Strategy for the Data Monetization in Tune with the Data Stream Processing

Mario José Diván
Economics and Law School, National University of La Pampa, Santa Rosa (CP 6300), Argentina
Engineering School, National University of La Pampa, General Pico (CP 6360), Argentina
mjdivan@[eco|ing].unlpam.edu.ar

Abstract — In this invited talk, a strategy for the data monetization in tune with the data stream processing is shown. The current organizations have as challenge not only the processing of the big data repositories, but also the data monitoring in real-time for guiding their commercial strategies. When an organization is closer of their customers, it has better knowledge of them and better results will obtain related to their action lines. In this sense, the real time data monitoring is a key asset in the now economy and to know how capitalize them could represent a competitive advantage for any company. In this proposal, the Measurement and Evaluation (M&E) projects oriented to data monetization are organized in terms of the C-INCAMI (Context – Information Need, Concept Model, Attribute, Metric and Indicator) framework for fostering the repeatability, comparability, consistency and extensibility of the measurement process. Therefore, a brief introduction to the M&E framework is done and the Processing Architecture based on Measurement Metadata (PAbMM) is introduced. The PAmbMM allows incorporating a detective and predictive behavior based on the data and metadata coming in real time. The interchanging of data between the data sources and PAmbMM is supported by the CINCAMI/MIS (Measurement Interchange Schema) which allows keeps the data traceability and collecting a set of complementary data (e.g. geographic data) associated with the entity under analysis and its context. In this sense, the CINCAMI/MIS 2 is summarized and the associated data monetization strategy is detailed. Finally, an application case for exposing the potential associated with the monetization strategy related to PAmbMM is shown.

Keywords — Data Monetization, Data Stream Processing, Big Data, Measurement & Evaluation

I. INTRODUCTION

The real time data monitoring is a key asset in the now economy and to know how capitalize them could represent a competitive advantage for any company. Every time, the mobile devices (e.g. cell phone) incorporate sensors that allow us get more information about the people and its context, but the heterogeneity between the devices (i.e. the data sources) is a challenge in terms of the data processing. On the one hand, the data coming from sensors associated with the mobile devices are a concern from the point of view related to the personal security, but on the other hand and for the companies, they represent a big opportunity for making business and getting better knowledge about the consumer behavior.

In this proposal, the Measurement and Evaluation (M&E) projects are organized in terms of the C-INCAMI (Context – Information Need, Concept Model, Attribute, Metric and Indicator) [1, 2] framework. C-INCAMI is a conceptual framework, which defines the concepts and their related components for the M&E. It provides a domain (ontological) model [1] defining all the terms, properties and relationships needed for designing and implementing the M&E processes. In this sense, the idea of the framework is fostering the repeatability, comparability, consistency and extensibility of the measurement process.

Based on the C-INCAMI framework, the CINCAMI/MIS (Measurement Interchange Schema) allowed the interchanging of data between heterogeneous data sources associated with M&E projects. This schema incorporated the possibility of: a) to interchange non-deterministic values, b) to incorporate metadata for guiding the processing strategy, and c) to use data and metadata jointly for detecting inconsistency in the processing (e.g. miscalibrations in the Weather Radar –WR-). Recently, a new version of CINCAMI/MIS [2] was proposed for incorporating complementary data in the measurement. Such complementary data are associated with geographical, audio, video or text data. In this sense, more exhaustive field studies are necessary for evaluating the impact in the performance of the processing. However, the obtained gain in terms of contextual information related to each measure is essential in terms of an eventual data monetization strategy.

The entities under monitoring (e.g. a consumer) are quantifies through an M&E project defined in terms of C-INCAMI. The interchanging of data between the processor and the data sources (e.g. an entity under analysis) requires homogeneity and consistency aligned with the M&E projects. Here is where the C-INCAMI/MIS uses the definition coming from the M&E project for structuring data and metadata jointly.

