# The Framework of SOA-GTDS

Yan Li and Christoph Meinel

Hasso-Plattner-Institute, University of Potsdam, Germany

*Abstract* — **The increasing number of cancer patients has raised more and more pressure on health care for nations. In order to improve caring quality of cancer patients, a comprehensive, interactive and flexible document management system on various cancer patients and their diagnosis requires more and more attentions. Germany has made huge progresses on tumor documentation during a long time. However, due to the lack of a feasible architecture, most tumor documentation systems geographically distributed in the nation cannot be fully accessed via Web service and their information cannot be inter-exchanged between each other. Therefore, it is necessary to build a unified nationwide architecture on managing the documents on the cancer registry, posterior diagnosis and quality controls. GDTS is a hospital cancer documentation system and used by more than 60 hospitals, clinics and tumor centers. This paper proposes a SOA-GTDS system, which supplies the fast access through Web service, the timely documentation exchange through homo- or heterogeneous tumor system and the easily analysis on quality controls.**

*Index Terms*— **GTDS, SOA-GTDS, Tumor, Web Service, Quality Control**

## I. Introduction

It is estimated that 1.45 million cancer patients are living in Germany, and their first cancer diagnosis were registered over 5 years ago. This number has increased since 1990 by 40% for women and 90% for men, which raised more pressure on the health care for the nation [1].

In order to improve caring quality for cancer patients, the comprehensive description of each cancer registry and diagnosis event is documented. Such long-term documentation collecting enables a trustable quality statement for the cancer treatment.

Germany has made more progresses on tumor documentation in a long time, and various documentation systems have been applied by hospital or tumor centers. German National Cancer Registry initiated the law on building a regional cancer registries network from 1995 to 1999, which legally guarantees a regulation on the data transfer from the registries in this network to a central institution for the further documentation process [2]. So, with the help of such network, there are nearly 426.800 total cancer cases in Germany in 2006, which is slightly lower than the 2004. However, this change does not mean the decrease of cancer patients, but due to the removals of duplicated documentation entries for the same patients collected by different registries in this network [3]. Unfortunately, this network does not have quality cancer registries [2]. The reason is that different states have different data protection laws on cancer registries. So the collected documentations by tumor registries aren't completed, and reliable. Therefore, it is required to build a flexible, scalable and secure architecture for a nationwide cancer registry system, which transparently provides the complete and reliable documentation. [2]

In this paper, we propose the framework of SOA-GTDS. GTDS (Giessner Tumordocumentationsystem) [4] is a hospital cancer register system, and was required since 1991 by the Federal Ministry of Health of Germany. Its goal is to develop a tool for hospital tumor registry and follow-up diagnosis of cancer patients. GTDS has the following two main features, through which we can benefit to implement a nationwide cancer registry system.

1. *Widely distributed hospital cancer registries*

GTDS is the mostly used hospital cancer registry system in Germany. It is used by over 60 hospitals, clinics and tumor centers, which are displayed in Figure 1.

2. *Full functionalities*

As a kind of hospital cancer registries the system, GTDS contains all the necessary functionalities.

Manuscript received December 13, 2010.

Y. Li. is a PhD student in Hassoa-Plattner-Institute at Potsdam University, Germany (e-mail: yan.li @hpi.uni-potsdam.de)

C. Meinel is the CEO of the Hasso-Plattner-Institute for IT-Systems Engineering and full professor for computer science of Potsdam University, Germany (e-mail: christoph.meinel@hpi.uni-potsdam.de)

