# Semantic Integration of Semi-Structured Distributed Data in the Domain of IT Benchmarking

## *Towards a Domain Specific Ontology*

Matthias Pfaff[1] and Helmut Krcmar[2]

[1]*fortiss GmbH, An-Institut der Technischen Universität München, Guerickestr. 25, 80805 München, Germany*

[2]*Technische Universität München, Boltzmannstr. 3, 85748 Garching, Germany*

Keywords: IT Benchmarking, Distributed Data Sources, Heterogeneous Data, Semantic Data Integration, Ontologies.

Abstract: In the domain of IT benchmarking a variety of data and information are collected. The collection of this heterogeneous data is usually done in the course of specific benchmarks (e.g. focusing on IT service management topics). This collected knowledge needs to be formalized previous to any data integration, in order to ensure interoperability of different and/or distributed data sources. Even though these data are the basis to identify potentials for IT cost reductions or IT service improvements, a semantic data integration is missing. Building on previous research in IT benchmarking we emphasise the importance of further research in data integration methods. Before we describe why the next step of research needs to focus on the semantic integration of data that typically resides in IT benchmarking, the evolution of IT benchmarking is outlined first. In particular, we motivate why an ontology is required for the domain of IT benchmarking.

## 1 INTRODUCTION

Benchmarking as a systematic process for improving organizational performance has gained great popularity worldwide since the 1980s. It is based on the insight that observing organizations and analyzing their acting and (measure) their performance is a powerful way to transform the own organization. This transformation is usually done by applying lessons learned form a benchmark (Camp, 1989; Peters, 1994). Moreover, benchmarking can help explaining value or cost aspects to stakeholders within the company while comparing for example their (IT) unit or only certain services of the IT with competitors (Spendolini, 1992).

Recent research in the Information Systems (IS) (e.g. (Slevin et al., 1991; Smith and McKeen, 1996; Myers et al., 1997; Gacenga et al., 2011)) focuses on the analysis and evaluation of performance measurement. Performance measurement in the IT context requires several prerequisites. Having a well-structured service oriented IT department and a consistent knowledge of IT services and their corresponding costs are, for example, important. Additionally these are basic requirements for circular comparisons and subsequently for improvements based on data analysis. Companies that are interested in benchmarking need to have valid definitions of the value and the costs for the objects selected to benchmark. (Rudolph and Krcmar, 2009) argues, that throughout increasing IT industrialization the standardization, documentation and definition of IT services are gaining more importance. They state, that IT service catalogues are an appropriate instrument to picture such a service structure. In addition, concepts for the identification of critical success factors for measuring the maturity level of service catalogues are developed by (Kütz, 2006) and (Rudolph and Krcmar, 2009). In detail, each IT service (object of IT benchmark) should encompass certain parts of deliverables and infrastructure components (Krcmar, 2010). Many of these studies omit facts such as data quality and data integration. Yet, in spite of this new interest, little work published in IS literature addresses the problem of data integration across different kind of IT benchmarks.

One difficulty in making data of different types of benchmarks comparable with each other is a result from the lack of an uniform description of any parameter that is measured. Moreover, a description of the relation in between two of such parameters is missing. This is not a particular issue in the domain of IT benchmarking. Other fields of research



Pfaff M. and Krcmar H..
Semantic Integration of Semi-Structured Distributed Data in the Domain of IT Benchmarking - Towards a Domain Specific Ontology.
DOI: 10.5220/0004969303200324
In *Proceedings of the 16th International Conference on Enterprise Information Systems* (ICEIS-2014), pages 320-324
ISBN: 978-989-758-027-7

are facing similar challenges in data integration, provided with some promising and practical approaches to solve them (Leser and Naumann, 2007). Thus, research on data integration methods for the specific field of IT benchmarking and its vocabulary should be intensified. Especially given the rising research in big data analysis, results from IT benchmarking should not be discarded because of an inadequate data management. A promising approach for data management lies in the use of a domain specific ontology, in order to make these kind of data meaningful (Uschold and Gruninger, 2004; Horkoff et al., 2012).

The next section gives an overview of benchmarking in general and data integration challenges in the domain of IT benchmarking in specific. Following Section 2 further research areas in semantic integration of IT benchmarking data are presented and discussed in Section 3. Furthermore, a first iterative approach for integrating data from different IT benchmarking initiatives is introduced in Section 3.

## 2 BACKGROUND

Most of the current research in IT benchmarking and the practical literature on this topic is only related to the implementation of IT benchmarks (e.g. (Dattakumar and Jagadeesh, 2003; Jakob et al., 2013)). All of these approaches have one thing in common: Neglecting the need for a sustainable semantic data integration and a unified structure for data management is left out of scope. Thereby most IT benchmarking initiatives are damned to exist side on side in siloed data storages. Consequently, they are incapable to be used a second time or in a different benchmarking context, except they have been collected for.