The Processing Architecture based on Measurement Metadata (PAbMM) is essentially the processor. It is that to say, it is a storm topology [3] oriented to the data stream processing. Thus, PAbMM receives streams organized under CINCAMI/MIS from the heterogeneous data sources (e.g. sensors on mobile devices associated with the customer) and it guides the processing using the data and metadata jointly.
As main contribution, it presents a new monetization strategy based on the CINCAMI framework and its associated measurement interchange schema. The strategy advances over the economical point of view associated with the data and its processing. In this perspective, PAbMM, CINCAMI and CINCAMI/MIS are needed tools for supporting a different way for making business.

This article is organized in seven sections. The section 2 summarizes the extended CINCAMI framework and the measurement interchange schema. The section 3 outlines the processing architecture. The section 4 introduces the monetization strategy. The section 5 presents an application case related to the monetization strategy. The section 6 discusses related works and finally summarizes the conclusions and future works.

II. THE EXTENDED C-INCAMI FRAMEWORK AND THE MEASUREMENT INTERCHANGE SCHEMA

C-INCAMI is a conceptual framework [4, 5], which defines the concepts and their related components for Measurement and Evaluation (M&E) area in software organizations. It provides a domain (ontological) model defining all the terms, properties and relationships needed for designing and implementing the M&E processes. This framework proposes an approach based on the information need. In this sense, the requirements specification, M&E, and analysis of results should satisfy a specific information need, in a given context. The concepts and relationships in C-INCAMI (e.g. Measure, Entity, Context, etc.) are intended to be used along all the M&E activities. This way, the idea is get obtain a common understanding of data and metadata shared among projects for fostering a more consistent analysis.

In [6] 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 [1, 7, 8] in relation with the measurement component was not enough because it considers the value associated with the measure just like a deterministic number. In figure 1, it is proposed 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, is is maintained with painted background the original concepts from C-INCAMI and with blank background the new incorporated concepts or extensions.

As shown in figure 1, 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 it is got a clearly defined value (DeterministicMeasure class in figure 1), or it could be associated with an estimated process in which it is obtained a set of <value, likelihood> pairs (See the classes called LikelihoodDistribution and EstimatedValue in figure 1). 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 it is could use the mathematical expectation for synthesizing the distribution in one representative value.

![Fig. 2. CINCAMI / Measurement Interchange Schema - Version 2 (A=All, S=Sequence and C=Choice).](image)

Each measure could have complementary data that help for better describing the context associated to the entity under analysis. In this sense, the complementary data incorporates an overhead in the data stream processing. Thus, before using them, it is important balancing the necessity of to get better descriptive data (e.g. in a data monetization strategy) in relation to the associated cost in terms of the processing overhead. As you can see in figure 1, the complementary data could be associated with one or more pictures, video, geographic information, audio tracks, or plain text. One important aspect in relation to the complementary data is that we can verify the data integrity through the hashControl attribute (See ComplementaryDatum class in figure 1).

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 1). 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 dsAdapterID attribute identifies the measurement adapter along the M&E projects. The name attribute is a friendly 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.

In general, the data sources are heterogeneous and there are not exists any control over them. Thus, the TraceGroup class represents the possibility of to get a more homogeneous group of the sources based on the similarity between them for better monitoring. For example, the weather radar of EEA INTA...
Anguil has many sensors in which each one measure a specific variable. In this sense, each sensor is a heterogeneous data source in relation to its neighbour sensor, but all they are group under the same trace group called “WR INTA EEA Anguil”. In this way, each sensor can has one or more properties that characterize it.

Finally, it is extends the metric definition incorporating the concept of device’s constraints through the Constraint class (See in figure 1). 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 1) and the kind of filter (filterType in Figure 1. e.g. mandatory or preferable). It is could limit the valid values associated with a data source property by pattern (PatternConstraint Class in Figure 1) or explicitly indicating a set of valid values (ConstraintValue class in Figure 1).

