

An experience of integration of service repositories

C. Batini, M. Castelli, G. Viscusi Department DISCo - University of Milano Bicocca, viale Sarca 336 - U14 - 20126 Milan, Italy {batini;castelli;viscusi}@disco.unimib.it

ABSTRACT

In the service development life cycle it is worthwhile to distinguish between a conceptual phase, that leads to model abstract services, and a production phase that produces concrete services. Both abstract and concrete services produced by a provider organization can be organized, for reuse purposes, in structured repositories of services. As occurs in database design for conceptual schemas, in a single provider organization several repositories of services may coexist, potentially characterized by heterogeneities and conflicting representations. In this paper we present an experience of integration of repositories of abstract services, based on a methodology currently under development.

Indexing terms/Keywords

Service science, service, abstract service, service repository, integration, correspondence.



Council for Innovative Research Peer Review Research Publishing System

Journal: International Journal of Management & Information Technology

Vol.4, No.2 editor@cirworld.com www.cirworld.com, member.cirworld.com



1. INTRODUCTION

In the last ten years Service Science [7] has evolved from a pioneering field to an emerging multidisciplinary research area, attaining a growing mainstream resonance and scientific recognition. The consolidation of the area is currently an on-going challenge reflected by the number of frameworks and modeling efforts available in the contributing disciplines, such as information systems, service marketing, service engineering and computing.

Notwithstanding the multidisciplinary efforts, the design and planning of services in digital service ecosystems still sees a focus on the technological perspective as the prevailing one. Indeed, the Service Oriented Computing paradigm and service oriented design and development methodologies support the realization of service-based ICT infrastructures [23], covering only one perspective among the many considered in initiatives aimed at designing and planning services in service ecosystems. As pointed out by [12], a service ecosystem is a marketplace for trading services that are developed, published, sold and used; accordingly, design and evaluation frameworks are required, considering all the different features of a service ecosystems: from technology to social and psychological issues implied by the service experience.

Due to the above limitations of actual research on services, several researchers are investigating multidisciplinary approaches to govern the service life cycle. Among them, we are developing a methodology called "Services and Meta-Services for smart eGovernment" (SMART in short), whose aim is to consider in an interdisciplinary framework all social, organizational, economic and technological issues that play a role in service design.

One of the paradigms that inspires the methodology, is that, as in other production processes of ICT artifacts such as e.g. database design, two representation levels for services are significant, that we call:

- Abstract services, established by law and having an owner (e.g. for a business, the service "obtain an authorization to open a cafeteria" whose owner are Chambers of Commerce) and
- Concrete services, provided by a specific provider, e.g. the previous service provided by a Chamber of Commerce or by a private provider.

Another relevant concept in SMART is the repository of services, that plays in the service life cycle the role of the software applications repository in software engineering. The repository of services is the integrated representation of all services produced or managed in the provider organization: in this paper we will focus on repositories of abstract services. In order to provide the repository with a semantic structure, highlighting similarities among services, two different semantic relationships are defined in the repository, the is-a and part-of relationships, see Section 3. Repositories of services structured in such a way can be used in the service life cycle as an effective tool for reuse, since service design is not performed from scratch, but is positively influenced and made more efficient by previously developed services.

Repositories of services show similar problems to those of databases and database schemas in data management in the life of an organization. Since databases are produced in time by different designers for different purposes, database schemas representing the same universe of discourse or overlapping universes of discourse show several types of heterogeneities. In order to produce a global schema of the information content of an organization, we need an integration step, that has been deeply investigated in the literature (e.g. [1] and [27]). Also for repositories of services an integration step is needed, to create a reconciled and integrated representation of all services produced and managed in the organization: this is the problem discusses in the paper. Besides reuse, integration of repositories has several interesting applications in the area of service provider companies mergers and acquisitions; the assets of companies are their repositories of services, and a relevant step in the merger/acquisition procedure is the integration of the assets.

This paper is organized as follows. We first analyze (Section 2) related work in the literature on repositories and integration methodologies. In Section 3 we provide basic definitions. Section 4 discusses the service repository integration methodology, and applies it to a real life case study. Section 5 discusses conclusions and future developments.

