second contribution is the updating of C-INCAMI/MIS for jointly supporting the geographic information, pictures, videos and audios as complementary aspect of the quantitative values. In this way and considering the heterogeneous data sources, this will allow us homogenizing and giving positional feedback to PAbMM for improving the statistical analysis and the classifiers.

This article is organized in six sections. Section 2 outlines the changes in C-INCAMI framework for supporting the pictures, audio and video under the measure concept. Section 3 shows the updated C-INCAMI/MIS schema for supporting these new kinds of data. Section 4 synthesizes the application of C-INCAMI/MIS from the Weather Radar of the EAS INTA Anguil (La Pampa, Argentina). Section 5 discusses related works and finally summarizes the conclusions.

2. C-INCAMI: SUPPORTING PICTURES, AUDIO AND VIDEO

2.1. The Original Measurement and Evaluation Conceptual Framework

C-INCAMI is a conceptual framework [7], [8], which defines the concepts and their related components for the M&E area in software organizations. It provides a domain (ontological) model defining all the terms [5], properties, and relationships needed to design and implement M&E processes. It is an approach in which the requirements specification, M&E, and analysis of results are performed for satisfying a specific information need in a given context. In C-INCAMI, concepts and relationships are meant to be used along all the M&E activities. This way, a common understanding of data and metadata is shared among projects fostering more consistent analysis.

C-INCAMI is structured in six components as follows: a) M&E Project definition, b) Non-functional requirements, c) Context, iv) Measurement, d) Evaluation and vi) Analysis and Recommendation.

The M&E Project definition (Not shown in Figure 1) defines and relates a set of project terms needed to deal with M&E activities, methods, roles and artifacts.

The Non-functional requirements (requirements in Figure 1): it allows specifying the information need of any M&E project. The information need identifies the purpose (e.g. predict, understand, etc.) and the user viewpoint (e.g. final user); in turn, it focuses on a Calculable Concept (e.g. quality system) and specifies the Entity Category to evaluate (e.g. a resource, system, etc). A Calculable Concept can be defined as an abstract relationship between attributes of an entity and a given information need. This can be represented by a Concept Model where the leaves of an instantiated model are Attributes. The attributes can be measured by metrics.

For the context package, one concept is Context, which represents the relevant state of the situation of the entity to be assessed with regard to the information need. We consider Context as a special kind of Entity in which related relevant entities are involved. To describe the context, attributes of the relevant entities are used –which are also Attributes called Context Properties (See [8] for details).

The Measurement component, includes the concepts and relationships intended to specify the measurement design and implementation. Regarding measurement design, a Metric provides a Measurement specification of how to quantify a particular attribute of an entity, using a particular Method (i.e. procedure), and how to represent its values, using a particular Scale. The properties of the measured values in the scale with regard to the allowed mathematical and statistical operations and analysis are given by the scaleType. Two types of metrics are distinguished. Direct Metric is that for which values are obtained directly from measuring the corresponding entity's attribute, by using a Measurement Method. On the other hand, the Indirect Metric value is calculated from other direct metrics' values following a formula specification and a particular Calculation Method.

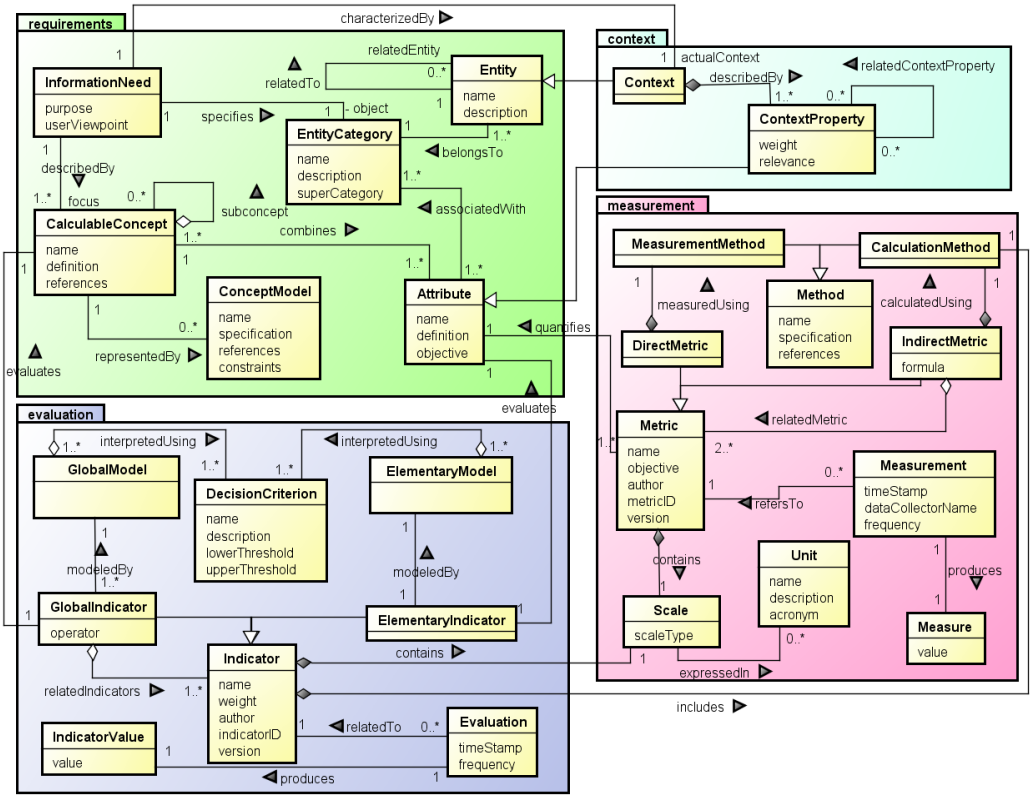

Figure 1. Main components, concepts and relationships for the C-INCAMI conceptual framework [3]

For measurement implementation, a Measurement specifies the task by using a particular metric description in order to produce a Measure value. Other associated metadata is the data collector name and the timestamp in which the measurement was performed.

