# Seamlessly and Securely sharing health care data with Triple Space Communication

E. Della Valle[1], D. Cerri[1], A. Ghioni[1], D. Cerizza[1]
[1]CEFRIEL – Politecnico di Milano, dellavalle@cefriel.it, cerri@cefriel.it, ghioni@cefriel.it, cerizza@cefriel.it
Via Fucini, 2, 20133, Milano, Italy

***Abstract*- The eHealth is an eBusiness scenario in which the integration problem is amplified by the intensive use of knowledge, by the need of accurately handling citizens' privacy and by live or death implications. eHealth has been seeking for semantic interoperability for more then a decade, but securely sharing health data among healthcare organizations remains an open challenge. Many standardization activities (such as HL7 CDA, openEHR, CEN ENV 13606 / EHRcom, DICOM, IHE) are addressing this problem but none of them has achieve the desirable level of flexibility.**

**In this paper we describe how sharing of health care data respecting parties' autonomy and citizens' privacy is addressed in TripCom, which is a EC funded project starting in April 2006) aiming at developing an highly scalable, semantically enhanced communication infrastructure. Such infrastructure is the result of the integration of Tuple Space, Semantic Web (triple), and Web Services technologies. Tuple Space and Web services and provides platform for application integration based on persistent publication. Semantic Web provide machine processable semantics in order to allow mechanized integration of services (data and processes). Decoupling communications TripCom will reduce (if not eliminate) the need for a priori knowledge of the partner and communication channel thus enabling multi-party interaction for free.**

## I. Introduction

The healthcare organizational structure in all countries is naturally distributed, being a geographical spread of centers at different levels of complexity: from the general hospitals down to individual physicians. The ultimate objective of such a structure is to build a network of complementary centers (hospitals, laboratories, ambulatories, coordination centers, etc.) spread over the territory, to meet effectively the social needs in the area.

This necessary distribution makes it very difficult for clinicians to capture a complete clinical history of a patient, because a patient's health information may be spread out over a number of different institutes or different departments within the same healthcare institute. As a matter of fact, the medical and economic impact of not knowing a patient's complete medical history is profound. Medical practice today still entails sorting through a stack of lab reports, trying to find information on a specific patient. But tens of thousands of people die each year due to lack of information. As reported in many studies, poor information is the "leading killer" in the Western World.

An open challenge in eHealth is (see strategic objective 2.4.11 of 2005-06 Work Programme of IST) to allow health professionals' timely interaction with heterogeneous and distributed medical databases. So, the key problem to address is exchanging patient records among healthcare organizations or among different units within the same organization: radiology, cardiology, neurology, etc.

The main purpose of such a mechanism is to provide healthcare organisations with a complete array of patient information. It should firstly identify a patient and then it should locate the patient's information, including clinical results and labs. The access to and the availability of this information have to be authorised, according to rules determined by the data owner through a data access security policy engine.

## II. State-Of-The-Art

A powerful integration technology, which allows for immediate access to distributed information, is needed in order to provide healthcare organisations with a complete array of patient information.

A number of standardization initiatives are progressing to address this interoperability problem such as:

- HL7 (Health Level Seven) [1], a non-profit, ANSI accredited Standards Developing Organization, founded in 1987, that provides standards for the exchange, management and integration of data to support patient clinical care and the management, delivery and evaluation of health care services;
- GEHR/openEHR [2], an initiative that fosters EHR interoperability started in 1992 as the "Good European Health Record" EU research project that is currently maintained by the openEHR Foundation;
- CEN/TC 251 [3], the technical committee on Health Informatics of the European Committee for Standardization, that, since 1998, is standardizing CEN EN 13606 / EHRcom [4, 5]; and

- IHE (Integrating the Healthcare Enterprise) [6], a not-for-profit initiative founded in 1998 that does not develop standards as such, but selects and recommends appropriate standards for specific use cases.

Most of those initiatives have been active for more than a decade and, after a first attempt in specifying the format of each of the message that can be exchanged among any pair of systems (e.g., HL7 v2.x [7]), they realized that they need to derive messages and interaction patterns from a common shared conceptual model.

In 1999, CEN/TC 251 was the first to introduce, with CEN ENV 13606 / EHRcom, a list of machine-readable terms to be used for structuring EHR content. The standard defines an EHR information model and a modeling approach for deriving concrete interoperable messages to be exchanged between heterogeneous EHRs. However, the single-level modeling approach, the big number of optionality and the high level of abstraction limited market uptake.