2. RELATED WORK

Information systems integration [15][5][10] represents a critical and strategic challenge for business, when dealing with complex organizational initiatives such as, e.g., mergers and acquisitions [9][11][17]. Likewise, the integration of services, database schemas, and software components is a key issue in all information systems where several levels of cooperation have to be established between different organizations or players [13][29][14]. Concerning schema integration [1][2], the research moved from the analysis of conflicts and heterogeneities in data bases [27][28] to a focus on semantic integration through ontologies [16], due also to an increasing interest in web services raised by Service Oriented Computing [23][24]. In this context, it is worth noting the renewed relevance and diffusion of repository technologies and registries of services such as the web services standard UDDI¹. Accordingly, repositories of web services are a central subject of research on service semantics and Ontological Matching [26].

¹ http://uddi.xml.org





The concept of repository has been investigated in the area of conceptual schemas. In [5] the repository is built through integration and abstraction primitives applied to schemas. Experiences of usage of schema integration models and large-scale repositories of conceptual schemas have been carried out in particular in the Public Administration domain [32][6][4].

Repositories of ontologies are proposed in several papers, in the area of Semantic Web [20] and in the Web Services area [18]. The alignment and integration of ontologies is investigated e.g. in [30], where information integration is enabled by adopting a standard common terminology.

Integrated service modeling is claimed in [31] to be the most critical and effective goal in eGovernment projects to achieve the "one-stop-shop" approach in interactions among Public Administration and citizens. The role of a service repository in platforms for one-stop-shop of e-Government services is discussed in [31]. An approach based on a semantic repository to support the interactions between service providers and customers in the Government to Business (G2B) context is discussed in [22], enabling semantic navigation and providing metrics and an algorithm to support businesses in selecting the most valuable services from their perspective.

3. BASIC DEFINITONS

We first recall a definition of service adapted from [8] and integrated with the perspective proposed in [19]:

Definition 1 –A *service* s consists in an activity or series of activities, of more or less intangible nature, that they have place in an exchange between a supplier and a customer, where the object of the transaction is an intangible good, so that both the supplier and the customer co-create and obtain value from the transaction.

A service can be distinguished along several classifications, here we introduce the following:

Definition 2 – An *elementary service* (es) is a service that is atomic, namely not made or else not perceived as a set of other simpler services.

Definition 3 – A composite service (cs), is a service modeled in terms of

- a set of 0 to n elementary services [Ø, es₁, es₂, ..., es_n], where Ø is the void service;
- a set of 0 to m composite services [Ø, cs₁, cs₂, ..., cs_m];
- an orchestration function (see [25]).

As an example, we consider the production of a certificate for a change of address as an elementary service (named "Certificate of residency"), since we are not interested to further represent it in terms of atomic components such as, say, the acquisition of the ID of the citizen, an access to the address data base, and the print of the certificate. Furthermore, an example of composite service is the (management of a) change of address of a single citizen from one municipality to a different municipality ("Change of Address (single)"); indeed, if we assume that addresses are represented in data bases managed in municipalities, the service is composed of at least three components that we may consider elementary services:

- a) the cancellation of the citizen address and record from the data base of the source municipality,
- b) the insertion of the citizen address and record in the data base of the destination municipality, and, in case
- c) the production of the residency certificate.

Besides this basic classification, we now introduce two different relationships, *part-of* and *is-a*, that provide a conceptual structure to services in the repository (see Figure 1). These relationships are inherited from conceptual models adopted in database and information systems design [21] and they are defined as follows:

Definition 4 – A part-of relationship holds between a service s_1 (the part) and a service s_2 (the whole) when the production of service s_1 contributes to the production of service s_2 .





ISSN 2278-5612



Definition 5 – An *is-a relationship* holds between two services s_1 and s_2 , called *child service* and *parent service*, when s_1 inherits all properties of s_2 .

Besides part-of and is-a, a service is characterized by other two types of properties, called in the literature functional and non-functional properties (see [24]); we will not consider explicitly such properties in the integration methodology.

We need also to adopt a graphic convention for representing small fragments of repositories, made of is-a and part-of structures. See in Fig.2 the representation of the change of address service and two groups of services that are related with it through the is-a and part-of hierarchies.