Figure 2 synthesizes the superior level related to the new version of the CINCAMI/MIS schema. This new schema aligns with the mentioned extensions associated with C-INCAMI framework. As you can see, it is possible interchange deterministic and non-deterministic measures with their associated contextual information and incrementing the description by the use of the complementary data. An important thing in this sense, it is that each CINCAMI/MIS stream has associated a footprint, which is associated with a logical window. Thus, PAbMM will verify the footprint each time that a cincami/mis stream is received. If the footprint does not match, then the logical window will be discarded without processing. This represents an important test for warranting the data consistency between the data origin and the responsible for its processing.

Moreover, on the one hand, with the measurementItem tag there is a dataSourceID attribute that identifies the sensor associated with the data origin. On the other hand, the root of CINCAMI/MIS has a dsAdapterID attribute, which give us information about the responsible for the translating between the original format and the CINCAMI/MIS schema. This is a key asset because now it is possible keep traceability between the sensor and each instance associated with the processing in PAbMM.

III. THE PROCESSING ARCHITECTURE BASED ON MEASUREMENT METADATA

The PAbMM is a manager of semi-structured measurements streams, enriched with metadata supported by C-INCAMI, specialized in M&E projects, which incorporates detective and predictive behaviour at online with the ability to manage and provide large volumes of data on demand.

As shown in figure 3, the conceptual model in terms of stream processing it is as follows. The measurements are generated in

MA incorporates together with the measured values, the metadata of the M&E project and reports it to a central gathering function (Gathering Function -GF) using Apache Kafka [9]. The GF distributes in parallel the CINCAMI/MIS streams between the subscribers, the smoothing and analysis function and the big data repository. On the one hand, the subscribers could receive the CINCAMI/MIS streams at the time when it is generated (for example, for INTA weather radar data, a consumer could be the National Meteorological Service). They consume the data stream in real time like a consumer in terms of Apache Kafka [9, 10]. On the other hand, the measurement stream is organized in a buffer by monitoring groups -dynamic way of grouping data sources defined by the M&E project manager- inside the smoothing and analysis function. This allows a consistent statistical analysis at level of monitoring group or by geographic region where the data sources are located, without incurring in additional processing overhead.

The data repository is useful for answering queries about the historical data; carry on different implementations of the machine learning algorithms, and for data visualization. This repository contains the formal definition of the M&E project and the organizational memory (OM) [11]. The formal definition allows that all the components along the architecture have the enough information about the focus in monitoring.

Whether the statistical analysis detects a deviation, or the online classifiers indicate an eventual risk situation, the alarm will be send to the decision maker (See figure 3). The decision maker can make a decision about the next steps, or discard the alarm. In this sense, the organizational memory contains the previous experience structured as key-value. Thus, through a case-based reasoning on the OM, it is possible recommend courses of action to the DM when it need.
The filled frame related to the storm topology in figure 3, it is detailed in figure 4. In terms of Apache Storm, a Spout is source of streams inside a topology, while Bolt is a representation for any querying, transformation and processing activity on the stream [12].

In this sense and when the topology start-up, a metadata spout (See figure 4) send the updated information associated with the M&E projects to the follows Bolts:

- **Metadata Analysis**: for monitoring the measures using the metadata on the fly,
- **Metadata-Us**: the synopsis is a data structure, which allows keep in-memory the last known state on an entity (See figure 5). In this bolt, the synopses are updated and their initial state is built.
- **Training Data set**: In parallel, a CINCAMI/MIS-based data stream is sent through the Training Dataset Spout to the Updating the HT Bolt. It will train the online classifier using the learned experiences from the organizational memory and the indicator’s value as target class.

Once the storm topology has started, the CINCAMI/MIS streams come from the GF (See figure 3) through the **Kafka Spout** and they are distributed in parallel to the follows Bolts in the topology (See figure 4):

- **Updating the HT**: where the Hoeffding Tree is updated with the new data,
- **Applying the current HT**: where the current Hoeffding Tree is applied and the classification from the new data is got, and
- **Metadata Analysis**: whereby means of the metadata inside the CINCAMI/MIS stream, each measure is compared with the M&E project definition for detecting eventual deviations.

