An Expert-Driven Approach for Collaborative Knowledge Acquisition

Leo Ordinez[§1], Rodrigo Santos[‡2], Damián Barry[§3], Sergio Firmenich[§4], Alejandro Sánchez[§5]

[§]Laboratorio de Investigación en Informática (LINVI), Departamento de Informática, UNPSJB, Bvd Brown 3050, 9120, Puerto Madryn, Chubut, Argentina
[‡]Dep. de Ing. Eléctrica y Computadoras, Instituto de Inv. en Ingeniería Eléctrica, UNS-CONICET, Av. Alem 1253, 8000 Bahia Blanca, Argentina
[1]`leo.ordinez@gmail.com` - [2]`ierms@criba.edu.ar` - [3]`damian_barry@unpata.edu.ar` - [4]`sergio.firmenich@lifia.info.unlp.edu.ar` - [5]`asanchez@unsl.edu.ar`

Abstract. Knowledge management is key for any organization. Huge amount of data is made available to organizations by pervasive technologies such as smart mobile devices. However, the knowledge to use such data is still missing, and organizations typically fail to exploit it. This paper proposes an architectural design that aims at addressing such problem. It focuses on knowledge management for collaborative systems in which complex and multicausal situations are presented to interacting actors on a large geographical area with possible low connectivity.

1 Motivation

In terms of information and communications technology (ICT), an effective information system is one that provides users with the necessary information to meet environmental requirements [1]. Web technology worldwide and electronic networks have created an environment where the physical location is not important. From the last two decades, advances in ICT have created the conditions to integrate and coordinate information from different actors and build knowledge from them. At this point it is important to distinguish between data, information and knowledge [2]. Data are facts of the physical world obtained through observation. The information adds to those data the necessary context to obtain a meaning of them. Although the information may be used to make decisions, by itself is not enough. It is required prior learning and experience to interpret this information (*i.e.*,contextualized data). This context, together with the corresponding information is knowledge, which allows to constitute a judgment on reality and take actions on it.

The purpose of this paper is the proposal of a reference model for describing and implementing collaborative systems based on ubiquitous computing. The primary motivation of this proposal is based on the field of application of the systems to be developed, which have to do with complex, multi-causal, possibly extraordinary situations involving the interaction of different actors, in a large geographical area, with potentially low level of connectivity. In a different way, these systems must interact with unfavorable contexts both in its technological infrastructure and its environmental particularities. The work focuses on the collection of data in a non-urban terrain, with sporadic visitors, like Patagonian region. In this regard, transformation of the data into information, will allow the building and extraction of knowledge of the Patagonian region. It is worth noting that in the context of this research knowledge is represented through statistical indicators that will support a wise decision making.

It is important to remark that the software application, for the previously described domain, are intended to be specified by *domain experts*, possibly with no computing skills. This is, according to the variability of the potential scenarios, it is necessary to count with a certain degree of experience in order to manipulate the collected data. In this regard, a domain expert is appropriate for this task. However, in general, the expert will not have sufficient computing or programming skills to develop a complete system, as it would a *computer expert*. Consequently, the reference model proposed should consider this aspect and provide the flexibility and mechanisms to allow an expert user to be able to define the software application.

In the here presented approach, collaborative systems are meant for tackling situations in large, heterogeneous and even unpredictable contexts. As mentioned above, a main actor is the *Domain Expert*, which is interested in obtaining information from the environment. Furthermore, he is the one who should define the software application in order to meets his requirements. Nonetheless, another necessary actor present in the system is the one in charge of collecting data, the *Data Collector*. This is an actor possibly inexperienced and yet with no particular interest in the information obtained, but in the process of gathering that information.