### 2.1 Benchmarking

In academic research benchmarking can be classified according to the nature of the object of study and according to the benchmarking type (e.g. process benchmarking, product benchmarking, and strategic benchmarking or generic benchmarking) (Carpinetti and Oiko, 2008). Benchmarking partners may include other units of the same organization, competitors in the same or different geographical markets and organizations in related or unrelated industries, in the same or different countries. So, a differentiation is made between internal and external comparisons of such a performance measurement.

Internal performance measurement focuses on the operations of a single company whereas external looks outside the firms industry. Nevertheless, both of them have a common foundation. An overview on the different types of benchmarks is given in Table 1.

Table 1: Types of benchmarks (Carpinetti and Oiko, 2008).

| Type | Description |
|---|---|
| Process Benchmark | Compares operations, work practices or business processes |
| Product Benchmark | Compares products or services |
| Strategic Benchmark | Compares organisational structures, management practices and business strategies. |
| Internal Benchmark | Compares similar products or services of similar business units within one organization |
| Competitive Benchmark | Compares performance with a direct competitor. Objects under investigation can be: Products, services, technology, research and development, personnel policies, etc. |
| Functional Benchmark | Comparisons between one or more non-competitive organizations of particular business functions or processes. |
| Generic Benchmark | Compares an organization or business unit with the best performing organisation, irrespectively conducted of the type of industry. |

An IT benchmark can be considered as passing through several phases. Starting with the initial conception by describing the object to investigate, up to optimizing and re-organizing internal (business) processes (cf. Figure 1). For each of these phases of a benchmark numerous data get collected in various data formats. The substance of these data are qualitative, as well as quantitative statements collected over the complete benchmarking cycle in every single benchmark. Furthermore these data get collected for every single participating company of a benchmark.

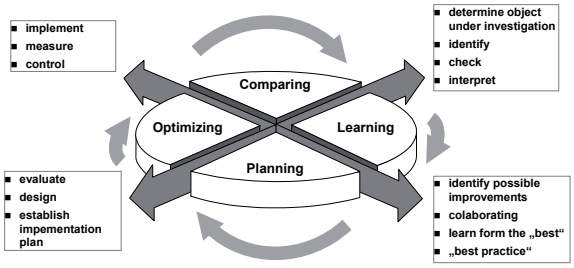


Figure 1: Phases of a benchmark (based on (Watson, 1993)).

## 2.2 Data Integration

As has already been presented by (Ziaie et al., 2012) and structural described by (Riempp et al., 2008) tool based data collecting is quiet common in the domain for IT benchmarking. Even if different benchmark types measure the same object from different perspectives a direct link in between these collected data is difficult to establish.

Next to various formats the data are stored no semantic information are machine readable persisted. But, in order to make the captured data comparable between different benchmarking approaches a semantic integration in a machine readable data format is crucial. Since concepts of such data integration methods are missing, most of the gathered data during a benchmark will stay only applicable for this specific one time performance measurement in its specific domain focus (e.g. cluster benchmarking by (Carpinetti and Oiko, 2008)). In other words, comparability of benchmarking data beyond the specific context of one specific benchmark is left out of research focus and actually impossible because of data separation.

Figure 2 shows the different scopes of data storing in benchmarking. Companies can participate on a specific benchmark (*Benchmark 1..n*) in a specific year. In other words, data storing is done yearly per participant. In addition, a benchmark itself can consist of several services (*Service A..n*) or specific strategic questions. Even if such benchmarks do have the same object of observation (f.i. same service or same product), no direct semantic information of these data are stored. Therefore, this kind of siloed storing information do inhibit further comprehensive analysis.

In the context of data integration particular requirements are demanded from the use of distributed context sensitive (i.e. heterogeneous) data. Since these are usually not solely for one field of research (e.g. IT benchmarking), approaches and methods to organize information are already applied in related fields of research. Ontologies which, by definition convey electronic or ”semantic meaning” are already used to structure unstructured data (e.g. (Cambria et al., 2011)) in the medical or in the information management sector (Riedl et al., 2009; Müller, 2010; Cambria et al., 2011). Thus, representing semantic knowledge with formal ontologies, as proposed by (Guarino, 1995) and (Brewster and O'Hara, 2007), seem to provide promising approaches for data integration techniques in the domain of IT benchmarking.