The Evaluation component includes the concepts and relationships intended to specify the evaluation design and implementation. It is worthy to mention that the selected metrics are useful for a measurement tasks as long as the selected indicators are useful for an evaluation tasks in order to interpret the stated information need. Indicator is the main term, and there are two types of indicators. First, Elementary Indicator that evaluates attributes combined in a concept model. Each elementary indicator has an Elementary Model that provides a mapping function from the metric's measures (the domain) to the indicator's scale (the range). The new scale is interpreted using agreed decision criteria, which help analyze the level of satisfaction reached by each elementary nonfunctional requirement, i.e. by each attribute. Second, Partial/Global Indicator, which evaluates mid-level and higher-level requirements, i.e. sub-characteristics and characteristics in a concept model. Different aggregation models (GlobalModel) can be used to perform evaluations. The global indicator's value ultimately represents the global degree of satisfaction in meeting the stated information need for a given purpose and user viewpoint. As for the implementation, an Evaluation represents the task involving a single calculation, following a particular indicator specification –either elementary or global-, producing an Indicator Value.

The PAbMM reuses the C-INCAMI conceptual base, in order to obtain a repeatable and consistent data stream processing. Thus, streams are basically measures (i.e., raw data usually coming from sensors), which are linked accordingly with the metadata based on C-INCAMI, such as the entity being measured, the attribute and its corresponding metric with the scale and measurement/calculation procedure, the trace group, among others. For a given data stream, not only measures associated to metrics of attributes are tagged but also measures associated to contextual properties as well. Thanks to each M&E project specification is based on C-INCAMI, the processing of tagged data streams are then in alignment with the project objective and information need, allowing thus traceability and consistency by supporting a clear separation of concerns. For instance, for a given project –more than one project can be running at the same time-, it is easy to identify whether a measure is coming from an attribute or from a contextual property, and also its associated scale type and unit. Therefore, the statistical analysis is benefited because the verification for consistency of each measure against its formal (metric) definition can be performed.

2.2. The Role of C-INCAMI in PAbMM

In the nutshell, the measurement stream is informed by each heterogeneous data source to the Measurement Adapter (MA). The MA incorporates the metadata (e.g. metric ID, context property ID, etc.) associated to each data source into the stream in order to transmit measures to the Gathering Function (GF). Such measures are organized in GF by their metadata and then sent to the Analysis & Smoothing Function (ASF). ASF performs a set of statistical analysis on the stream in order to detect deviations or problems with data, considering its formal definition (as per C-INCAMI DB). In turn, the incremental classifiers (i.e. the Current and Updated Classifiers) analyze the arriving measures and act accordingly triggering alarms in case a risk situation arises.

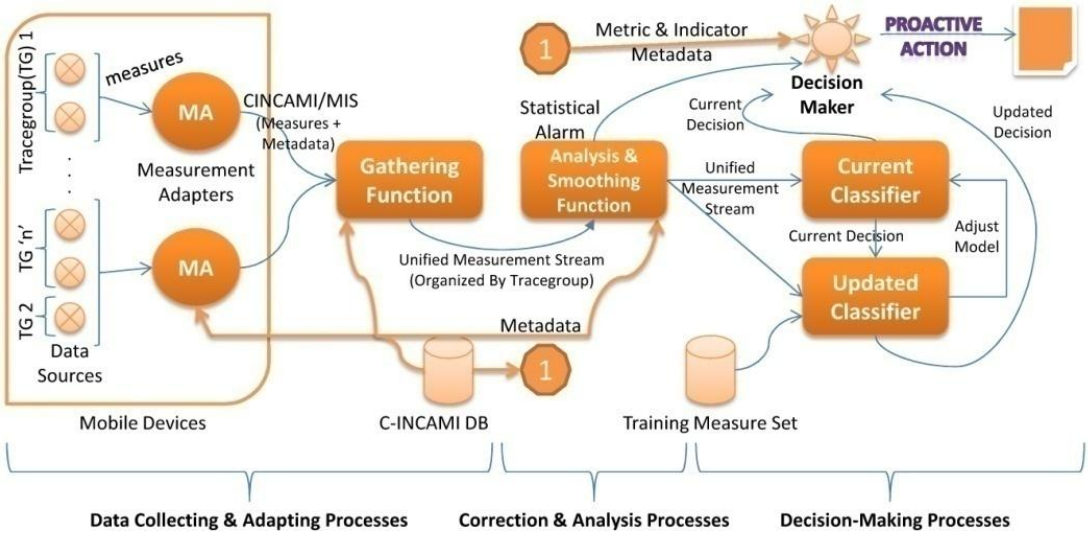

Figure 2. The General Processing Perspective of PAbMM

As you can see in Figure 2, the data sources can be gruped under the idea of the tracing group. In this sense, nothing warranty the homogeneity between different data sources belonging to the same tracing group (e.g. two WR from different manufacturers, or even, from the same manufacturer but with different models). In [4] we saw that the WR could provide us not only with deterministic values but also with pictures or even pictures sequences. In this sense, the original proposal from C-INCAMI (See section 2.1.) in relation with the measurement component was not enough because it considers the value associated with the measure just like a deterministic number (See in the Measure class inside the Measurement package in Figure 1, just is able to represent one value).

2.3. Extending C-INCAMI for Supporting Complementary Data

In Figure 3, we propose the extension of the concept called measure and the use of complementary data as a new point of view over each measure. For better differentiation, we maintain with painted background the original concepts from C-INCAMI and with blank background the new incorporated concepts or extensions.

As shown in Figure 3, the QuantitativeMeasure class inherits from the Measure class and it represents the situation in which the measure is exclusively numeric, but not necessarily deterministic. In this sense, the QuantitativeMeasure class could be associated with a deterministic measurement process in which we get a clearly defined value (DeterministicMeasure class in Figure 3), or it could be associated with an estimated process in which we obtain a set of <value, likelihood> pairs (See the classes called LikelihoodDistribution and EstimatedValue in Figure 3). For this reason, the QuantitativeMeasure class incorporates the idea of synthesisAlgorithm as attribute, because even when is trivial in a deterministic value (the value is the same); it does not trivial in likelihood distribution in which we could use the mathematical expectation for synthesizing the distribution in one representative value.