With all the contribution of this paper is the proposal of an architecture for collaborative systems, addressed to expert users of a particular domain, who require to manage information in order to make decisions that:

- P1. Allow *Domain Experts* define the data they were seeking and the way in which these data are converted into displayable information automatically (no third-party development, only involve the intervention of the expert or team of experts in the domain).
- P2. Allow changing the above definition at any time, due to volatile information requirements, which emphasizes the variability of the final applications both for displaying information (for *Domain Experts*) as for the implementation of data collection (for *Data Collectors*).
- P3. Allow data collection regardless of connectivity or other conditions that are already mentioned in the article.
- P4. Allow data sources to be managed in a certain agnostic way, allowing to obtain:
  - (a) Data collected by *Data Collectors*, mainly focused on the use of mobile devices
  - (b) Data from sensors and other such devices
  - (c) Data obtained from repositories

P5. Allow deriving automatically the mobile data collection applications for use by *Data Collectors*
P6. Allow users to reward collaborators, for example, by creating a scenario of gamification.

The rest of the paper is organized as follows. In Section 2 a broad review of related works is performed; the proposal of an architecture to develop this kind of systems is done in Section 3; in Section 4 an application case is presented to illustrate the proposal; finally, conclusions are exposed in Section 5.

2 Related Works

Some works propose different levels of standardization of communication protocols to ease the interoperability of heterogeneous systems. A generic framework for cooperative multi-issue one-to-many negotiations is proposed in [3]. The need for a common way to abstract the heterogeneity of devices is outlined in [4], when working on collaboration policy generation technique to govern dynamic collaboration in Service Oriented Architecture (SOA) systems. Dynamic collaboration allows two parties not knowing each other to establish their collaboration protocol at runtime [5]. In [6] the authors employ code generation techniques and use common abstractions, operations and mechanisms at all system levels to support uniform system of systems composition. The complexity of the communication in collaboration system depends on the integration and negotiation of diverse and heterogeneous community societies like companies, governments, universities, ONG, etc. Some of them in the direction of infrastructure towards coordination, switching and choreography of information in heterogeneous systems [3,5,7], others in specifying semantics and standardizing protocols to communicate devices and systems [4,6]. The two previously mentioned alternatives need the concept of data transformation like a graph-based transformation approach proposed in "Towards Choreography Model Transformation via Graph Transformation" [8] or a XSLT transformation model based on XML syntax proposed in "Sociedad de la Información-Switch Transaccional" [7]. The works mentioned above are focused in the description of a particular part of a middleware. They proposed different alternatives to amalgamate heterogeneous messages, transactions or another communication unit more abstract than a packet.

As was mentioned in Section 1, in this paper a distinction is made between a *Computer Expert* and a *Domain Expert*. Following some relevant works oriented to the former are contrasted. In [9] a new interpretation of the Internet-of-Things (IoT) is presented. The authors allow end users to manage the sensing resources populating the involved devices, not just the mere data output they produce, by using virtual devices in the cloud. The model is focused in sensing/actuating devices without human intervention. In this regard, heterogeneity is given by the different kinds of resources (devices) available, which in our proposal would constitute one data source. A similar approach concerning management of heterogeneity is presented in [10] and [11]. The proposal is centered in Cyber-Physical Systems, which report data from sensors to different databases. Although the architecture proposed is very interesting and could be extended to other data sources, the objective of the papers was to reduce energy consumption by lowering the communication rate. In [12] the authors proposed an architecture similar to the one proposed in this paper. One of the aspects they focused is maintain consistency between transactions. To this, they based their approach in graphs and logic clocks.

Other important aspect of the collaboration system architecture is the adaptive functionality of the components needed to interconnect systems to solve a domain specific problem. In this regard an interesting approach is presented in [13], which provides a generic software and middleware components to ease the design and development of mass market context-aware applications built above the Internet of Things. In [14] the authors made a proposal to ease the development process of Agent-Based Pervasive Systems, providing the user with a set of abstractions that ease the implementation of Pervasive Systems and the deployment of a platform for their execution. Finally, also in IoT application domain, [4] presents a semantic web knowledge representation technologies and interaction with devices and the physical world to achieve access and modifying their virtual representations.