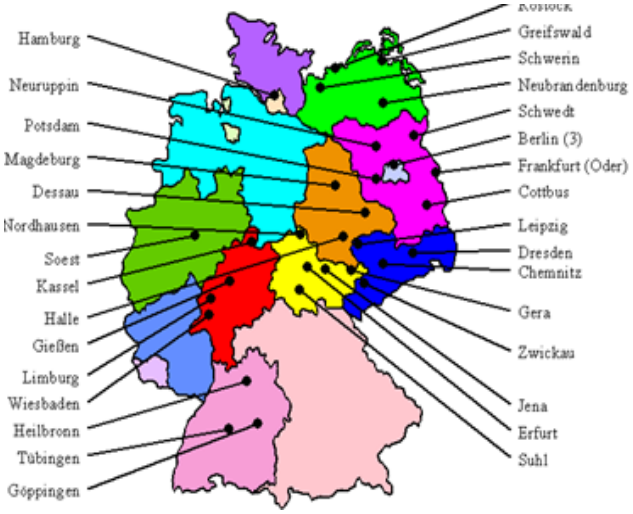


Figure 1 Installation of GTDS in Germany

**However, because of the incompleteness of the software architecture, the current GTDS system stays as a standalone tumor documentation system, which has a low scalability and poor flexibility, and lacks of the opportunity to apply the fast access and reliable data exchange through the Web service.**

Based on two main features and the existing problem shown above, this paper proposes a SOA framework solution for the current GTDS system. SOA, as a software architecture mode, is used for a cross- enterprise system. Based on the benefit of SOA-Architecture, the SOA-GTDS gives the opportunity to supply the fast access through Web service, the timely documentation exchange through homo- or heterogeneous tumor system and the easily analysis on quality controls**.**

This Paper is structured in the following sections: the **Section 2** shows the related documentation used in GTDS; the **Section 3** presents in details the design of SOA-GTDS system. The **Section 4** discusses four different cases studies of SOA-GTDS, and the conclusion and future work is given in **Section 5**.

## II. Related Work: GTDS and SOA

In this part, we will present two related works: GTDS and SOA. The documentation creation and the related data collection used by GTDS will be firstly stepwise discussed, and then we will describe the key concept of SOA.

### *A. Documentation Integration of GTDS*

#### *A.1. Goal of documentation collection*

In the hospital cancer registries, the goal of documentation collection is to get completed and accuracy data from different data sources to improve quality of cancer registries. All these collected documentations record the patients over a long period from his/her first diagnosis to the death. The collected documentations are integrated for the quality analysis of the medical treatment and the follow-up diagnosis and caring.

#### *A.2. Classification of documentation collection*

Hospital cancer registries include two types of documentation collections. One is process-oriented documentation collection, and the other is system-based documentation collection. A process-oriented documentation collection is collecting the documentations of a patient along his/her diagnosis and caring process; while a system-based documentation collection produces the new documentations by calculating other process-oriented documentations collected by different tumor and non-tumor registries.

***Process-oriented documentation collection***

Process-oriented documentation collection collects the data through reporting forms, which are defined by the tumor diagnosis process. We use three features to describe the collection process, which are showed in Table 1.

TABLE 1
FEATURES OF THE PROCESS-ORIENTED DOCUMENTATION COLLECTION

| Feature | Denotation |
|---|---|
| Location | Tumor center, clinic; private medical praxis |
| Role | Colleagues of tumor center, Test-doctor of tumor center, Doctor of clinic or private medical praxis |
| Content | Diagnosis and caring data |

The content of a process-oriented documentation is produced by the following 7 steps:

***Step 1:*** *A Doctor fills first reporting/ diagnostic reporting.*

1) A patient has made a check-up with his family doctor because of painless swelling of lymph nodes;
2) The diagnosis from his family doctor points to Hodgkin's disease;
3) The doctor fills first reporting/ diagnostic reporting;
4) The doctor sends the reporting to the nearby tumor center.

***Step 2*****:** *Tumor center registries the reporting to the GTDS.*

1) Correctness of reporting/ diagnostic reporting was checked by a colleague of the tumor center and test-doctor
2) If the GTDS hasn't the information of the patient, the colleague enters the reporting to the GTDS.
3) If the patient already exists in GTDS, the *colleague* checks, whether the master data of the patient in the GTDS is still valid or whether the patient in the meantime, for example, is moved.