Each measure could have complementary data (See the classes ComplementaryData) which allow us complement the measure itself, for example, one picture could describe the habitat under analysis when we want to measure the wind velocity of the place. So, the complementary data are a composition of at least one complementary datum, which is represented as an abstract class called ComplementaryDatum in the Figure 3

and it incorporates three attributes: 1) mimeVersion: the mime version associated with the complementary datum, 2) mimeContentType: it identifies the type of content (e.g. image/jpeg, audio/mpeg, video/3gpp, etc.), and 3) hashControl: it is a footprint for verifying the integrity of the content (e.g. MD5 footprint). As you can see in Figure 3, from the ComplementaryDatum class inherit five classes:

1. GeographicComplementaryData: it allows sending a document under the Geography Markup Language (GML) [9], [10] as a complement of the measure. This class has two attributes, GMLDocument and GMLVersion. The GMLDocument attribute contains the document organized in terms of GML and the GMLVersion attribute refers to the GML's version. It is worthy to mention because it allow us incorporates the possibility for establishing a relationship between the positioning and the image, the audio track or the video;
2. PlainTextData: It allows incorporating textual information associated with the measure or the measurement device, for example, the device's log at the moment of the measure. This class incorporates the attributes textValue and language. The textValue attribute contains the data in text plain and the language attribute indicates the idiom in which the text is written by the use of ISO 639 (11).
3. PictureData: This class incorporates the possibility of using a photo as complement of a measure. The timestamp attribute indicates the moment in which the picture was taken, and the pictureValue attribute store the data associated with the image itself. As you can see in Figure 3, one PictureData's object could be associated with positional data through the association with the class GeographicComplementary.
4. AudioTrackData: the class allows us use an audio track as a complement of the measure. The timestamp attribute indicates the moment in which the audio track started the recording. The duration attribute is associated with the track's longitude and the audioValue attribute represents the audio itself. In this case, we can establish a relationship between an AudioTrackData's object and a GeographicComplementaryData's object considering the timestamp. That is to say, if we can synchronize the timestamp associated with the starting of the audio track with a described instant in the GML data, then it's highly possible analyze the audio and geographic information jointly.
5. VideoData: it represents the possibility of using a video as a complement of the measure. The timestamp attribute indicates the moment in which the video started the recording. The duration attribute is associated with the video's longitude and the videoValue attribute represents the video itself. Like the audio track, we can establish a relationship between a VideoData's object and a GeographicComplementaryData's object considering the timestamp.

In this sense and trough the incorporation of the classes associated with the complementary data, we now be able to manage a new perspective of data for complementing each measure and its processing. The C-INCAMI framework is extended for supporting complementary data in metrics associated with entity attributes and/or context properties. For taking advantage of the complementary data in the measurement processes, we need incorporate in C-INCAMI the specific Figure of the data source because each measurement project has different requirements and the data collector is essential in terms of reliability, precision, etc.

With the aim of focusing the concept associated with the data collector, we replace the dataCollectorName attribute from the Measurement class (Inside of Measurement package in Figure 1) for a new DataSource class (See in Figure 3). The DataSource class represents the concept associated with the measurement device, which allows us to get the measures. In this sense, the DataSource class incorporates four representative attributes: dataSourceID, name, type and dataFormats. The dataSourceID attribute identifies the data source along the M&E projects for fostering the traceability. The name attribute is a friendlier way for referencing the device (e.g. an alias). The type attribute allows us to know if the device sends the measures in predictable or unpredictable way (e.g. the weather radar regularly sends data for processing. So, it is predictable). The dataFormats attribute allow us knowing about the different ways that the data source could organizes the content. Each data source sends the measures always through a measurement adapter (See DataSourceAdapter in Figure 3). The measurement adapter translates from the original data format associated with the data source to the C-INCAMI/MIS stream. In this way, each data source adapter incorporates three attributes: dsAdapterID, name and supportedFormats. The dsAdaptedID attribute identifies the measurement adapter along the M&E projects. The name attribute is a friendlier way for referencing the measurement adapter. The supportedFormats attribute allow us knowing about the kind of formats that the adapter could translates to C-INCAMI/MIS.

When we have different data sources monitoring the same entity under analysis (e.g. different weather radars making complementary monitoring for the same region), we could grouping them under the TraceGroup class (See in Figure 3). It is important because even when this concept there not existed before in C-INCAMI, its incorporation is useful in the processing as you can see in Figure 2.

The DataSourceProperty class (See in Figure 3) inherits from the Attribute class (requirements package in C-INCAMI, See Figure 1) and represents each property that allows us characterizing a data source (or measurement device). In this way, the DataSourceProperties contains the characteristics that describe to the data source, indicating the relevance and the associated value. Additionally, we incorporate a new association between Metric and DataSource called candidates. It represents what data sources are able for getting measures in terms of the metric definition. Moreover, through the association called performedBy between the Measurement class and Data Source class, we determine the origin of the data and we keep the traceability.

Finally, we extend the metric definition incorporating the concept of device's constraints through the Constraint class (See in Figure 3). The constraints allow defining the minimum requirements that the measurement devices must satisfies before implementing the metric (e.g. minimum accuracy). A constraint is associated with one data source property but a metric may has a set of constraints linked. Each constraint incorporates the procedure for filtering (filterAlgorithm attribute in Figure 3) and the kind of filter (filterType in Figure 3. e.g. mandatory or preferable). We could limit the valid values associated with a data source property by pattern (PatternConstraint Class in Figure 3) or explicitly indicating a set of valid values (ConstraintValue class in Figure 3). Therefore, considering the constraints associated with the metrics and the available data sources, we could know the candidates data sources and the data sources linked with the measurement.