In the academic literature of ontologies there exist several types of ontology development strategies. (Wache et al., 2001) distinguishes between three main types of ontologies (cf. Figure 3). A *single ontology*

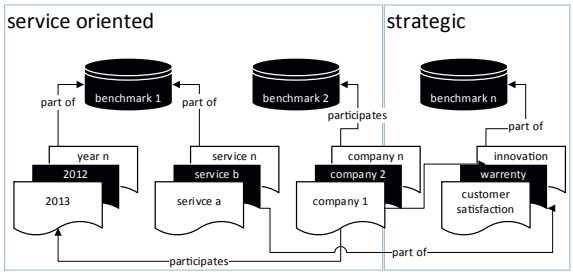


Figure 2: Data dispersion in benchmarking.

(Figure 3(a)) uses a shared vocabulary for describing the semantic information of data . The main advantages of this approach is its quick development process. Managing a single complex and large ontology is one of the main disadvantages, while every change is generating potentially sweeping ontology-wide inconsistencies. *Multiple ontologies* (Figure 3(b)) are based on several independently build ontologies for every source of information. The complexity of a single ontology is only dependant from its corresponding data source and therefore in general less complex. One major disadvantage is the lack of a shared vocabulary when comparing these ontologies. In order to achieve such comparisons *hybrid ontologies* (Figure 3(c)) are used. This kind of ontologies use a shared vocabulary with basic terms of the domain related information of its local ontologies.

On the basis of the existing data of IT benchmarking collected within the last four years, it has to be checked first which type of ontology being the most likely to leverage data integration. Particularly bearing in mind that most of the collected data during an IT benchmark were only meant to be used in their single case of measurement. Thus, existing data form questionnaires presented by (Ebner et al., 2012) and (Ziaie et al., 2012) are used to identify possible start-

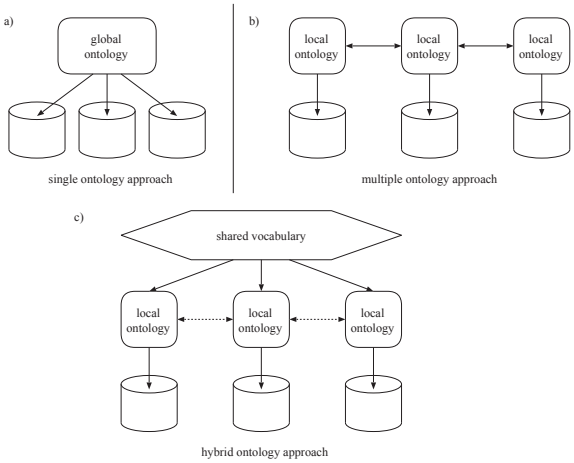


Figure 3: Types of ontologies (Wache et al., 2001).

ing points for a benchmarking ontology.

## 3 CONCLUSIONS

Identifying potential performance improvements within organisations by the use of IT benchmarks suffers from the quality of the collected data. This quality of data is strongly dependent on a precise specification of every single key performance indicator.

There is not only a demand of a precise description of these indicators on the questionnaires side, the underlying contextual connection should be taken into account for data management. This is especially important when trying to analyse benchmarking data beyond the specific scope they were collected for.

In order to achieve a comparison across different kinds of benchmarks a consistent semantic description of the collected data is essential. Consequently, future research on semantic data integration should be conducted for the domain of IT benchmarking.

For the development of a suitable solution for the data integration in IT benchmarking, already available data and service descriptions of different IT benchmarks serve as sources. These data were collected from 25 large and medium size companies during strategic and service oriented IT benchmarks over the last years. Previously implemented online IT benchmarking systems (c.f. (Ziaie et al., 2012)) and frameworks to structure and asses strategic IT/IS management (c.f. (Riempp et al., 2008)) are used for the data acquisition. Building up on these data the specific requirements that need to be met by a concept for data integration are identified.

Using a common vocabulary, such as based on (ITIL, 2013) might ensure broad acceptance of different domains of benchmarking or IT service management. Derived from this, a domain specific ontology for IT benchmarking will be developed iteratively according to (Noy and McGuinness, 2001).

In a next step, a concept of a system to re-integrate and organize benchmarking data needs to be developed and prototypically implemented. To this end, the previously used data and service descriptions of a strategic and service oriented benchmark can be restructured according to the previous elaborated ontology. This in turn allows a direct inclusion of the ontology and the restructured data into the existing capturing mechanisms for the data collection process during an IT benchmark. Therewith, not only an ontology for IT benchmarking is elaborated but also the seamlessly fit into the existing benchmarking tools is pointed out, with all its added value in terms of comparability of data collected.

Moreover, already existing benchmarking data become significantly enhanced by establishing a link across boards of different benchmarking initiatives.

At least the collected data become comparable and integrable across different benchmarking domains. This enables the development of new assistance system and further statistical analysis on such structured IT benchmarking data.

In addition, already existing data sets can be integrated into a uniform data representation structure and thus be used for further statistical analysis which is actually not possible.