***Step 3:*** *A Doctor fills treatment reporting.*

1) The doctor fills treatment reporting of the patient;
2) The doctor sends the reporting to the tumor center.

***Step 4:*** *Tumor center registries the treatment reporting to the GTDS.*

1) Correctness of treatment reporting was checked by a colleague of the cancer center and test-doctor;
2) The colleague enters the reporting to the GTDS.

***Step 5:*** *the GTDS generates a scheduled follow-up reporting.* After completion of primary treatment, GTDS provides monthly a scheduled follow-up reporting for a patient.

***Step 6:*** *A Doctor of the patient executes the scheduled follow-up reporting*.

***Step 7:*** *Ending reporting is performed.*

1) The patient is died in a clinic;
2) A doctor of the clinic fills an ending reporting;
3) The doctor sends the reporting to the nearby tumor center;
4) The tumor center registries the information of ending reporting in the GTDS.

***System-based documentation collection***

Besides collecting the documentations through the reporter following the defined process, GTDS generates documentations automatically by calculating the process-based documentations collected by various hospital information systems and tumor registry systems.

*1) Hospital information system*

HIS is an import data source of a hospital cancer registry system, which provides not only tumor documentations, but other information. GTDS can process this information and generate new documentations such as diagnosis and treatments to improve the complement of data from one hospital system.

*2) Other Tumor registry system*

Moreover, GTDS can import, process the documentations of related patients from other tumor registry systems and produce new documentations.

*B. SOA – Service Oriented Architecture*

SOA is a an architectural style, whose purpose is to address the requirements of loosely coupled, standards-based, and protocol-independent distributed computing, mapping enterprise information systems (EIS) appropriately to the overall business process flow[5].

We restructure GTDS by introducing SOA due to the two main features of SOA required by GTDS: interoperability and reusability.

**Interoperability**

SOA provides GTDS the interoperability of documentation integration between homogenous tumor registry systems, which was the challenge for providing a flexible, scalable and secure software structure for a nationwide cancer registry system as well as GTDS. Over SOA, GTDS benefits the interoperability for documentation integration via standards-based adapters and interfaces.

**Reusability**

GTDS needs the ability to be easily extended to a nationwide cancer registry system. Therefore the medical logic of the system needs to be reused as much as possible during this extension process. SOA provides the reusability through a fundamental concept service. Service defined as a discrete unit of business functionality that is made available through a service contract [6]. The service contract specifies all the interactions between the service consumers and service providers. Different services can be reused and accessed through standard interfaces.

## III. STRUCTURE OF THE SOA- GTDS SYSTEM

In this section, we will discuss the new SOA based architecture designed for GTDS System. Figure 2 shows the abstraction of a layer-architecture. Fiver layers are precisely defined in this new SOA architecture, which contains a set of service components, as well as the interactions among components and layers.

*A. Operation system layer*

This layer represents all custom and packaged application assets in the application portfolio running in an IT operating environment, supporting business activities [7]. To reduce the cost of implementing SOA-GTDS, the GTDS database system is maintained as a legacy system in this layer. Besides, this layer includes service applications communicating with other systems such as epidemiological cancer registry, other GTDS Systems and Hospital information system, which supply the collaborations for other medical logic with GTDS.

*B. Component layer*

This layer contains all software components, each of which implements a service or a part of a service. A “legacy system connection” component is used to contain not only the connection of the fundamental business database, but also the reused functionalities extracted from the current GTDS. The “interoperability” component implements collaborative information transformation with other systems.