This extension of the C-INCAMI framework is meaningful in relation to the original version instroduced in section 2.1. Now, we can consider the measures not just from the point of view quantitative but also spatial and temporal too (e.g., for the WR of the EAS INTA Anguil, we can keep the geographic data/plain text/picture/video/audio and the quantitative measures jointly for each sampling point). Moreover, we can define constraints that allow us to be more selective in terms of the available data sources for implementing different metrics. However, a concern to consider is the overhead in the data stream processing of these new incorporations. In this sense, we must highlight that the complementary data are optionals and not mandatories. It is that to say, they just will be used on demand (e.g. in M&E projects oriented to data monetization).

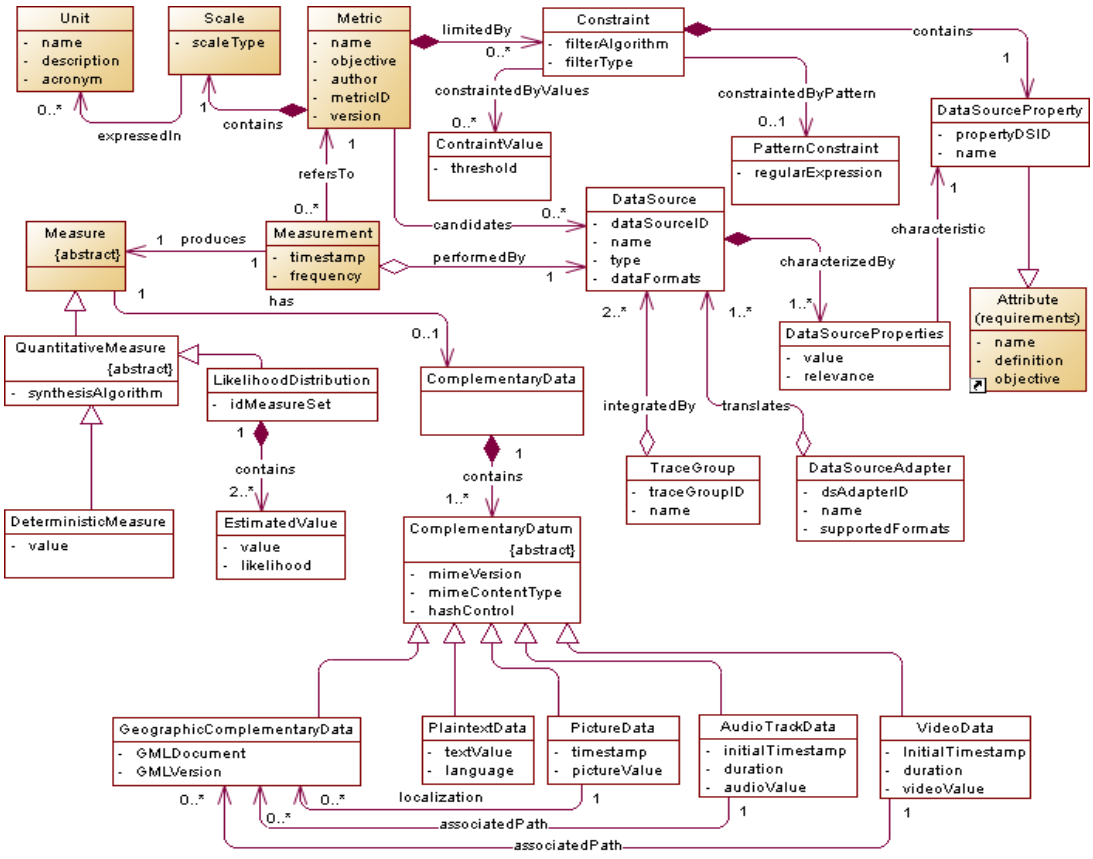

Figure 3. Extensions for the Measurement Component of C-INCAMI

3. THE UPDATED MEASUREMENT INTERCHANGE SCHEMA

The first version of C-INCAMI/MIS [6] is based on the original version of C-INCAMI [7], [8]. From the introduced changes in the section 2.3., we managed to update the measurement interchange schema for supporting the quantitative measures (deterministic or estimated) and complementary data jointly.

Figure 4 shows the new organization of C-INCAMI/MIS for the version 2. The tag CINCAMI_MIS incorporates two attributes: version and dsAdapterID. The version attribute refers to the edition of the schema and the dsAdapterID refers to the data source adapter used for translating the measures from the raw data (e.g. from the radar) to the CINCAMI/MIS stream. Under the measurementItemSet tag, we could have many measurementItem tags. Each measurementItem tag incorporates three attributes: 1) dataSourceID: it represents the identification code for the data source along the M&E projects, 2) originalDataFormat: the data format coming from the sensor (e.g. the weather radar) before translating through the measurement adapter, and 3) footprint: it allows verify the integrity of the content. In this way, using the tags dataSourceID and dsAdapterID in the CINCAMI/MIS stream, we know the origin of the data (e.g. the INTA Anguil Weather Radar) and the responsible for its translation (e.g. a mobile device) before the processing.

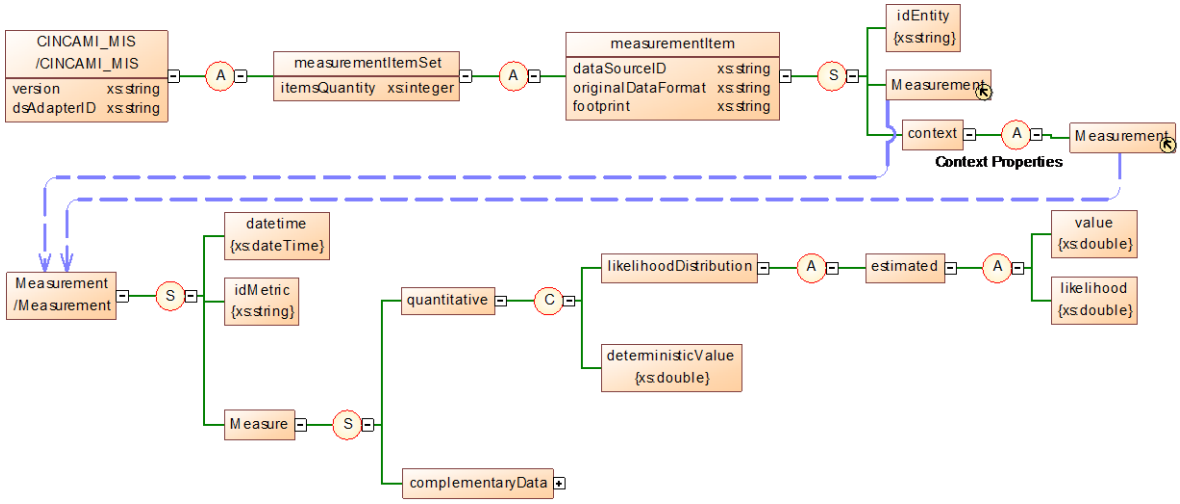

Figure 4. CINCAMI / Measurement Interchange Schema - Version 2 (A=All, S=Sequence and C=Choice)