The approaches described in the previous paragraph are centered in a specific domain. In what follows, a set of articles which pursuit the objective of being oriented to Domain Experts, are analyzed. A recent work can be found in [15]. The authors report an ontology-based visual query system to specify and extract data of interest for non-technical users. The article is centered in usability and consequently not the architecture nor the middleware are described. In addition, the work does not consider heterogeneous sources in the sense that they are established here. A full architectural description that manages heterogeneous data sources is presented in [16]. Finally, another detailed description of an architecture and its components is proposed in [17]. An interesting contribution is given by a file format used for exchanging multimedia content. Another contribution is the inclusion of a Message agent, that handles message queuing through a "store and forward" mechanism and that translates messages, providing the usable information to upper layer applications.

To sum up, it is important to emphasize the difference between the different approaches presented in this section and the proposal of this paper. In particular, some of the approaches do not consider the expert user as playing a "developer role" (also known as *end-user development*). On the contrary, the user is taken as a passive receiver of the information, but not as the one who set the basis, parameters and design the application to gather that information. Having an *Domain Expert* acting as a developer and counting with an architecture and mechanisms to support that task, extends the number of possible scenarios and domains to consider. Consequently, the experts reduce their dependence on *Computer Experts*, when their information requirements slightly change or new decision scenarios are presented. In addition, another actor plays a key role, the *Data Collectors*, which are not interested in the results nor in the development, but in the process.

3 An Architecture for Collaborative Systems

A Collaborative System is one in which its components are inherently distributed, data is collected from different sources, information is obtained by processing data in a particular context and knowledge is produced to fulfill the requirements of an expert in a particular field of interest. Any entity or role played by a person or another system consisting in gathering data through automatic, electric, mechanical, observed or indirect means, satisfying sources D1., D2. or D3. is considered a *Data Collector*.

As mentioned in Section 1, in general three data sources are considered:

D1. Data collected by collaborating users
D2. Data from sensors and other such devices
D3. Data obtained from repositories

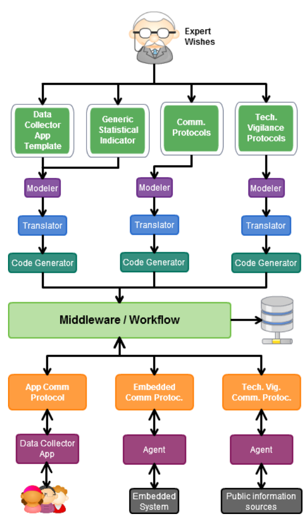

Fig. 1. Architecture for Collaborative Systems based on Ubiquitous Computing.

Figure 1 shows the architecture proposed in this paper for tackling Collaborative Systems. From top to bottom it can be seen how the expert's wishes traverse multiple phases until being concretely implemented. The boxes represent components which perform data transformations or products which are the generated by those transformations. The arrows represent the direction of the information flows. The idea behind it is as follows.

The expert in a particular field intents to obtain information from its application domain. That information comes from data that is gathered from multiple sources that is eventually contextualized. The different sources from which it can obtain the information gives the context in a certain level of detail. For example, if information is taken from sensors they must be normalized and electrical measures turned into physical quantities. On the other hand, if information retrieval from a patents repository is chosen (so technological vigilance protocols are used), the gathered information is already contextualized.

In addition, the expert defines a series of statistical indicators the he wants to get from the information obtained.

Note that the previous paragraphs also establish a workflow for developing a Collaborative Systems. Figure 2 shows a sketch of that workflow.

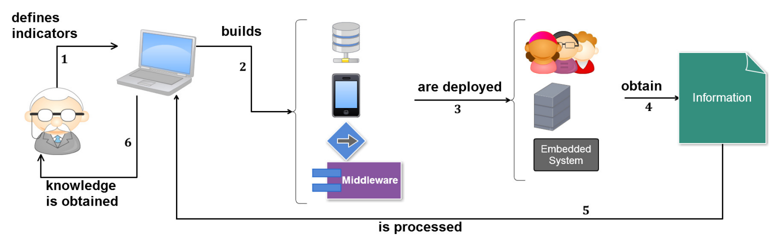

Fig. 2. Workflow for Developing a Collaborative Systems.