Figure 2: A graphical representation of a fragment of the repository

4. A METHODOLOGY FOR SERVICE REPOSITORY INTEGRATION

Figure 3 shows an input output representation of an activity of repository integration, where Is-a and Part-of in cells represent the is-a and part-of relationships among the services represented in rows and columns.



Figure 3: Input output representation of a repository integration process

In its more general form, a methodology for service repository integration is organized in three steps:

- 1. Correspondence analysis;
- 2. Conflict analysis;
- 3. Repository integration;

We now describe the three steps, applying the methodology to a real life case we developed within two domains, the Smart-City and the Tourism domains. The main goal of SMART is to design services that add value both to citizens (in this case the value in-use of the service is relevant), and to providers (value in exchange). The main heuristic adopted to add





value to services is to build from elementary services composite ones; in such composition, bundles of services offered by mixed public and private providers are created. Initially the value-oriented analysis of services in the two domains has been performed by two distinct groups of analysts, which produced two repositories characterized by the following structure.

- The Smart-City (SC) repository is made of 63 services, among them 58 are elementary services, and 5 are composite. 15 services are in Is-a relationships, and 29 are in Part-of relationships. As to providers, in 33 elementary services the provider is a Municipality, in the other 30 it is a private provider (Museum, Accommodation Manager or Local Operator).
- The Tourism (T) repository is made of 93 services, among them 84 are elementary services, and 9 are composite. 37 are in Is-a relationships, and 66 are in Part-of relationships. As to providers, in 29 elementary services the provider is a Municipality, in the other 54 it is a private provider.

We omit the representation of the two repositories for reasons of space.

4.1 Correspondence analysis

In this step pairs of services s(SC) and s(T) of repositories Smart-Cities and Tourism are analyzed in order to identify the presence of correspondences among them. With the term *correspondence* among two services s(SC) and s(T) we mean one of the following semantic relationships:

- 1. *Identity*, when s(SC) and s(T) correspond to the same service.
- 2. Similarity, when s(SC) and s(T) are recognized as different services sharing common properties.
- 3. Part-of, when s(SC) and s(T) are such that s(SC) part-of s(T).
- 4. Is-a, when s(SC) and s(T) are such that s(SC) is-a s(T).

Correspondence analysis consists of two sub-steps that look for potential correspondences among repositories and decide for their presence and type, according to the following procedure:

1. Correspondence search - Decide the strategy of visit of the two repositories.

- 2. Correspondence discovery For each pair of services s(SC) and s(T) encountered in the visit:
 - 2.1 consider their description and properties (namely, names of services, and structure of is-a and part-of relationships in which s(SC) and s(T) are involved).
 - 2.2 decide for the existence of a correspondence, and its type.

Once the correspondences have been discovered, they may result, with reference to the integrated repositories, in two possible types:

- *Conflicts*, when some property of one or both services has to be modified in order to merge them in the integrated schema. Conflicts are considered in Conflict Analysis, see Section 4.2.
- Interschema properties, when the two services can be represented in the integrated schema without applying transformations to them, but new (is-a or part-of) relationships have to be included, that were not present in the two source repositories. Interschema properties are considered in Repository integration, see Section 4.3.

In Figure 4 we represent a set of correspondences and related conflict types. Notation s_i-(Is-a) means the set of services s_j such that s_i Is-a s_j, similar meanings for notations (Is-a)-s_i, s_i-(Part-of), (Part-of)-s_i.

#	Correspondence	Property of services si, i=[1,2]	Type of conflict	
1	Identity	Name	Synonymy	
2	Similarity	Name	Homonymy	
3	Identity	Туре	Different types	
4	Identity	si-(Is-a) relationships	Nonoverlapping is-a relationships	
5	Identity	si-(Is-a) relationships	Partially overlapping is-a relationships	
6	Identity	(Is-a)-si relationships	Nonoverlapping is-a relationships	
7	Identity	(Is-a)-si relationships	Partially overlapping is-a relationships	
8	Identity	si-(Part-of) relationships	Nonoverlapping part-of relationships	
9	Identity	si-(Part-of) relationships	Partially overlapping part-of relationships	
10	Identity	(Part-of)-si relationships	Nonoverlapping part-of relationships	
11	Identity	(Part-of)-si relationships	Partially overlapping part-of relationships	

Figure 4: Correspondences and related conflict types



4.2 Conflict analysis

In this section we highlight examples of the first three types of correspondences in Figure 4, related conflicts and possible solutions that we have encountered in our project and inspired the above approach.