Under the measurementItem tag, we have the tags idEntity, Measurement and context. The idEntity tag identifies the entity under analysis. The Measurement tag describes the measurement associated with the attribute of the entity under analysis. Under the context tag, we grouped the context properties associated with the context of the entity under analysis as a set of Measurement tags. Therefore, we have two different Measurement tags but with the same internal structural organization.

The Measurement tag organizes the new concepts in terms of the updated C-INCAMI framework (See the section 2.3.). Here we have three tags: 1) datetime: it exposes the moment in that the device gets the measure, 2) idMetric: it indicates the metric associated with the attribute of the entity under analysis or the context property related to its context, and 3) Measure: it organizes the data discriminating between the quantitative measure and the complementary data.

The quantitative measure can have an estimated or deterministic value. On the one hand and if the value is estimated, we will have a likelihood distribution expressed as set of <value, likelihood> pairs; and on the other hand and if the value is deterministic, we will have just one numerical value.

Additionally, each measure could have complementary data under the complementaryData tag. For better understanding associated with the Figure, the complementary data are synthesized by its tag in Figure 4 and they are detailed in Figure 5. In Figure 5, under the complementaryData tag, we can have one or more complementaryDatum tags which represent each particular complement. The complementaryDatum tag has associated three attributes: mimeVersion, mimeContentType and duration. The attributes mimeVersion and mimeContentType have the same meaning such as we defined in section 2.3. (when we introduced the ComplementaryDatum class). The attribute duration indicates the time in seconds associated with audio or video, but in the case of picture, it always will be cero.

Under the complementaryDatum tag we have five possibilities for choosing the kind of complementary datum; it could be geographic information (GML tag), a photo (pictureData tag), a plain text

(plainText tag), an audio track (audioTrackData tag) or a video (videoData tag). Just one possibility can be chosen for the complementaryDatum tag. Eventually, if we need send more than one complementary datum, then we will use as many complementaryDatum tags as we need under the complementaryData tag. As shown in Figure 5, on the one hand, the GML tag is a possibility to be chosen under the complementaryDatum tag, and on the ohter hand it can be reused from the geographyData tag associated with the tags pictureData, audioTrackData and videoData. It is important because allow us establish a relationship between the audio, video or picture and the geographic information.

This new version of C-INCAMI/MIS extends the previous version giving supports to the extended C-INCAMI framework shown in the Section 2. Therefore, the new version of the measurement interchange schema incorporates new possibilities not considered before, such as:

a. Traceability: we incorporate the responsible (the measurement adapter) for translating from the raw data generated by the data source, to the C-INCAMI/MIS stream,
b. Complementary Data: we include new data types associated with the measures (e.g. audio, video or pictures). Moreover, the complementary data could be related with geographic information for bettering the localization in maps and the traceability. It is worthy to mention that this complementary data are optionals. It is that to say, when we need perfomance in the data stream processing we coul omit the complementary data and make particullary focus on the quantitative measures.
c. Integrity Verification: Under each measurementItem tag, the integrity of the content can be verified using the footprint, and
d. Supporting for new kind of measures: The context and the entity under analysis can be monitored jointly using quantitative measures and now, these measures can be estimated or deterministic. The new schema is able for managing likelihood distributions allowing the use of synthesis algorithms for getting a representative value of the data series.

Therefore and as we showed before, the version 2 of CINCAMI/MIS incorporates important changes which allows extending the possibilities of monitoring for the entities under analysis and their contexts. It is important because opens new opportunities oriented to different data monetization strategies.

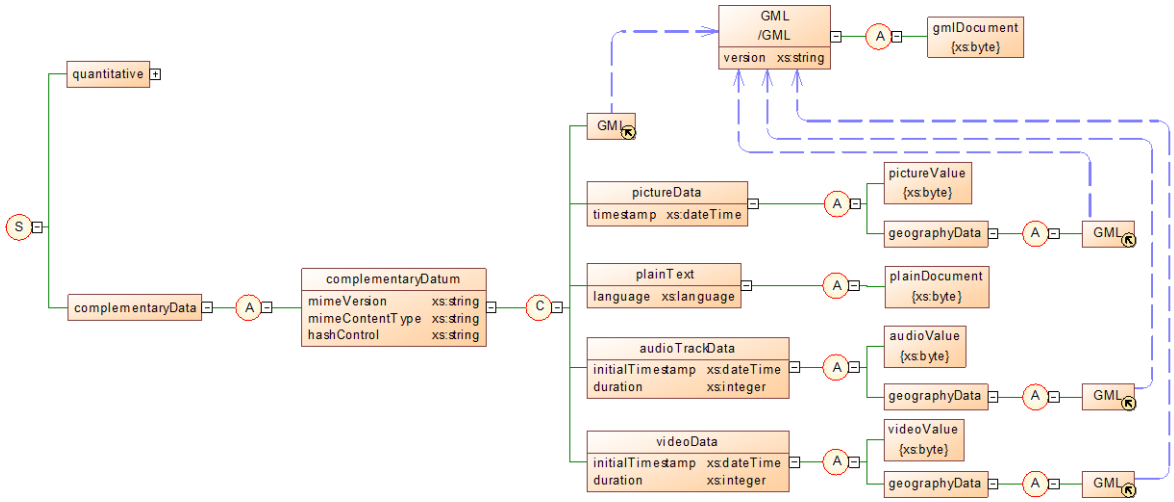

Figure 5. The Detailed Complementary Data from Figure 2 for the CINCAMI / Measurement Interchange Schema - Version 2 (A=All, S=Sequence and C=Choice)