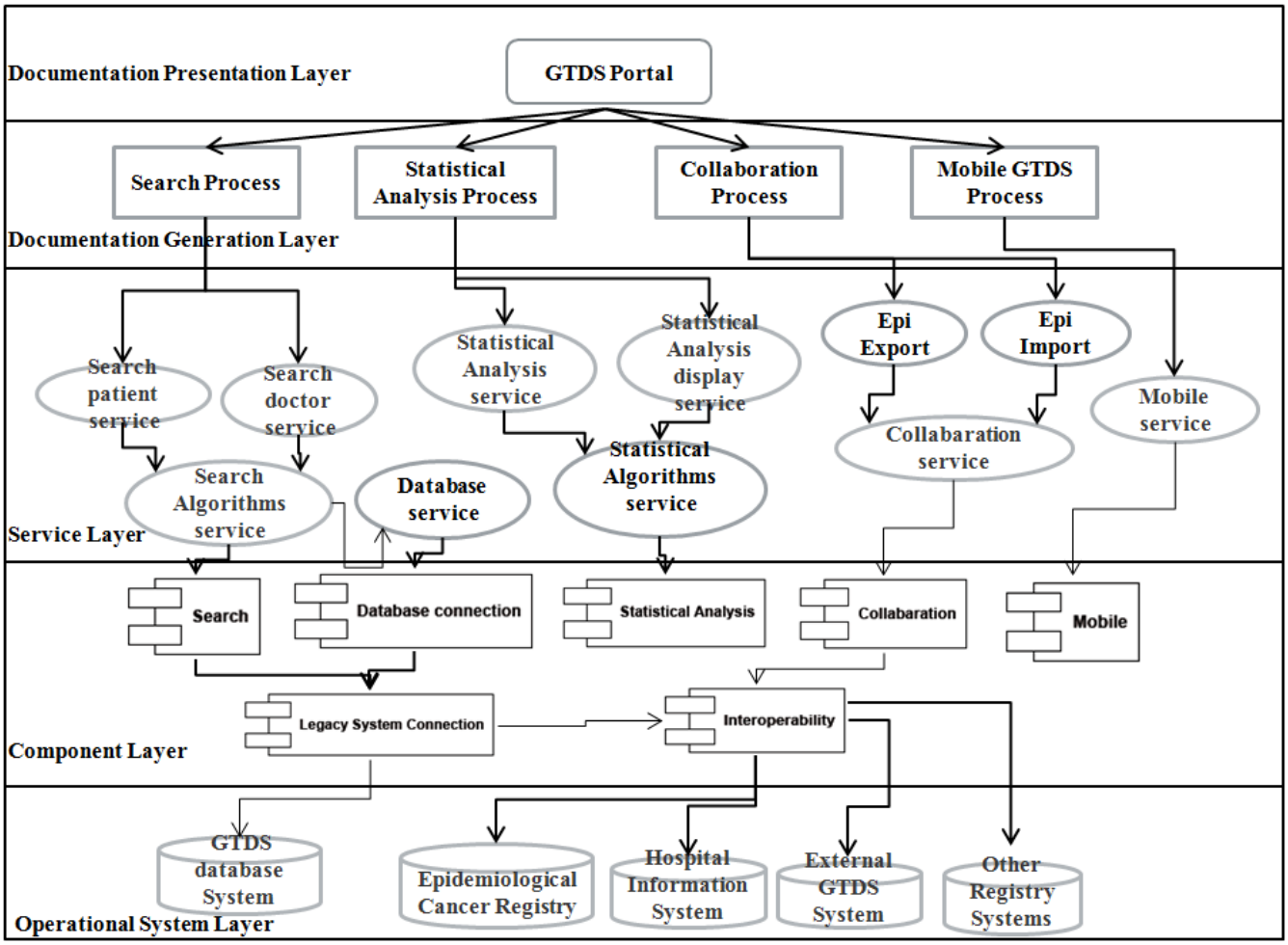


Figure 2 The architecture of the SOA-GTDS

*C. Service layer*

This layer contains all the services defined within the SOA-GTDS. Services can be reused to build other services. For example, "search patient service" and "search doctor service" reuse "search algorithm service".