On the other hand and once the Hoeffding Tree (HT) has been updated, the new tree is applied for getting a new classification. Followed, the current and updated HT are compared by a ROC (Receiver Operating Characteristic) curve [13] for knowing which model has mayor gaining. If the updated HT is better than the current HT, the updated HT will replace the current model and it will be the new current model, else the current HT will continue with its role.

On the other hand and from the metadata analysis, information about the eventual deviations is sent to the statistical analysis (KS Test, CorrA, PCA & DA in figure 4) and the recommender.

From the statistical point of view, four analyses are performed in parallel using the R Software [14] from the Storm topology:

- **HT**: Hoeffding Tree
- **ROC**: Receiver Operating Characteristic
- **PCA**: Principal Component Analysis
- **CorrA**: Correlation Analysis
- **KS**: Kolmogórov-Smirnov
- **DA**: Descriptive Analysis
- **KS-US**: Updating Synopsis (for example, KS-US means that we update the synopses with data from KS)
a) The Kolmogorov-Smirnov Test with the Liliefors correction for knowing the distribution [15] (KS Test in figure 4),
b) The correlation analysis for detecting dependencies between variables (CorrA in figure 4),
c) The Principal Component Analysis for studying the incidences of each variable in the variability of the system [21] (PCA in figure 4), and
d) The descriptive analysis (DA in figure 4). It allows characterizing the measures and the last known state.

With this information, the corresponding synopses are grouped by trace group in Redis [16] are updated from the Storm Topology (See KS-US, CorrA-US, PCA-US & DA-US in figure 4). All the synopses will update the last known state for each trace group and entity under monitoring (See Updating State in figure 4) allowing answer queries about the entity’s state even when the streams are interrupted.

Thus, the recommender may receive information from (See figure 4):

a) The decision maker, e.g. a risk situation classified following the decision criteria establishes in the M&E project definition,
b) The metadata analysis bolt if an entity has associated a deviation with respect to the expert criteria in the M&E project definition, and
c) The organizational memory with similar cases,

The storing associated with the synopsis is kept in memory using Redis. In this way and as you can see in figure 5, a multilevel hash table for accessing and updating the values is used. Each associated value is in reality another hash table integrated for the entities that compose the trace group and you can access it using the ID of each entity. Finally, each value in the hash table of entities is an array that contains two hash tables. The first hash table is organized by the ID of the metric and keeps in memory the last known measure for the metric. The second hash table is organized by kind of analysis, keeping in memory the last known analysis for KS (Kolmogorov-Smirnov), PCA (Principal Component Analysis), CorrA (Correlation Analysis) and DA (Descriptive Analysis). This new kind of organization for the synopses allows keep in memory the last state of each trace group and entity for an agile answering, even allowing inform the states when the measure streams are interrupted.

IV. THE DATA MONETIZATION STRATEGY

As shown in section III, the PAbMM integrates from the measurement and evaluation project definition and its associated information need, to the online decision making process. In this spectre, you can choose the data sources (and their grouping way by trace groups) to monitor and establish the measurement-processing road (i.e. statistical analysis, classifiers, data replication for the subscribers, etc.).

Initially, this opens two scenarios: a) One scenario is where the data processing happens from the public data; b) but other different scenario happens when the data processing is made on demand from a private project. In the last case, the data are private and corresponds to the customer. It is that to say, when we use public equipment (e.g. a weather radar), it is possible that the data are accessible for anyone who want to use them. However, when the entities under monitoring are trucks from a logistic firm, the data are exclusive to the customer.

The Data Monetization Strategy (DMS) is organized around the possibilities and functionalities given by the PAbMM. The
first big difference happens around the data ownership. As you can see in figure 6, one thing is the private measurement and evaluation project (upper region in figure 6); and other very different things are the public measurement and evaluation projects (lower region in figure 6). In the private projects, the data ownership corresponds to each customer, and it is important protect the data confidentiality. In the public projects, it is possible that the data are public (not always) and an organization could subscribe to one given service. In this sense, a private project is made on demand, while a public project could be made on demand or from a particular public interest. For this reason, the private projects are all post-paid, while the public projects could be pre-paid or post-paid.