4. A PRACTICAL CASE: THE WEATHER RADAR

As shown in Figure 6, WR are active sensors of remote sensing that emit pulses of electromagnetic energy into the atmosphere in the range of microwave frequencies. Their measurements are based, first, on the electromagnetic radiation as it propagates in the atmosphere is scattered by the objects and particles existing, and secondly, the ability of the antennas for emitting directed radiation and capturing the radiation incident from a certain direction. These sensors are tools to monitor environmental variables, and specifically, the identification, analysis, monitoring, forecasting and evaluation of hydro meteorological phenomena, as well as physical processes that these involve, given the risk that can cause severe events. The scanning region associated with the WR is dynamic and it changes its radius from 120km to 360km,

considering the radar as the centre point of the circle. Additionally, the scanning happens in different angles of elevation, from 0 to 85° along the scanning territory given by the scanning radius. It is important to say that the 5° immediately upper the radar, it cannot be monitor and for that reason is called silence cone (See Figure 6).

The main applications of the WR are: a) weather description, forecasting and nowcasting, b) forecasting and monitoring of environmental contingencies (e.g. hail, torrential rain, severe storms, etc.), c) Security in navigation and air traffic control, d) Studies of atmospheric physics, e) studies of agro climatic risk, f) Provision of basic data for scientific and technological research, and g) Provision of input data for hydrological models (e.g. floods) [4]. The information recorded by the WR is collected through volumetric scans and each sampled cell has a 1 km^3 as you can see in Figure 6. The sample units are defined as 1 km^2 and 1º. The data contains the different variables: reflectivity factor (Z), differential reflectivity (ZDR), polarimetric correlation coefficient (RhoHV), differential phase (PhiDP), specific differential phase (KDP), radial velocity (V) and spectrum width (W) (12).

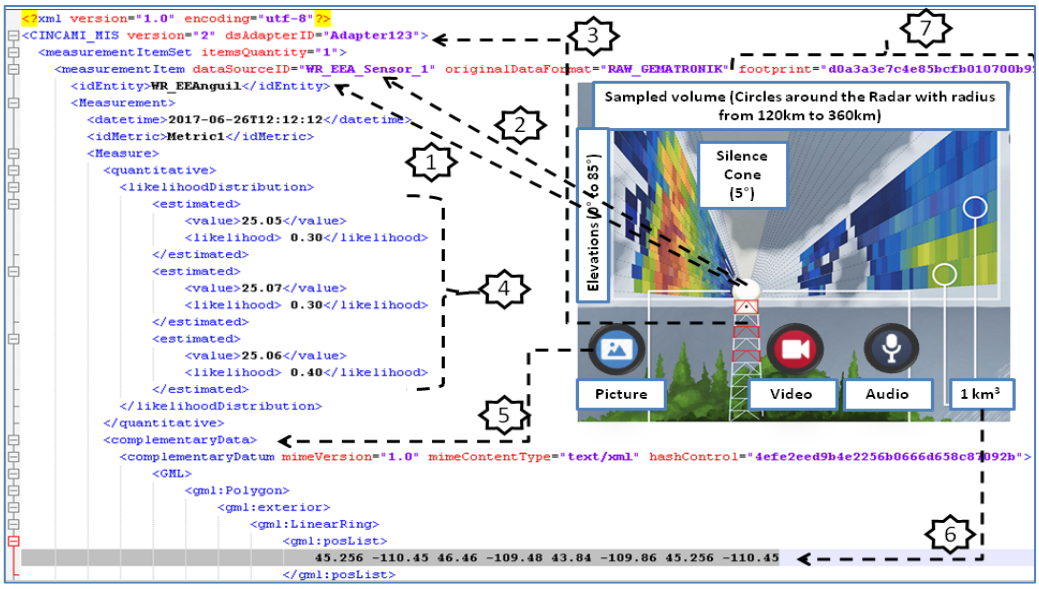

Figure 6. Partial View of a CINCAMI/MIS version 2 Document

Figure 6 show a partial view of a XML document and its relationship with the weather radar. This document is organized in terms of the CINCAMI/MIS version 2. The star "1" identifies the entity under analysis in the document (e.g. the weather radar of EEA, Anguil –Argentina-). However, the weather radar could have many sensors associated, and for that reason the star "2" identifies the data source associated with the data origin. Because the data sources could generate the data under propietary format (e.g. RAW_GEMATRONIK), a data source adapter is necessary for keeping the interoperability between the heterogeneous data sources and PAbMM. For example, the star "3" indicates that the measurement adapter is identified as Adapter123 (e.g. it could be located under the sphere of the WR in the example Figure).

The star "4" in Figure 6 shows a quantitative measure expressed as likelihood distribution, it is that to say, there are three values with their associated likelihood which belong to the same measurement. The star "5" refers to the tag under which many complementary data could be expressed. In fact, the star "6" tries to express the geographic region associated with the sampled volume by the WR and related to the quantitative measure.

The star "7" shows the footprint, through which allows us checking the data integrity in relation to the transmitted document. In terms of data stream, each footprint is associated with a logical window for processing in PAbMM. Thus, with the extension of C-INCAMI framework and the updating of CINCAMI/MIS (See Section 2.3. and 3 respectively), we incorporate important updating in the measurement interchange schema, such as:

a. We are able to send environmental audio coming from the WR context inside the data stream, e.g. for audio monitoring using audio classifiers (e.g. useful for wind, electrical activity, etc.),

b. We are able to use descriptive picture associated with the region in which the sensor is located, e.g. for graphical monitoring of the phenomena,
c. We are able to interchange videos describing the WR region, e.g. for storm monitoring,
d. We can verify the integrity of the multimedia data using the footprint. It is important because the complementary data (e.g. picture, audio & video) are associated with the quantitative measures, and they give a new perspective for the same phenomenon,
e. The schema incorporates an additional tool for data traceability through the identification of the measurement adapter. Therefore, we identify the responsible for generating each C-INCAMI/MIS stream from the raw data coming from the WR,
f. We can associate the complementary data (i.e. video, audio, picture and text plain from logs) with the measures and their geographic information using GML. It is important because now we improve the notion of measure positioning and its associated traceability.