GEHR/openEHR in 2002 moved a step forward proposing the archetype concept [8] and the respective two-level methodology. The first level specifies a health care domain reference model [9, 10] that contains concepts such as role, act, entity, participation, observation, etc., while the second level specifies health care and application specific concepts such as patient, GP, lab result, modelling them as archetypes. Each archetype constrains a set of concepts in the reference model (e.g., "Observation") to a specialized data structure (e.g., "Blood Pressure") and defines the vocabulary, such as SNOMED [11] or LOINC [12], to be used within instances of the archetype. The formal language for expressing archetypes, introduced by the openEHR initiative, is the Archetype Definition Language [13]. A complete example of the "Blood Pressure" Archetype is available in [14].

An alternative approach is offered by HL7 Reference Information Model (RIM) [15] which is the ultimate source from which all HL7 v3 protocol specification standards [7] draw their information related content. The RIM model is an explicit data semantics model by which the messages can be implemented locally and top-down, emphasizing reuse across multiple contexts. Moreover, RIM offers a formalism for vocabulary support that permits to get domain concepts from the best terminologies (SNOMED, LOINC, etc.).

## III. Current Trend Towards Semantic Interoperability

All the proposals for standardizing an application protocol for the health care sector may differ in the progress achieved, but they are all similar in concept and capabilities. All of them try to address the interoperability problem by introducing a shared conceptual model (i.e., an ontology). This is very similar to the Semantic Web Services approach in which "semantic interoperability" is achieved by modeling, at a conceptual level, Web Services and the domain they are deployed in. In all eHealth standardization efforts, data structure and sequencing information are enhanced with semantic information that encodes the definition of each element of data including its relationship with other elements. Differently from Semantic Web Services, all eHealth standardization efforts are focusing on developing a horizontal ontology to capture the health care information reference model, which can be linked to the most appropriate vertical domain ontology specifying domain vocabularies. In this sense, eHealth standardization efforts lack:

- the possibility of dealing with systems that commit to different horizontal (e.g., one uses HL7's RIM in CDA, the other uses openEHR archetypes based on EN 13606 RIM) and vertical ontology (e.g., one uses SNOMED, the other some proprietary coding), and
- a comprehensive model for automating service usage such as discovery, choreography and mediation, at both data and process level.

So, even if a clear trend toward a harmonization can be perceived and many people expect a unification of the reference information models, nevertheless such result will only be achieved in the long term and many systems, implemented following different version of all these standard protocols, will be on-line even longer. For all these reasons, we believe that eHealth could greatly benefit from the adoption of Semantic Web Service technology [22].

## IV. Triple Space Computing

Triple Space Computing (TSC) is an innovative paradigm [16] that is taking a significant move towards a new era of the Internet. As reported on the home page of TripCom [17], the leading European project in this area, since the invention of the Internet in the 1960's, the two major evolution steps were email and The Web. **Email** changed the communication processes of humans by providing instant communication over any geographical distances in an asynchronous fashion based on the **message-exchange paradigm**. The **Web** changes communication processes of humans by providing instant publication over any geographical distances in an asynchronous fashion. It is based on broadcasting via **persistent publication** of information. So, the two major asynchronous styles of human communication (direct communication via mail and indirect communication via publication) have been significantly improved through email and Web.

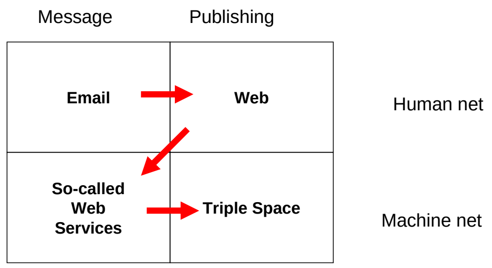


Figure 1. Evolution of communication means for humans and machines as presented in [18]

TSC [18] asserts that the next step for the Internet is likely to be the direct integration of applications and computers via Web service technology. This network no longer directly interlinks humans but interlinks applications and programs to provide integrated services to the human end-user. However, current Web service technology has only very little to do with the Web. It is based on the message exchange paradigm similar to email communication. Truly Web-enabled Web services will communicate via persistent publication of information (see figure 1).

Realizing this vision and a new technology is the mission of TripCom with the result of the integrating Tuple Space [19], Semantic Web [20] (i.e., the triple based data model of RDF [21]), and Web service technologies (see figure 2).

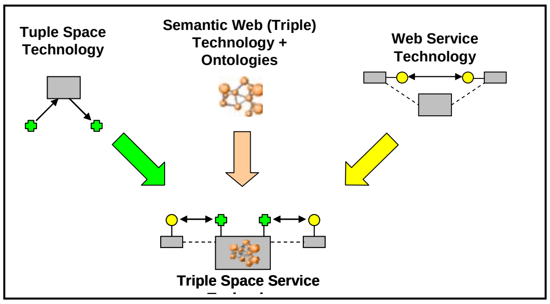


Figure 2. Triple Space Computing results from the integration of three existing and well known technologies.