Whenever the project is public or private, there are common services with different levels of services (See figure 7). The level of services are organizes such as essential, professional and enterprise. The essential level is oriented to particular persons, which they wish use the services without high rates of pay and with control of the budgets. The professional level is oriented to small firms or even people that use the data as part of the main activity. Finally, the enterprise level is focused in firms that require full data processing and scalability in their associated operations with the measurement.

### Common Services

<table>
<thead>
<tr>
<th>Items</th>
<th>Essential</th>
<th>Professional</th>
<th>Enterprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projects</td>
<td>qProject</td>
<td>3*qProject</td>
<td>Unlimited</td>
</tr>
<tr>
<td>MB/day</td>
<td>mbday</td>
<td>3*mbday</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Availability in the month</td>
<td>Static</td>
<td>Dynamic</td>
<td>Dynamic</td>
</tr>
<tr>
<td>Extending of the MB/day</td>
<td>Unavailable</td>
<td>Available</td>
<td>Dynamic</td>
</tr>
<tr>
<td>Processing Threshold per second</td>
<td>0.003*mbday</td>
<td>0.009*mbday</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Concurrent Users</td>
<td>cu</td>
<td>3^cu</td>
<td>10^cu</td>
</tr>
<tr>
<td>Extending the concurrent users</td>
<td>Unavailable</td>
<td>Available</td>
<td>Available</td>
</tr>
<tr>
<td>Support – Time Response</td>
<td>72 hs</td>
<td>24 hs</td>
<td>Online</td>
</tr>
<tr>
<td>Web UI – Monitoring</td>
<td>Basic</td>
<td>Standard</td>
<td>Full</td>
</tr>
<tr>
<td>Alternation between data sources and trace groups</td>
<td>Unavailable</td>
<td>Unavailable</td>
<td>Available</td>
</tr>
</tbody>
</table>

### Fig. 7. Common Services by level of service

Thus, for better describing the levels of services in figure 7, it is uses a set of variables for establishing the differentiation between levels:

- **qProject** (Quantity of public projects): it represents the quantity of public projects available for online using at the same time. In this case, the projects are accessible and the data are public.
- **qOwnProject** (Quantity of private projects): it represents the quantity of private projects available for online using at the same time. The data ownership corresponds to each customer. Just the users authorized by the customer have access to the measurements.
- **Mbeday** (MB/day): it represents the global allowed limit of MB by day.
- **Smb** ($/MB): it represents the price for each additional MB.
- **cu** (Concurrent Users): it represents the available quantity of concurrent users.
- **Scu** ($/concurrent users): it represents the payment related to each additional concurrent user. It is applicable when the extending of concurrent users is available.
- **ds** (data sources): it represents the global available quantity of data sources.
- **tg** (trace groups): it represents the global available quantity of trace groups.

The limitations between levels of service are organized in turn of the follows items (See figure 7):

- **Projects**: it represents the quantity of projects globally available.
- **MB/day**: it represents the limitations of MB by day in terms of processing.
- **Availability in the month**: it establishes the possibility of change the hour range in which the user could use the service. If it is static, then the user defines an hour range and it is immovable for the entire month. However, if it is dynamic, then the user could change in real time the hour range for connecting to the service.
- **Extending of the MB/day**: it allows adding MB/day when the diary block has been consumed.
- **Processing threshold per second**: it establishes a maximum limit in the processing rate.
- **Concurrent users**: it represents the maximum quantity of users using the service at the same time.
- **Concurrent user extension**: it refers to be able to extend the quantity of concurrent users.
- **Support – Time response**: this characteristic is associated with the response time to the user, from the requesting to the answering.
- **Web UI – Monitoring**: it is associated with the different items to be used inside of the web application in relation to the level of service.
- **Alternation between data sources and trace groups**: it refers to change dynamically the selection of the data.
origins between data sources and trace groups, without take importance the quantity of them.