As shown in Figure 6, the quantitative measures (estimated or deterministic) and the complementary data (audio, video, text plain, picture & geographic information) could be associated. They allow us getting a better description of each situation under analysis in terms of the entity under monitoring and its context. In this sense, with better descriptions of each situation, it is highly likelihood improving the posterior analysis associated with the online processing strategy

5. RELATED WORKS

There are recent works which make focus on the data stream processing from a syntactic point of view. In this context, the data model of the stream is based in a key-value structure and incorporates techniques for the adaptive management of high-rate streams [13]. Our previous proposal with C-INCAMI/MIS version 1 incorporated data and metadata jointly inside the same data stream. It allowed us interchange different kind of data and we managed to interpret them in base on the M&E Project Definition. Now, with the version 2 of the measurement interchange schema, we incorporate integrity control and new data types such as video, audio or picture inside the data stream, which allow us getting a better description about the entity under analysis and its context.

Apache Storm incorporates the data such as tuples in analogy with the relational model and interchanges the tuples under a fixed structure between Spouts and/or Bolts [1], [14]. In C-INCAMI/MIS, the characteristics associated with the entity under analysis (attributes) or its context (context properties), are defined in the M&E project and they are not incorporated such as part of the topology as happen in Apache Storm. It is important, because the measures (data and metadata) and the phenomena under analysis guide the processing strategy in base on the monitoring necessities incorporated in the M&E Project. In this sense, an interesting and usable aspect is associated with Kafka Stream [15], [16] because it includes a framework-free stream processing, so we can process the CINCAMI/MIS stream in a transparent way.

The Resource Description Framework (RDF) is a standard model for data interchange from the semantic web [17], [18]. RDF has features that facilitate data merging and it supports schemas and semantic verification. CINCAMI/MIS is a specialized measurement interchange schema based on the C-INCAMI conceptual framework. Therefore, we just need define the information need, the entity under analysis to monitor and its context inside the M&E project using the concepts, terms and relationships of C-INCAMI. Followed, we use the data stream (data and metadata) coming from different sensors for guiding the processing strategy [19] in a more consistent way.

The ISO 19156 [20] defines a conceptual schema for observations, and for features involved sampling when making observations. These provide models for the exchange of information describing observation acts and their results. The standard exposes many interesting points such as the incorporation of the dimensions temporal, spatial or spatial-temporal. In this sense, the new C-INCAMI/MIS schema reuse the idea but extending it for supporting estimated measures. So, using C-INCAMI/MIS we can inform deterministic and estimated measures and they could be related with geographic data for bettering the associated positioning using GML.

6. CONCLUSION

In this work, we showed an extension for the C-INCAMI conceptual framework that allows us incorporating a new perspective for the concept of measures. Now, the measures could be complemented trough video, geographic information, text plain (such as radar logs), pictures or audio as support of the quantitative measures. Even, the quantitative measures could be deterministic or estimated, opening the possibility for dealing with likelihood distribution in transparent way. It is important, because the quantitative

measures and the complementary data work together for better describing of the entity under analysis and its context.

We introduce the data source adapter concept for improving the traceability between the processing strategy and the data source. Now we know not just the origin of the data (data source), but also the responsible for the translating between the raw data and the C-INCAMI/MIS stream.

The idea of tracing group incorporated in C-INCAMI, allows us joint monitoring between different data sources. Therefore, when different data sources are monitoring the same entity under analysis and/or context, we could introduce the tracing group in the M&E project definition, like logical grouping of the data sources for better joint data analysis.

In this way and with the extension of C-INCAMI, using the quantitative measures and their complementary data, the integrity control (e.g. using the footprint), the tracing group for joint analysis and the data traceability, it is possible to build new views such as the temporal, spatial and/or spatial-temporal.

We showed the updating of C-INCAMI/MIS and its relationship with the new concepts incorporated in the extension of C-INCAMI. This is a key asset because allow us dealing with heterogeneous data sources, homogenizing and giving positional feedback to PAbMM for improving the statistical analysis and the classifiers.

From the point of view associated with the processing method, we synthesize the general strategy in the section 2.2. This allowed introducing the relationship between the CINCAMI/MIS and the data stream processing. Because the processing strategy is guided by metadata (e.g. the data origin, the entity under analysis, etc), the gathering function, the statistical analysis and the classifiers are benefited given than better information on the stream implies better contextualization at the moment of the processing. Even, the possibility of managing likelihood distributions make more consistent the statistical analysis, because before the measures always was determinic. The processing method was not able to discriminate one situation from other. This is key in the statistical analysis, because a simple correlation calculation could be affected for this detail and, for example, it would be a concern for detecting miscalibration in a weather radar. Moreover, the C-INCAMI DB (See Section 2.2.) keep the training data for the start-up of the online classifiers and use the previous experiences organized in cases in a organizational memory with case-based reasoning (CBR). In this organizational memory, each item of knowledge is based in CINCAMI/MIS. So, the new schema allows getting a wider charcterization very useful in a case-based reasoning. This is important because the CBR is oriented to online recommend courses of actions (e.g. when the miscalibration of the weather radar happens).

A practical case about the application of C-INCAMI/MIS in the Weather Radar (as data source) of the EAS INTA Anguil (La Pampa, Argentina) was synthesized. In the practical case, we exposed how the complementary data (picture, audio, geographic information, text plain and video) could contribute for better describing the meteorological phenomena. Even, we incorporate new information in the measurement interchange schema (e.g. the footprint, the measurement adapter, etc.), which provides us a better way for implementing the integrity and data traceability control.