To this end, on the one hand TripCom plans to improve Tuple Space technology by adding semantics and means to structure and relate tuples in a scalable and linkable Triple Space architecture. On the other hand TripCom plans to improve Web service technologies by adopting the flexible and powerful asynchronous communication model of Tuple Spaces. Furthermore TripCom plans to improve business data exchange standards by use of our new technology and demonstrate the usefulness of this approach in several practical use cases. Finally, TripCom plans to establish a proper security and trust model for the Triple Space to ensure safe communication and data handling, as well as distributed trust models. As the result of the project the combination of these building blocks could give ground to a novel Semantic Web service paradigm.

TSC aims at offering an infrastructure that scales conceptually on an Internet level. Like the Web supports the publishing of Web pages for humans to read, TSC supports publishing of machine-interpretable data. The main advantages in TSC approach, w.r.t. current message-based solution are:

- **time autonomy –** providers of data can publish data at any point in time;
- **location autonomy –** once published the data becomes independent of providers internal storage (thus available even if the providers is not on-line);
- **reference autonomy –** providers are independent of the knowledge about potential readers; and
- **schema autonomy –** the data are represented independently for any provider internal data schema.

## V. Sharing Health Care Data: Message-based vs. TSC-based

As we described in section II, the state of the art in sharing health care data is achieved by standard application protocols (e.g., HL7, openEHR, EHRcom and IHE), which define meaningful components of the messages to be exchanged, and domain vocabularies (e.g., SNOMED, LOINC, etc.), which define the meaning of the data transported by each message. On the contrary TSC enables communication via persistent publication of the information

Such message-based communication has proven efficient and effective for certain activities in this area (i.e., hospital administration), but has shown some problems to effectively and seamlessly collecting and integrate data from electronic health records. When addressing such a need, at least two are the possible solutions. On the one hand a possible solution is to build **centralized databases** that would contain all the medical records on every patient. It would also incorporate all of the different access rules and policies regarding different users and different levels of access. But this kind of efforts has **four weak points**:

- the *cost* of building the infrastructure and collecting the data is enormous,
- the centralized repository approach creates *competitive and security issues* about who controls and has access to the information on a specific patient,
- the *difficulty in maintaining up-to-date* a repository originating from a large number of independently evolving systems, and
- last but not least a message once sent (especially in an asynchronous scenario) gives the owner a *sense of disengagement instead of* strengthening the sense of *ownership.*

On the other hand a possible solution is to **exchange messages only when needed**. In this way no central repository is required and the ownership of the data seems respected, but this solution has **several weak points too**:

- each recipient must *know in advance where to look for information*,
- each recipient must *know in advance the terminology* (e.g., SNOMED, LOINC) *to use* when asking for a specific record content,
- each recipient ends up maintaining *a specific interface for each system* it has to interact with, and
- *mining* (for disease prevention, early diagnosis, pharmaceutical research, enhancement of patient safety) *becomes almost impossible* due to the large amount of messages to be exchanged

On the contrary, Triple Space Computing provides an innovative solution to health and medical data sharing among heterogeneous, distributed environment, because TSC is focused on persistent publishing of knowledge (information augmented with semantics) and not on its collection and distribution. Such a new technology will allow authorized users to identify which health care data are available and where they are located in order to share them when necessary (e.g. authorized physicians will have a complete view of the treatments their patients are receiving, and this is very important for chronic diseases as diabetes). Practical strong points in using TSC are:

- **it is a realistic solution for the data ownership problem** because healthcare organizations will not loose their control over resources and they will be able to share information only with those that are authorized,
- **it provides a simple way to guarantee consistency** because health data won't be neither transmitted or copied but simply used,

- **it supplies a straight forward way to deal with integrity** because data won't be transmitted and it should be impossible for anybody, but the owner, to modify the data,
- finally, **it is a cost-effective solution** because additional storage resources (and related management cost) are drastically reduced.

## VI. Conclusions

Like the Web changed the networking of humans from email exchange to persistent publication, Triple Space Computing aims at revolutionize the networking of machines. The EC is fostering this activity in TripCom, a FP6 funded project starting in Aprile 2006.