To the aims of this proposal, information could be obtained from three main sources. The first one is collected through human intervention by means of a smartphone app (D1.). This app could be as simple as a data collector similar to filling in a spreadsheet; or more complex like a game [18] or citizen science [19] app. The second information source is from an embedded system, in which case the expert has to define the communication protocol with that system (D2.). The third source is constituted by online repositories (D3.), which are accessed through technological vigilance protocols [20].

The former concerns the requirements definitions made by an expert in a particular application domain. In the specific case of this paper, the architecture has to consider some non-functional constraints, which have to do with the environmental conditions of Patagonian region itself. These aspects include communication mechanisms in a wide area, energy-aware devices, data consistency and availability of data and devices, among others.

Independently of the information source, the decision made by the expert is then traversed through a chain of tools that perform transformations on each step resulting in products that feed the next step. In general, the graphical model defined by the expert is then transformed into a syntactical model, which is then transformed into an executable code. The generated code would be downloaded to a smart device, integrated to the embedded system or deployed in a server to retrieve information.

4 Collaborative Knowledge in the Wool Production

In this section, an application case in which the authors are working is presented. It is focused in a particular productive sector in the agriculture field, which is the wool sector.

Currently the process of shearing is done manually and with registration of all information with low-tech methods: analog scales, bale processing cards made by hand, recording of all information made on handwritten forms. In order to establish a set of best practices, the Ministry of Agriculture implemented the PROLANA Program for the purpose of regulating the activity, but mainly to ensure the wool export process, improving international ratings of Argentinian wool. By improving this rating, producers are enabled to obtain better international prices.

PROLANA has detected that the main problems encountered in the shearing process are:

1. Errors in weighing the bale;
2. Failures in registering the burden weighing on the scoresheet wool book;
3. Inability to record the weight of the burden on the individual card thereof;
4. Mismatch between the kind of wool recorded on the card and the one registered on the scoresheet wool book;
5. Registration of kinds of "non-existent" wool;
6. Letter and / or illegible numbers;
7. Forget to include the identification card of each bale;
8. Creating bundles "twins", by doubling the number of burden;

The actors involved in the shearing process are:

Wool Conditioner: His main purpose is to supervise the overall process of shearing and quality control of wool packaging. PROLANA certifies its work and is responsible of capturing data in field.

Shearing Company: Its purpose is to execute the overall shearing process in the field. PROLANA enables his work and guarantee and regulate their contracts.

Producer: His main purpose is to offer his wool to the association of the wool export system. PROLANA guarantee the quality process control of the shearing and proper stowage and storage of wool until exporting process.

PROLANA: Its main purpose is to guarantee the overall quality process of shearing wool, packaging, stowage and storage, and properly export of it.

In addition to the multiplicity of actors involved in the shearing process, interested organizations such as the Ministry and the Livestock Associations are added. It is therefore necessary to have a platform to work with all the actors mentioned in the first instance and to ensure the reliability of information acquired at different points in the process.

This project seeks to provide sheep producers in the Patagonian region a semiautomatic electronic recording mechanism. This device (commonly a rugged handheld) can be connected to an electronic scale to automatically record the weight of the bales. In addition, the equipment must be compatible with a printing mechanism which has its own battery and battery charger. This would allow to comply with all registration requirements that PROLANA has and minimizes errors in loading and transferring of information.

The information stored by a team will be loaded in a software specially designed to automatically make the summary of the wool book with electronic signature (Wool Conditioner, Producer, Shearing Company) and then move to print worksheets and their respective summary wool book with the necessary information.