*D. Documentation generation layer*

This layer integrates services to tumor documentation managing processes of SOA-GTDS.

*E. Documentation Presentation layer*

The layer consumes services provided by the documentation generation layer. The layer implemented the final web GTDS portal, through which the users can enjoy the fast, reliable and comprehensive documentation services.

## I. Use Cases Studies

*A. Use Case 1: Statistical analysis*

This use case realizes the functionally of supplying data for two kinds of analysis tasks: *descriptive* and *analytical statistical analysis*. *Descriptive statistical analysis* describes the populations of patients can be evaluated by descriptions according to different characteristics of patients such as age, gender, tumor stage, treatment, etc. *Analytical statistical analysis* gives some evaluations in a quantified way such as the comparisons of the survival time of patients [8]. Having these two analysis functionality is the fundamental goal of building a tumor documentation system.

*A.1. One scene for the use case 1*

In a regional tumor center, a quality report is published every year for the quality assurance of cancer care. For example, a type of report "Qualitätssicherung durch klinische Krebsregister -- Qualitätbericht Onkologie" is published by the tumor center Brandenburg in 2009. This report describes the result of the quality assurance by calculating frequency, severity and differences in the follow up for patient between hospitals and administrative districts. This analysis process can be implemented in a GTDS System by *Analytical statistical analysis* which compares the statistical analyzed anonym tumor registry data. Figure 3 shows an example of such *Analytical* results with the graphic images [9]. This Figure presents an instance of analyze image, that numbers of cancer incidence per annum corresponding age group are illustrated. In SOA-GTDS, such graphically analytical results can be directly produced and displayed through a web-portal.

*A.2. Complexities for Implementing this Use Case*

***Restructuring data***

Statistical analysis is based on tumor-based data set, in which the documentations are anonyms, but the old GTDS only accept the collected documentations having detailed patient personal information, which is discussed in Section 2. In a single GTDS, the documents are indexed and recognized based on unique patient information, which means the documentations are patient-center organized. Such data organization cannot support the quality analysis on tumor system level. So in SOA-GTDS, the documentations are restructured by establishing an exchange-process module. The key concept of this module is the exchange-process from patient-based data set to tumor-based data set.

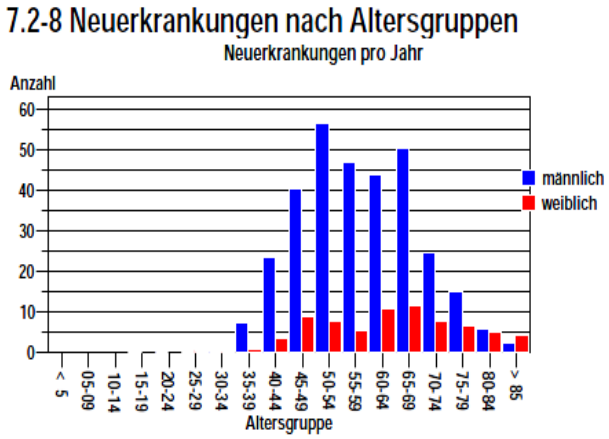


Figure 3 An example of an analytical result

***Implementation of Statistical analysis with R***

The old GTDS provides only a standard interface for the SPSS based data set, which does not have the statistical functionality. However, in order to reduce cost of analyze process, the statistical analyze in the new GTDS will be through R language based module realized. R is a language and environment for statistical computing and graphics [10].

In SOA-GTDS, such functionalities implemented by statistic components though different layers

*B. Use Case 2: Transformation of Collaborative Information*

Transformation of collaborative information is the significant requirement for GTDS system. Improving the quality of the cancer data needs the collaborations of complemented cancer information. The completeness of cancer information is not only required by GTDS, but also by other connected systems with GTDS. For example, the Figure 4 presents the four kinds of collaborative medical systems. On the one hand, GTDS can accept data from another epidemiological cancer registry system; on the other hand, this epidemiological cancer registry requires the provided cancer reporting from GTDS. It is found that 95% of the tumor reporting of epidemiological cancer registry will be provided by the hospital cancer registries [11]. It is necessary that we design a mechanism for transforming collaborative information among different tumor systems.