Figure eight shows the particularity for the level of services in case of prepaid or post paid modality just for public projects. As you can see, when the modality is prepaid, the customer can choose between trace groups or data sources. When the level of service is enterprise, the customer could freely change between them. However and when the modality is post-paid, the customer receives the same services that pre-paid with the initial pricing, being possible the extension. Depending of the level of services, each extension will be charged at the end of month in addition to the base pricing. In this sense, the base pricing is given by the prepaid modality.

<table>
<thead>
<tr>
<th></th>
<th>Level of service</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Essential</td>
</tr>
<tr>
<td>Data source</td>
<td>ds</td>
</tr>
<tr>
<td>Pricing</td>
<td>MF</td>
</tr>
<tr>
<td>Trace group</td>
<td>tg</td>
</tr>
<tr>
<td>Pricing</td>
<td>2,5*MF</td>
</tr>
<tr>
<td>Automatic Extension</td>
<td></td>
</tr>
<tr>
<td>of concurrent users</td>
<td></td>
</tr>
<tr>
<td>Automatic Extension</td>
<td></td>
</tr>
<tr>
<td>of M/B/day</td>
<td></td>
</tr>
<tr>
<td>Automatic fit of</td>
<td></td>
</tr>
<tr>
<td>threshold</td>
<td></td>
</tr>
<tr>
<td>Web UI - Administrative</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 8. Public Projects. Pre-paid versus post-paid**

In the case of private projects (See figure 8A), the modality is post-paid only. Thus, it inherits the possibilities of the post paid modality associated with the public projects, and incorporates specific items related to the M&E project definition. As you can see, the differentiation between level of services is directly associated with the capacities of PAbMM. For example, the last state known is limited in the level “essential” because this level does not have the correlation analysis. Even when the professional level has the correlation analysis, it does not have the principal component analysis, and for this reason, the last state known is limited.

The private projects are oriented for bringing support to the measurement and evaluation activity. In this sense, the extensibility is a current coin, and the difference between levels of services is associated with the used capacities of PAbMM. For example, a logistic firm could use the enterprise level for monitoring the consuming associated with the trucks, the used routes, etc.

**An Application Case of the Data Monetization Strategy**
Fig. 9. The INTA Weather Radar [6]

The National Meteorological Service maintains the WR (See figure 9) of the National Institute of Agricultural Technology (in spanish, Instituto Nacional de Tecnología Agropecuaria – INTA-), Experimental Anguil Station (EAS) and it is located in Anguil (province of La Pampa, Argentine).

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. Its main applications are: a) Weather description, forecasting and nowcasting, b) Forecasting and monitoring of environmental contingencies (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 (i.e. floods) [6]. The data recorded 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) [17].

Two types of data are distinguished: a) raw data and b) some level of data processing or “products”. In both cases, the sampling units are 1 km² and 1º and each variable are stored in separate files [17].

In terms of the data monetization strategy shown in section IV, the proposing of data monetization is associated with a public project in where any person or firm could have access by subscription. The subscription could be related to essential, professional or enterprise level. In addition, the subscription could be pre-paid or post-paid.

In this case, the data are publics and they are generated by the WR [17]. The data ownership corresponds to National Institute of Agricultural Technology and at now, they are accessible. Thus, here it is possible monetarily capitalize the extension of CINCAMI framework and the new CINCAMI/MIS schema from the incorporated complementary data (See section II).

The idea associated with the essential level, it is make accessible the daily data for anyone who needs a sample, e.g. a particular researcher, a student, etc. In this case, the pricing could be established in a symbolic value for getting access easily.

The professional level is accessible for professionals (such as agricultural engineers), who needs more diary data volume for their professional activity (e.g. crop monitoring). In this sense, it is logical that the pricing to be higher than the essential level.

The enterprise level is oriented to the firms that need data volume for specific activities, for example, the insurance firms. In this sense, it is possible that the firms require a number of users for consuming the data in real time. It is highly likelihood that the data consuming is related with internal applications of the firms with a specific purpose. Therefore, the data consuming will have a commercial finality.