## REFERENCES


Brewster, C. and O'Hara, K. (2007). Knowledge representation with ontologies: Present challenges - future possibilities. *International Journal of Human-Computer Studies*, 65(7):563–568.

Cambria, E., Hussain, A., and Eckl, C. (2011). Bridging the gap between structured and unstructured health-care data through semantics and sentics. In *Proceedings of ACM*.

Camp, R. (1989). *Benchmarking: The search for industry best practices that lead to superior performance*. Quality Press, Milwaukee, Wis.

Carpinetti, L. and Oiko, T. (2008). Development and application of a benchmarking information system in clusters of smes. *Benchmarking: An International Journal*, 15(3):292–306.

Dattakumar, R. and Jagadeesh, R. (2003). A review of literature on benchmarking. *Benchmarking: An International Journal*, 10(3):176–209.

Ebner, K., Riempp, G., Müller, B., Urbach, N., and Krcmar, H. (2012). Making strategic it/is management comparable: Designing an instrument for strategic it/is benchmarking. In *Proceedings of the Multikonferenz der Wirtschaftsinformatik (MKWI 2012)*, Braunschweig, Germany.

Gacenga, F., Cater-Steel, A., Tan, W., and Toleman, M. (2011). It service management: towards a contingency theory of performance measurement. In *International Conference on Information Systems*, pages 1–18.

Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, 43(5-6):625–640.

Horkoff, J., Borgida, A., Mylopoulos, J., Barone, D., Jiang, L., Yu, E., and Amyot, D. (2012). *Making Data Meaningful: The Business Intelligence Model and Its Formal Semantics in Description Logics*, volume 7566 of *Lecture Notes in Computer Science*, book section 17, pages 700–717. Springer Berlin Heidelberg.

Jakob, M., Pfaff, M., and Reidt, A. (2013). A literature review of research on IT benchmarking. In *11th Workshop on Information Systems and Service Sciences*, volume 25.

Krcmar, H. (2010). *Informationsmanagement*. Springer, Berlin, 5 edition.

Kütz, M. (2006). *IT-Steuerung mit Kennzahlensystemen*. dpunkt.verlag, Heidelberg.

Leser, U. and Naumann, F. (2007). *Informationsintegration: Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen*. dpunkt-verlag, Heidelberg.

ITIL (2013). The official ITIL website. http://www.itil-officialsite.com.

Müller, M. (2010). *Fusion of Spatial Information Models with Formal Ontologies in the Medical Domain*. Thesis.

Myers, B. L., Kappelman, L. A., and Prybutok, V. R. (1997). A comprehensive model for assessing the quality and productivity of the information systems function: Toward a contingency theory for information systems assessment. *Information Resources Management Journal*, 10(1):6–25.

Noy, N. and McGuinness, D. (2001). Ontology development 101: A guide to creating your first ontology. Technical report, Stanford knowledge systems laboratory and Stanford medical informatics.

Peters, G. (1994). *Benchmarking Customer Service*. Financial Times Management Series. McGraw-Hill, London.

Riedl, C., May, N., Finzen, J., Stathel, S., Kaufman, V., and Krcmar, H. (2009). An idea ontology for innovation management. *International Journal on Semantic Web and Information Systems*, 5(4):1–18.

Riempp, G., Müller, B., and Ahlemann, F. (2008). Towards a framework to structure and assess strategic IT/IS management. *European Conference on Information Systems*, pages 2484–2495.

Rudolph, S. and Krcmar, H. (2009). Maturity model for it service catalogues an approach to assess the quality of IT service documentation.

Slevin, D. P., Stieman, P. A., and Boone, L. W. (1991). Critical success factor analysis for information systems performance measurement and enhancement: a case study in the university environment. *Information & management*, 21:161–174.

Smith, H. A. and McKeen, J. D. (1996). Measuring IS: How does your organization rate? *ACM SIGMIS Database*, 1996(1):18–30.

Spendolini, M. J. (1992). *The benchmarking book*. Amacom New York, NY.

Uschold, M. and Gruninger, M. (2004). Ontologies and semantics for seamless connectivity. *SIGMOD Record*, 33(4).

Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., and Hübner, S. (2001). Ontology-based integration of information - a survey of existing approaches. In Stuckenschmidt, H., editor, *IJCAI–01 Workshop: Ontologies and Information Sharing*, pages 108–117.

Watson, G. (1993). *Strategic benchmarking: how to rate your company's performance against the world's best*. J. Wiley and Sons, New York.

Ziaie, P., Ziller, M., Wollersheim, J., and Krcmar, J. (2012). Introducing a generic concept for an online IT-Benchmarking System. *International Journal of Computer Information Systems and Industrial Management Applications*, 5.