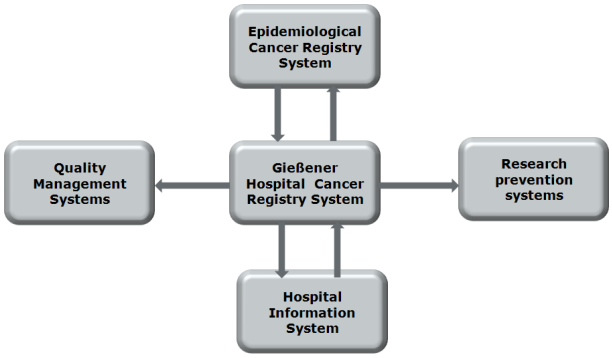


Figure 4 The four kinds of collaborative medical systems with GTDS

*B.1. One scene for the use case 2*

In this part we discuss a scenario for transforming collaborative information between GTDS and an epidemiological cancer registry (ECR) system. GTDS requires importing death certificate information of patients from ECR to supplement the key life status information in GTDS.

The GTDS exports the following datasets to ECR [12].

1. Recording of diagnosis within the reporting period
2. Newly added primary therapy data
3. Modification of the crucial tumor-date
4. Date of death within the reporting period
5. Five years previous date of diagnosis Tumor.

On the other hand, ECR requires import death certificate from GTDS. GTDS provides death certificate as workable format to ECR.

*B.2. Complexities for Implementation of the Use-Case*

***Data transmission between different data formats***

GTDS and ECR use difference data formats for registry. We require using medical transaction standards and formats (HL7) for exchanging documentations between two systems.

***Completeness and consistency of data sources***

If the GTDS accepts the data of the dead patients from an epidemiologic cancer registry, we need to ensure completeness and consistency of the new accepted data through checking rules.

***Timely data communication***

Real-time data communication of reporting improves data quality of registry and efficient.

This use case can be implemented through the interoperability module of the SOA-based structure, which is shown in Figure 2.

*C. Use Case 3: Integrated Query strategic*

Integrated Query strategic is an essential function in GTDS which requires the necessary extension for querying the related tumor information efficiently. However, the current

GTDS only gives solid SQL query on the information of dedicated patients and doctors, and applies the limited spelling correction. A structured SQL query is shown in the current GTDS, which means that the user has to clearly know what query fields should be used for different query terms. For example, if a user wants to query a patient named "Ms. Alice Schmidt", he has to input "Alice" in querying the first name field and "Schmidt" in querying the second name field. If a user wants a more complex query such as one "Alice Schmidt" born on December 1980 in Berlin, he has to input the related query term in the right query fields. Moreover, the current GTDS does not supply the functionality querying the diagnosis documentations with the combinations on patients, doctors and diagnosis terms.

*C.1. One scene for the use case 3*

For the UI consistence of query, SOA-GTDS keeps the old structure query mechanism. However, SOA-GTDS supplies a more powerful query mechanism. A user can query "Ms. Alice Schmidt December 1980 Berlin Dr. Müller" in one query box to query the tumor diagnosis documentations for "Ms. Alice Schmidt" who was born (or dead) on December 1980 and related with "Dr. Müller".

Moreover, SOA architecture gives the possibility to apply the same query on distributed GTDS and other tumor systems. For the same query, it is impossible for the current GTDS to retrieve the documentations on other tumor systems, but SOA-GTDS will automatically search on other tumor systems by using query web service for the same query.