Within the timeframe of the project TripCom will be put at work for sharing health care data respecting both parties' autonomy and citizens' privacy. We believe that this eHealth scenario is a very demanding test-bed for TripCom because it poses significant challenges in terms of

- **Interoperability**: TripCom should provide a distributed infrastructure that enables maximum decoupling (i.e. time, space, information schema and terminologies) between the various recipients that own the information (e.g. labs, GP's patient file, hospital information systems, etc.) and those recipients that need to elaborate such information (e.g. an application on board ambulance). In other words, TripCom is required to support different eHealth services (ranging from general practitioner electronic patient records to hospital information systems, including mobile devices like those used by nurses in homecare) in writing information in a way that other eHealth services can later access such information without regards to the standard they implement (HL7, ENV13606, etc.).
- **Information security and trust**: TripCom should enable the enforcement of Authentication and Authorization rules in a distributed way which is not commonly available and quite worth in highly decentralized scenarios such as healthcare, in which every party involved is responsible for keeping the ownership of the data, but the health of citizens depends on the ability to trustworthy sharing data.

Shortly we expect TripCom will provide a basis for **wide, trustable and confidential access to patient information**.

## Acknowledgment

The work described in this paper was partially funded by the European Commission under the projects TRIPCOM (IST-4-027324-STP), COCOON (FP6 IST-507126) and by Italian government under the ITEA project Nomadic Media.

## References

[1] HL7 Version 3 Message Development Framework. http://www.hl7.org/library/mdf99/mdf99.pdf .
[2] openEHR Community. http://www.openehr.org/ .
[3] European Committee for Standardization – Technical Committee on Health Informatics. http://www.centc251.org/ .
[4] CEN ENV 13606. Medical Informatics – Electronic healthcare record communication. European Prestandard ENV 13606, European Committee for Standardization, Brussels, Belgium, 2000.
[5] CEN prEN 13606-1. Health informatics – Electronic health record communication – Part 1: Reference model. Draft European Standard for CEN Enquiry prEN 13606-1, European Committee for Standardization, Brussels, Belgium, 2004.
[6] Integrating the Healthcare Enterprise. http://www.ihe.net/.
[7] The HL7 Version 3 Standard: Clinical Data Architecture, Release 2.0, ANSI Standard, 2005.
[8] T. Beale. Archetypes: Constraint-based Domain Models for Future-proof Information Systems. In OOPSLA 2002 workshop on behavioural semantics, 2002. http://www.deepthought.com.au/it/archetypes/archetypes_new.pdf.
[9] T. Beale and S. Heard. The openEHR EHR Service Model, Revision 0.2. openEHR Reference Model, the openEHR foundation, 2003.
[10] T. Beale, S. Heard, D. Kalra, and D. Lloyd. The openEHR Data Structures Information Model, Revision 1.6rc1 (Release 1.0 draft). openEHR Reference Model, the openEHR foundation, 2005.
[11] SNOMED (The Systematized Nomenclature of Medicine) Clinical Terms. http://www.snomed.org/snomedct_txt.html.
[12] Logical Observation Identifiers Names and Codes (LOINC). http://www.loinc.org/.
[13] T. Beale and S. Heard. Archetype Definition Language (ADL), Revision 2.0rc1 (Release 1.0 draft). openEHR Specification, the openEHR foundation, 2005.
[14] Complete Blood Count Archetype ADL Definition. http://www.openehr.org/repositories/archetype-dev/adl_1.1/adl/archetypes/openehr/ehr/entry/openehr-ehr-observation.haematology-cbc.draft.adl.html.
[15] HL7 Reference Information Model. http://www.hl7.org/library/data-model/RIM/modelpage_non.htm.
[16] D. Fensel: *Triple-space computing: Semantic Web Services based on persistent publication of information*, Proc. of IFIP Int'l Conf. on Intelligence in Communication Systems, Bangkok, Thailand, November 2004: 43-53
[17] The TripCom Project. http://www.tripcom.org
[18] R. Krummenacher, M. Hepp, A. Polleres, Ch. Bussler, and D. Fensel: *WWW or What is Wrong with Web Services*, Proc. 3rd European Conf. on Web Services ECOWS2005, Växjö, Sweden, November 2005: 235-243.
[19] D. Gelernter: Generative Communication in Linda, ACM Transactions in Programming Languages and Systems (TOPLAS), 7(1), 1985: 80-112.
[20] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web", The Scientific American, May 2001.
[21] G. Klyne and J.J. Carroll (eds.), "Resource Description Framework (RDF): Concepts and Abstract Syntax", W3C Recommendation, Feb 2004, http://www.w3.org/TR/rdf-concepts/
[22] Della Valle E., Cerizza D., Bicer V., Kabak Y., Laleci G., and Lausen H. The need for semantic web service in the ehealth. In W3C workshop on Frameworks for Semantics in Web Services, 2005.