The information will be transmitted to the PROLANA server located in Rawson, the capital city of Chubut. This will allow each wool book information, to be available immediately in the PROLANA's central of information.In addition to the information for the producer, PROLANA builds statistical indicators that allow a conscious decision making process and political planning.

For the particular case of a *D1.-based system*, such as the one presented in this section, the architecture becomes a double case of a 3-tier architecture [21], as it is shown in Figure 3: a global 3-tier, which includes a server for collecting information and processing it; and a local 3-tier, deployed in several smart devices.

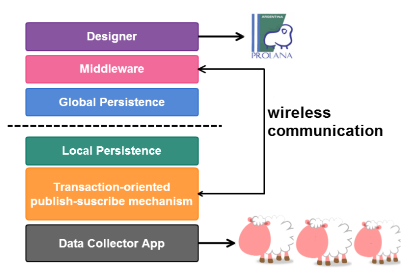

Fig. 3. Two 3-tier architectures.

The main objective is to avoid loading the wrong information about wool production of the Province of Chubut. Nonetheless, PROLANA also uses the information to plan the actions to be taken in the next years. Thus, building knowledge. Note that an inherent consequence of it, is the variability of the information requirements that may exist for the next production of wool, because of the knowledge acquired. This exposes a mutual reliance between the information system and the physical world. To sum up, the presented case shows how the architecture proposed adapts to the following premises enumerated in Section 1: P2., P3., P4.(a), P4.(b), P5. and P6.

5 Conclusions and Future Works

The fact of having a reference model to obtain knowledge from field data is extremely useful in decision making. Knowledge is built from different information sources, which in terms are contextualized data gathered from the field. In the context of this paper, the system must interact with unfavorable contexts both in its technological infrastructure and its environmental particularities. The work focuses on the collection of data in a non-urban terrain, with sporadic visitors, like Patagonian region. These kinds of systems can be classified as Collaborative based on Ubiquitous Computing, since the collection of data is done in an environment of sparse population and, in many cases, by means of embedded autonomous systems.

In order construct that knowledge, a reference model is needed. In this paper, the contribution made is the proposal of an architecture to manage those kinds of collaborative systems. Finally, a representative application case was exposed. The case illustrated the importance and applicability of the proposal done in this paper. Our approach could fit with unexpected and variable scenarios where information could be defined by Domain Experts without computing skills. For instance, urban traffic information, businesses' statistical information or biological observation.

The future works are related to two lines that we are working on. In the first one, we are detailing the proposed architecture under the 4+1 architectural view model, defining formally the model for supporting this kind of systems and applying a consistency protocol for the middleware. Secondly, and considering some good practices of the end-user development literature, we are working in the domain experts tool prototype. With it, he would be able to define the mappings of the model, from establishing communication protocols to setting the data he wants to collect, even when he does not have programming skills.

References

1. El-Bibany, H., Katz, G., Vij, S.: Collaborative information systems. Technical Report TR048, Stanford University (april 1991)
2. Bellinger, G.D.C.A.M.: Data, information, knowledge, and wisdom. (2004)
3. Scafes, M., Badica, C., Pavlin, G., Kamermans, M.: Design and implementation of a service negotiation framework for collaborative disaster management applications. In: 2010 International Conference on Intelligent Networking and Collaborative Systems, IEEE (2010) 519–524
4. Kiljander, J., D'elia, A., Morandi, F., Hyttinen, P., Takalo-Mattila, J., Ylisaukko-Oja, A., Soininen, J.P., Cinotti, T.S.: Semantic interoperability architecture for pervasive computing and internet of things. Access, IEEE **2** (2014) 856–873
5. Tsai, W.T., Huang, Q., Xiao, B., Chen, Y., Zhou, X.: Collaboration policy generation in dynamic collaborative soa. In: Autonomous Decentralized Systems, 2007. ISADS'07. Eighth International Symposium on, IEEE (2007) 33–42
6. Blair, G.S., Bromberg, Y.D., Coulson, G., El Khatib, Y., Réveillère, L., B Ribeiro, H., Rivière, E., Taïani, F.G.D.: Holons: towards a systematic approach to composing systems of systems. (2015)