Beside, SOA-GTDS plans to give a more intelligent spelling correction. Because SOA gives the functionality to executing the same query on distributed GTDS and other tumor systems, it brings the possibility on selecting the more accurate documentations on misspelled queries.

*C.2 Complexities for Implementation of the Use-Case*

There are some interesting techniques required by the powerful query mechanism in SOA-GTDS. For example, one box query, which replaces the solid query on multi boxes, requires the efficient query parser and analysis to send the sub queries to different query fields. Different from the free text query on Internet, most documentation in GTDS are structured and semi-structured. With the structural features and the relatively closed and limited domain on tumor information, it is possible to develop a light but powerful query parser.

Fuzzy string searching illustrates a type of searching algorithms, called "approximates matching of strings", which calculates edit distance between two strings. This type of Algorithms search string und allows errors [13]. This searching method can improve fault tolerance and efficiency of system query.

This use case is implemented in search module on component, service and documentation generation layers, which converts the relatively free input query to the classical SQL database query.

*D. Use Case 4: 4G-Mobile GTDS*

With the increasing demand of users for tumor registry systems, traditional data collection methods, for example paper-based form or desk computer, are not suitable for documentation collections used by GTDS. Mobile phone improves the efficiency and effectiveness of data collection in resource-poor environments. SOA-GTDS, supplies the channel for the mobile users to upload and retrieve tumor documentation via mobile web service.

*D.1. One scene for the use case 4*

In Section 2 we described collection process-based data. Let's imagine a patient treatment happening emergently on an airplane, the doctor needs quickly entering and reporting diagnosis information to a tumor center. It is necessary, that the doctor use 4G-mobile phone to enter cancer reporting of a patient.

*D.2 Complexities for Implementation of the Use-Case*

***User Interface(UI)- Design***

Two critical success factors of mobile services are mobile user experience–Design and UI-Design. UE describes all aspects of a user's experience when interacting with a product, service, environment or facility. In a cancer registry system large quantity information will be collaborated with users by a mobile phone GUI. During the process, the information selection, process and displaying is totally different from the traditional desktop service. Therefore, the efficiency of displaying and saving information and information selection are key problems. Due to the highly reusability and Interoperability supplied by SOA, it requires low expense on extending the GTDS to mobile service. However, extra efforts have to be paid on improving the quality of mobile user experiences.

This use case is implemented through the mobile-GTDS module in SOA-architecture.

## II. Conclusion and Future Work

In this paper, we have discussed the framework of the SOA-GTDS system, which provides a flexible and scalable SOA-based software structure for tumor documentations. As a result, tumor documentations between homo- or heterogeneous tumor systems can be fast and timely exchanged and easily analyzed on quality controls. Moreover, a unified nationwide architecture on managing the documents on the cancer registries, posterior diagnosis and quality controls will be possibly realized by means of the SOA-GTDS system. Eventually, this unified nationwide architecture can apply high caring quality of cancer patients for nations.

Currently, more detailed structure of SOA-GTDS, which is based on functionalities on different layers, are being implemented. Because the structure of SOA-GTDS can be extended, new functional components are easily generated by dividing the defined ones in the layer-architecture shown in Section 3. However, GTDS is a huge and complex tumor

documentation management system. We have to fully investigate, decouple and enrich the functional requirements in the old GTDS, e.g. management of tumor therapy has to be analyzed as a new use case. These are the tasks that we have to do in the near future.