Case 1: Correspondence = Identity; Property = Name; conflict = Synonymy

In Figure 5 we have the case of two services that are considered identical, and have different names (*Hotel reservation* and *Inn reservation*). Here, we have changed the name of one of the two services, and merged them. Another possibility is to change both names.

SC-Repository	T-Repository	Integrated Repository		
Hotel reservation	Inn reservation	Hotel reservation		

Figure 5: Synonymy among services

Case 2: Correspondence = Similarity; Property = Name; Conflict = Homonymy

In this case we have two services that share the same name *City tour on two wheels*. An analysis of the properties of the two services led to the conclusion that they are different, while sharing some properties. So, we have to rename one of the two or both of them. In the case shown in Figure 6, we decided to rename both, and created a new service, to which we assigned the original common name.

City tour on two wheels	City tour on two wheels	City tour on two wheels is-a City tour by bike City tour by tandem

Case 3: Correspondence = Identity; Property = Type; Conflict = Different types

In this case we have two services with the same name *Wine tour*, that in the SC-Repository is a composite service, while it is an elementary service in the T-Repository. An analysis of the properties of the two services led to the conclusion that they are the same service, but in the SC-Repository *Wine tour* is modeled more properly, because a child service of *Wine tour* has been represented named *Reservation of a wine tour*. Accordingly, in the Integrated Repository we maintained the SC-Repository representation (Figure 7).

SC-Repository	T-Repository	Integrated Repository	
Wine tour PART-OF Reservation of a wine tour	Wine tour	Wine tour PART-OF Reservation of a wine tour	







4.3 Repository integration

Besides correspondences that give rise to conflicts, other correspondences result in adding to the pair of repositories new concepts and/or semantic relationships that in our methodology are classified as interschema properties. This is what happens everytime the correspondence is of type Is-a or Part-of; other interschema properties arise from similarity correspondences. In Figure 8 we show a case in which a similarity correspondence is discovered for two services, s6 and s9, that differ only in the presence in Repository B of service s8 part-of s9. In this case we can resolve the correspondence creating an interschema is-a between s9 and s6.

Once the interschema properties are added to the integrated schema, a consistency check has to be made to avoid the creation of cycles in Is-a and Part-of relationships.

We show in the following an example of interschema property in our case study.



Figure 8: An example of similarity correspondence and related interschema property.

<u>Case 4: Correspondence = Similarity; Interschema property: a new service, common parent of a service in SC and a service in T</u>

In this case (see Figure 9) we have services *Hotel reservation* in the SC-Repository and *Bed & breakfast reservation* in the T-Repository. The two services can be related in Is-a with a new service *Accommodation reservation;* common part-of properties between the two services have to be associated with the new service.







The final integrated schema produced at the end of the integration step is made of 153 services, among them 137 are elementary services, and 16 are composite. 57 services are in Is-a relationships, and 105 services are in Part-of relationships (see Figure 10). Notice that the total number of elementary services decreases, due to the presence of several identical services, and the number of composite services increases, due to the existence of part-of interschema properties.

		# of elementary	# of composite	# of services in a	# of services in a PART-OF
Repository	# of services	services	services	IS-A relationship	relationship
Smart Cities	63	58	5	15	29
Tourism	93	84	. 9	37	66
Integrated repository	153	137	16	57	105

5. CONCLUSIONS AND FUTURE WORKS

In this paper we have presented a methodology for service repository integration that was conceived in the course of a project, whose aim was to design services in various application domains that provide value to users and providers. We focused on Smart-Cities and Tourism domains. Although the integration methodology has revealed effective, it needs several improvements. First, we have to generalize the methodology to the case of n repositories. This generalization leads to the investigation of a preliminary step, whose aim is to choose an optimal order of integration among the n repositories, where optimality corresponds to minimizing the whole effort needed in the integration activity. Then, in order to extend the methodology to concrete services, we have to enrich the representation of services including functional and non-functional properties. Further, we have to produce a complete set of correspondences, related conflicts, transformations and interschema properties. Finally we are investigating a tool that helps the designer in discovering correspondences on the basis of structural and semantic properties of services in repositories.