The three levels of subscription are oriented to the online processing. However, the collected data by the WR must store for future reference. In this sense, it is possible capitalize the big data repository with historical measurements for generating additional incomes. It is that to say, an insurance firm could need ten years of history related to a specific region for building a hail predictive model. Thus, the consuming of the historical repository could be commercialized by data volume.
Finally, the idea behind the data monetization strategy, it is generate additional incomes which allow extend the current services of INTA and improve the existents.

V. 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 [18, 19]. In the perspective of PAbMM, each data stream is a measurement stream organized under the CINCAMI/MIS 2 schema (See Section II). This measurement interchange schema allows using the metadata for guiding the processing. Thus, PAbMM can maintain the last state known associated with the entities under monitoring and to give to thirds an approximated answer when the original measurement streams are interrupted. It represents a key asset in relation to the data monetization strategy, because even when the current measurements are unavailable, PAbMM is able to provide the last known state and even the last analysis made on the entities under monitoring.

There are works oriented specifically to the processing techniques over data streams, such as online outlier detection [20], the join processing [21], etc. These works are based on raw data (or plain tuples), however the PAbMM manage CINCAMI/MIS streams. In this sense, even when the CINCAMI/MIS add an overhead in the processing, it incorporates specific knowledge on the entity under monitoring, which benefits to the data monetization strategy. It is that to say, when more information around the entities under analysis and their contexts, more possibilities for completing the associated data ecosystem.

In [22], an approach that conduce a semantic segmentation from a real-time sensor data stream to recognize an elderly persons complex activities is shown. PAbMM is based on a measurement and evaluation framework such as C-INCAMI, which allows getting a comparative, repeatable and consistent measurement process. Thus, it is possible to use jointly data and metadata in PAbMM for incorporating detective and predictive behavior.

VI. CONCLUSIONS AND FUTURE WORKS

In the invited talk, we have introduced a monetization strategy based on the CINCAMI framework and its associated measurement interchange schema. The strategy advances over the economical point of view associated with the data and its processing.

In this perspective, PAbMM, CINCAMI and CINCAMI/MIS are needed tools for supporting a different way for making business. The CINCAMI framework fosters the repeatability, consistency and comparability in the measurement process. The CINCAMI/MIS Schema facilitates the measurement interchanging between heterogeneous data sources and PAbMM, and it is guided by a formal project definition sustained in CINCAMI. PAbMM allows providing data in two service modalities: real time and on demand. In the real time modality, the data are provides by subscription via replication over Apache Kafka. In the on demand modality, the data are provides from Apache Hive via query and the payment happens through the interchanged data volume.

In this way, it is possible differentiate between two kinds of monitoring projects: privates and publics. On the one hand, the public projects are oriented to the public without data restrictions. Any person or firm could subscribe via pre-paid or post-paid at the wished level of service. On the other hand, the private projects keep the data ownership in each firm, and it is a determinant factor because it is necessary guarantees the confidentiality of the measurements and the associated complementary data.

An application case in the weather radar of the National Institute of Agricultural Technology located in the Experimental Agricultural Station of Anguil (province of La Pampa, Argentina) was shown in section V. Because the weather is an interesting thing for many activities (e.g. agricultural, insurance, etc), the concrete possibilities of subscription are very interesting. Even, considering that the weather radar is a public heritage, in the level of service called “essential”, a free of charge alternative was proposed. However, the others levels of service is related to a more intensive use of the data and with commercial finality. Therefore, the highest levels of service (professional and enterprise) allow supporting the lowest level (essential) for getting a more wide scope in the society.

As future work, an analysis on the customer behavior by profile related to the proposing of data monetization will be carry on.

ACKNOWLEDGMENT

This research is supported by the project 09/F068 of the Engineering School, the project Res.CD 278/16 of the Economics and Law School and partially, through the technical cooperation contract between the Engineering School and the EAS INTA Anguil (Argentina).

REFERENCES