## REFERENCES


[1] Umfassende Darstellung des Krebsgeschehens in Deutschland, http://www.rki.de/cln_178/nn_205760/DE/Content/Service/Presse/Pressemitteilungen/2010/03__2010.html

[2] Karen Dente: Cancer meeting in Germany highlights need for national registry. Nature Medicine 15, 825 (2009) doi:10.1038/nm0809-825

[3] Krebs in Deutschland http://www.rki.de/cln_160/nn_204124/DE/Content/GBE/DachdokKrebs/KID/gesamt__bKindern/C00__97__07,templateId=raw,property=publicationFile.pdf/C00_97_07.pdf

[4] Gießener Tumordokumentationssystem: http://www.med.uni-giessen.de/akkk/gtds/

[5] M. P. Papazoglou , W-J van den Heuvel : Service oriented architectures: approaches, technologies and research issues

[6] M. Rosen, B. Lublinsky, Kevin T. Smith Marc, J. Balcer: Applied SOA-Service-Oriented Architecture and Design Strategies

[7] Ali Arsanjani, Liang-Jie Zhang, Abdul Allam, Michael Ellis, et al., "Design an SOA solution using a reference architecture", IBM developerWorks, http://www.ibm.com/developerworks/library/ar-archtemp/

[8] J. Dudeck, E. Grundmann, G Wagner: Basisdokumentation für Tumorkranke. Springer, Berlin; Auflage 5

[9] Qualitätsbericht Onkologie 2009 http://www.tumorzentrum-brandenburg.de/pwp/uploads/Sachbericht_2009_040310.pdf

[10] R language: http://www.r-project.org/

[11] K. Strelocke, O. Förster, A. Kluttig, C. Werner, A. Niedostatek, H. Kunath: Das Onkologieinformationssystem MADOS4 – ein etabliertes Werkzeug zur Tumordokumentation im Klinischen Krebsregister und in kooperierenden Einrichtungen des Tumorzentrums Dresden

[12] Export und Unterstützung der Vergütung von Dokumentationen ab Version 2 des Exports zum Gemeinsamen Krebsregister. Available: http://www.med.uni-giessen.de/akkk/gtds/

[13] Gonzalo Navarro: A Guided Tour to Approximate String Matching, ACM Computing Surveys, Vol. 33, No. 1, March 2001, pp. 31-88


**INSERT YOUR PICTURE**

**Yan Li** is a researcher in the Internet Technologies and Systems group at the Hasso-Plattner-Institute for IT-Systems Engineering (HPI) at the University of Potsdam, Germany.

Her main research interests are telemedicine and service-oriented architecture (SOA).

**Prof. Dr. sc. nat. Christoph Meinel** is President and CEO of the Hasso-Plattner-Institut for IT-Systems Engineering (HPI) and full professor (C4) for computer science at the University of Potsdam.

His research field is Internet and Web Technologies and Systems (www.hpi.uni-potsdam.de/meinel). Beside he is a teacher at the HPI School of Design Thinking, a visiting professor at the Computer Science School of the Technical University of Beijing (China) and a research fellow of the interdisciplinary center SnT at the University of Luxembourg. Since 2008 he is program director of the HPI–Stanford Design Thinking Research Program. Since 2010 he chairs the Steering Committee of the HPI Future SOC lab.

Christoph Meinel is author or co-author of 8 text books and monographs and of various conference proceedings. He has published more than 350 per-reviewed scientific papers in highly recognised international scientific journals and conferences. His high-security solution Lock-Keeper is international patented and licensed by Siemens AG. His tele-TASK system provides an innovative mobile system for recording and Internet broadcasting lectures and presentations used in many universities all-over the world. The virtual tele-lab for Internet Security provides the possibility to get hands-on experiences in practical issues of internet and information security. The recently developed Tele-Board supports remote work of creative teams.

Furthermore, Christoph Meinel is chairman of the German IPv6 council, and chairs the advisory board of UTD Meraka in South Africa. In 2006, he hosted together with Hasso Plattner the first German "National IT-Summit" of the German Federal Chancellor Dr. Angela Merkel. From 1998 to 2002 he was the founder and CEO of the Research Lab "Institute for Telematics" in Trier. Christoph Meinel is chief editor of the scientific electronic journals "ECCC – Electronic Colloquium on Computational Complexity" and "ECDTR – Electronic Colloquium on Design Thinking Research", the "IT-Gipfelblog", and the "tele-TASK"-archive.