7. Barry, D.P.: Sociedad de la información-“switch transaccional”. In: XV INTERCON – Trujillo 2008, IEEE (2008)
8. Han, F., Kathayat, S.B., Le, H., Brek, R., Herrmann, P.: Towards choreography model transformation via graph transformation. In: Software Engineering and Service Science (ICSESS), 2011 IEEE 2nd International Conference on, IEEE (2011) 508–515
9. Distefano, S., Merlino, G., Puliafito, A.: A utility paradigm for iot: The sensing cloud. PERVASIVE AND MOBILE COMPUTING **20** (2015) 127–144
10. Mundra, A., Rakesh, N., Tyagi, V.: Query centric cps (qcps) approach for multiple heterogeneous systems. CoRR **abs/1306.6397** (2013)
11. Mundra, A., Rathee, G., Chawla, M., Rakesh, N., Soni, A.: Transport information system using query centric cyber physical systems (QCPS). CoRR **abs/1401.3623** (2014)
12. Ives, Z.G., Green, T.J., Karvounarakis, G., Taylor, N.E., Tannen, V., Talukdar, P.P., Jacob, M., Pereira, F.: The orchestra collaborative data sharing system. SIGMOD Rec. **37**(3) (September 2008) 26–32
13. Arcangeli, J.P., Bouzeghoub, A., Camps, V., Canut, M.F., Chabridon, S., Conan, D., Desprats, T., Laborde, R., Lavinal, E., Leriche, S., et al.: Income–multi-scale context management for the internet of things. In: Ambient Intelligence. Springer (2012) 338–347
14. Agüero, J., Rebollo, M., Carrascosa, C., Julián, V.: Mdd-approach for developing pervasive systems based on service-oriented multi-agent systems. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal **2**(3) (2013) 55–64
15. Soylu, A., Giese, M., Schlatte, R., Jimenez-Ruiz, E., Ozcep, O., Brandt, S.: Domain experts surfing on stream sensor data over ontologies. In: Proceedings of the 1st Workshop on Semantic Web Technologies for Mobile and Pervasive Environments (SEMPER@ESWC 2016). Volume 1588., Heraklion, Greece, CEUR-WS.org (2016)
16. Prasad, S.K., Madisetti, V., Navathe, S.B., Sunderraman, R., Dogdu, E., Bourgeois, A., Weeks, M., Liu, B., Balasooriya, J., Hariharan, A., Xie, W., Madiraju, P., Malladi, S., Sivakumar, R., Zelikovsky, A., Zhang, Y., Pan, Y., Belkasim, S. In: SyD: A Middleware Testbed for Collaborative Applications over Small Heterogeneous Devices and Data Stores. Springer Berlin Heidelberg, Toronto, Canada (2004) 352–371
17. Su, X., Prabhu, B.S., cheng Chu, C., Gadh, R.: Middleware for multimedia mobile collaborative system. In: in Wireless Telecommunications Symposium, 2004. (2004) 112–119
18. : Online citizen science games: Opportunities for the biological sciences. Applied & Translational Genomics **3**(4) (2014) 90 – 94 Global Sharing of Genomic Knowledge in a Free Market.
19. Rogers, D., McCann, T., Cooper, A.: Integrating paper-based habitat mapping with mobile electronic field recording procedures. Ecological Informatics **5**(6) (2010) 479 – 483
20. Barry, D., Cortez22, J.M., Aita12, L.I., Stickar, R.: Extracción automatizada para la unidad de vigilancia tecnológica e inteligencia competitiva de la patagonia
21. Eckerson, W.W.: Three Tier Client/Server Architecture: Achieving Scalability, Performance, and Efficiency in Client Server Applications. In: Open Information Systems 10. (January 1995)