As future works, we will implement the proposed changes in C-INCAMI and the version 2 of C-INCAMI/MIS for incorporating them in our processing strategy (i.e. PAbMM). Additionally, we will carry on a performance benchmark between both versions of the measurement interchange schema.

ACKNOWLEDGEMENTS

This research is supported by the PICTO 2011-0277 project of the National Agency of Technology and Science (Argentina) and the project 09/F068 of the Engineering School. This research integrates the technical cooperation agreement between Engineering School and the EAS INTA Anguil.

REFERENCES

[1] A. Jain and A. Nalya, "Learning Storm. Create real-time stream processing applications with Apache Storm," Packt Publishing Ltd., Birmingham, England, 2014.
[2] M. Frampton, Mastering Apache Spark, Packt Publishing Ltd., Birmingham, England, 2015.
[3] M. Diván and L. Olsina, "Process View for a Data Stream Processing Strategy based on Measurement Metadata," *Electronic Journal of Informatics and Operations Research*, vol. 13, nº 1, pp. 16-34, June 2014.
[4] M. Diván, et. al., *"Towards a Data Processing Architecture for the Weather Radar of the INTA Anguil,"* Proceedings of International Workshop on Data Mining with Industrial Applications, IEEE, Edited by S. Gómez, E. Hochsztain & D. Romero. Pp. 72-78. Asunción, Paraguay, 2015.
[5] L. Olsina and M. Martín, Ontology for Software Metrics and Indicators. Journal of Web Engineering (JWE), vol. 3, nº 4, pp. 262-281, 2004.
[6] M. Diván, "Strategy for Data Stream Processing based on Measurement Metadata (In Spanish)," PhD Thesis. Computer Science School. National University of La Plata, Buenos Aires, Argentina, 2011.

[7] L. Olsina, et. al., "How to Measure and Evaluate Web Applications in a Consistent Way", Chapter 13 in Web Engineering, pp. 385–420, Edited by G. Rossi, O. Pastor, Daniel Schwabe & L. Olsina. Springer, London, England, 2008.
[8] H. Molina and L. Olsina, *"Towards the Support of Contextual Information to a Measurement and Evaluation Framework,"* Proceedings of 6th International Conference on the Quality of Information and Communications Technology (QUATIC). IEEE, edited by R. Machado, F. Brito e Abreu & P. da Cunha, pp. 154-166. Lisbon, Portugal, 2007.
[9] Open Geospatial Consortium and International Standard Organization (ISO), ISO 19136:2007. Geographic Information - Geography Markup Language. ISO, 2007.
[10] Open Geospatial Consortium and International Standard Organization (ISO), ISO 19136-2:2015. Geography Markup Language (GML) -- Part 2: Extended schemas and encoding rules. ISO, 2015.
[11] International Standard Organization (ISO), ISO 639-2:1998. Codes for the representation of names of languages -- Part 2: Alpha-3 code. ISO, 1998.
[12] Gematronik, Rainbow® 5 Products & Algorithms, Gematronik GmbH. Neuss, Germany, 2005.
[13] M. Lee, et.al., "Load Adaptive and Fault Tolerant Distributed Stream Processing System for Explosive Stream Data," *ICACT Transactions on Advanced Communications Technology*, vol.5 n° 1, pp. 745-751, January 2016.
[14] J Samosir, et.al., "An evaluation of data stream processing systems for data driven applications," *Procedia Computer Science*, vol. 80, pp. 439-449, June 2016.
[15] R Estrada and I Ruiz, The Broker: Apache Kafka, Chapter 8 in Big Data SMACK: A Guide to Apache Spark, Mesos, Akka, Cassandra, and Kafka. pp. 165-203. Apress, Berkeley, CA, USA. 2016.
[16] M Noll, "Introducing Kafka Streams, the new stream processing library of Apache Kafka," Session in Berlin Buzzwords Conference. Video available at http://2016.berlinbuzzwords.de/session/introducing-kafka-streams-new-stream-processing-library-apache-kafka. Berlin, Germany, June 2016.
[17] P. Szeredi, et.al. "The Semantic Web explained: the technology and mathematics behind Web 3.0," Cambridge University Press, New York, USA, 2014.
[18] D Brickley, et.al.(Eds.). RDF Schema 1.1. W3C recommendation, vol. 25, pp. 2004--2014, January 2014.
[19] M. Diván, *"Processing Architecture based on Measurement Metadata,"* Keynotes proceeding of 5th International Conference on Reliability, Infocom Technologies and Optimization. IEEE, edited by B. Shukla, S. Khatri & P. Kapur, pp.9-18. Noida, India, 2016.
[20] Open Geospatial Consortium and International Standard Organization (ISO), ISO 19156:2011. Geographic Information - Observations and Measurements. ISO, 2013.

BIOGRAPHIES OF AUTHORS

Mario José Diván is full professor in the Law and Economic School at National University of La Pampa (Argentina). He is director of the Data Science R&D Group at the Engineering School. He has obtained a PhD in Computer Sciences from La Plata University (Argentina) and a MBA from the National Technological University (Córdoba, Argentina). Additionally, he managed to three specialities: a) High Performance and Grid Computing from La Plata University (Argentina), b) Data Mining and Knwoledge Discovery from the Buenos Aires University (Argentina), and c) Managerial Engineering from the National Technological Unviersity (Córdoba, Argentina). He is academic committee member in the Computer Science PhD and Master in the Technological National University (Córdoba Argentina). ACM Professional Member. His research interests include Data Stream Processing, Stream Mining, Data Mining, Big Data, Data Quality, Measurement and Evaluation.

Maria de los Angeles Martin is an Associate Professor in the Engineering School at National University of La Pampa, Argentina, and researcher of software and Web engineering R&D group. Her research interests include Web engineering, particularly, Web knowledge management, Organizational Memory, Case Based Reasoning, Semantic Web and Ontologies. She is a doctor in computer sciences, and Magister in the software engineering area. In the last 10 years, she has published over 30 refereed papers.