ACKNOWLEDGMENTS

The work presented in this paper has been partially supported by the Italian PON project PON01_00861 SMART (Services and Meta-services for smART eGovernment).

REFERENCES

- 1. Batini, C., Lenzerini, M., A methodology for data schema integration in the Entity Relationship model. IEEE Transactions on Software Engineering (1984).
- 2. Batini, C, Lenzerini, M & Navathe, S.B., 1986. A comparative analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys*, 18(4), pp.323–364.
- 3. Batini, C., Di Battista, G. & Santucci, G., 1993. Structuring primitives for a dictionary of entity relationship data schemas, *IEEE Transactions on Software Engineering*, 19, pp.344–365.
- 4. Batini, C., Comerio, M. & Viscusi, G., 2012. Managing Quality of Large Set of Conceptual Schemas in Public Administration: Methods and Experiences. In *MEDI* Conference, pp. 31–42.
- 5. Bernstein, P.A. & Haas, L.M., 2008. Information integration in the enterprise F. Özcan, ed. Communications of the ACM, 51(9), p.72.
- Bertoletti, M., et al., 2003. The Service to Businesses Project: Improving Government-to-Business Relationships in Italy. In R. Traunmüller, ed. *Electronic government: Second international conference, EGOV 2003, Prague, Czech Republic, September 2003: proceedings*. Berlin / Heidelberg: Springer, pp. 468–471. Available at: http://www.springerlink.com/openurl.asp?genre=article&id=HKH73LRGAQRGE5RD.
- 7. Chesbrough, H., and Spohrer, J., A research manifesto for services science, Communications of the ACM, vol. 49, no. 7, pp. 35–40, 2006.
- 8. Grönroos, C., Service Management and Marketing. Managing the Moments of Truth in Service Competition. Lexington Books, 1990.
- 9. Giacomazzi, F., et al., 1997. Information systems integration in mergers and acquisitions: A normative model. *Information & Management*, 32(6), pp.289–302.
- 10. Hasselbring, W., 2000. Information system integration. Communications of the ACM, 43(6), pp.32-38.



- 11. Hwang, M.I., 2004. Integrating Enterprise Systems in Mergers and Acquisitions. In AMCIS 2004 Proceedings. Association for Information Systems. Available at: http://aisel.aisnet.org/amcis2004/12/.
- 12. Khadka, R., Saeidi, A., Jansen, S., Hage, J., and Helms, R., "An Evaluation Of Service Frameworks For The Management Of Service Ecosystems," in PACIS 2011 Proceedings, 2011.
- Kamal, M.M., Themistocleous, M & Khosrow-Pour, M., 2009. Investigating Enterprise Application Integration Adoption in the Local Government Authorities. In C. G. Reddick, ed. *Handbook of Research on Strategies for Local E-Government Adoption and Implementation: Comparative Studies*. Hershey, PA; London, UK: Information Science Reference, pp. 661–686.
- 14. Khoumbati, K. & Themistocleous, Marinos, 2006. Integrating the IT Infrastructures in Healthcare Organisations: A Proposition of Influential Factors . *Electronic Journal of e-Government*, 4(1), pp.27–36. Available at: http://www.ejeg.com/volume-4/vol4-iss1/v4-i1-art4.htm.
- 15. Lam, W., 2007. Information Systems Integration in E-Government. In A.-V. Anttiroiko & M. Mälkiä, eds. *Encyclopedia of digital government*. Hershey, PA: Idea Group Reference, pp. 1069–1075.
- 16. McGuinness D., L., 1998. Ontological Issues for Knowledge-Enhanced Search. In *Frontiers in Artificial Intelligence and Applications,*. IOS Press, Washington, DC.
- 17. Mehta, M. & Hirschheim, R., 2007. Strategic Alignment In Mergers And Acquisitions: Theorizing IS Integration Decision making. *Journal of the Association for Information Systems*, 8(3), pp.143–174. Available at: http://aisel.aisnet.org/jais/vol8/iss3/8.
- 18. Mika, P. et al., 2004. Foundations for service ontologies: aligning OWL-S to dolce. In S. I. Feldman et al., eds. WWW. ACM, pp. 563–572.
- 19. Normann R., and Ramìrez, R., "From Value Chain to Value Constellation: Designing Interactive Strategy %J Harvard Business Review," 1993.
- 20. Oberle, D et al., 2005. Supporting Application Development in the Semantic Web. ACM Transactions on Internet Technology, 5(2), pp.328–358.
- 21. Olivé, Conceptual Modeling of Information Systems. Berlin- Heidelberg: Springer Verlag, 2007.
- 22. Palmonari, M., Viscusi, G. & Batini, C., 2008. A semantic repository approach to improve the government to business relationship. *Data & Knowledge Engineering*, 65(3), pp.485–511.
- 23. Papazoglou, M.P., Traverso, P., Dustdar, S. and Leymann, F., "Service Oriented Computing: a Research Road map," International Journal of Cooperative Information Systems, vol. 17, no. 02, pp. 223–255, 2008.
- 24. Papazoglou, M. P & Heuvel, W.-J. Van Den, 2006. Service-oriented design and development methodology. International Journal of Web Engineering and Technology, 2(4), p.412.
- 25. Peltz, C., Web services orchestration and choreography. Computer 36 (2003), 46-52
- 26. Syeda-Mahmood, T. et al., 2005. Searching service repositories by combining semantic and ontological matching, ICWS 2005, pp- 13-20.
- 27. Spaccapietra, S., Parent, C., and Dupont, Y., Model Independent Assertions for Integration of Heterogeneous Schemas VLDB Journal, 1, pp. 81-126 (1992).
- 28. Spaccapietra, S., Parent, C., View Integration: A Step Forward in Solving Structural Conflicts. IEEE Trans. Knowl. Data Eng. 6(2): 258-274 (1994).
- 29. Themistocleus, M. & Chen, H., 2004. Investigating the integration of SMEs' information systems: an exploratory case study. *International Journal of Information Technology and Management*, 3(2/3/4), pp.208–234.
- Wang, J. & Gasser, L., 2002. Mutual online concept learning for multiple agents. In AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems. New York, NY, USA: ACM Press, pp. 362–369.
- 31. Wimmer, M. and Tambouris, E., 2002. Online One-Stop Government: A working framework and requirements. In Proceedings of the IFIP World Computer Congress, August 26-30, 2002, Montreal.
- 32. Ye, F., Li, H. & Hu, M., 2011. The Construct of Data Integration Model of Heterogeneous E-Government System Based on Topic Maps. In *Intelligent Computation Technology and Automation (ICICTA), 2011 International Conference on.* pp. 263–266.



ISSN 2278-5612

Author' biography with Photo



Carlo Batini is full professor of Computer Engineering at University of Milano Bicocca, Italy. His research interests include cooperative information systems, information systems, data base modeling and design, data and information quality, web services design, eGovernment planning methodologies, and services and data repositories. He has published more than 30 papers in international journals and books, including IEEE and ACM Transactions, and about 100 papers in international conferences. He is associate editor of the Journal on Data Semantics and associate editor of the Journal

on Data and Information Quality. He published more than 15 books, in particular one on conceptual database design with Benjamin and Cummings (1987) and one on data quality with Springer Verlag (2006). Since 1993 to August 2003 he has been on leave from university, being a member of the executive board of the Italian Authority for Information Technology in Public Administration, where he leaded significant projects related to the Italian e-Government initiative. In 2010 he has co-authored with Gianluigi Viscusi and Massimo Mecella the book Information Systems for eGovernment: a quality of service perspective (Springer, Heidelberg).



Marco Castelli is currently research assistant at the Department of Informatics, Systems and Communication (DISCo) of the University of Milan Bicocca. His research interests cover Information Value and Service Value. In past he played some professional experience including two years in the IT Department of a large company of the manufacturing sector.



Gianluigi Viscusi is currently post-doc research fellow at the Department of Informatics, Systems and Communication (DISCo) of the University of Milan Bicocca: the research title is "Strategic planning and IT alignment of eGovernment systems and services." Research interests concern methodologies for information systems planning, eGovernment, information value, business modelling and IS strategy alignment. He has published more than 40 referred papers in books, conference proceedings, and journals as Data & Knowledge Engineering and Government Information Quarterly. In 2010 he has co-authored with Carlo Batini and Massimo Mecella the book Information Systems for eGovernment: a quality of service perspective (Springer, Heidelberg).