THE D-SDA REPORTING SYSTEM: REPORTING AND USER ACCESS AT DLR-EOC

Johanna Senft, Cristian Chereji, Mario Winkler, Katrin Molch, Eberhard Mikusch

German Aerospace Center
German Remote Sensing Data Center
Oberpfaffenhofen, D-82234 Weßling, Germany

Abstract

In recent years, Earth observation missions and projects have put ever-increasing emphasis on reporting. Detailed reports are being requested on system performance, user behaviour and processing workload. At the same time reporting is gaining significance for data centres' internal purposes as well – for identifying bottlenecks and optimizing system performance and reliability. For future planning, Earth Observation data centres increasingly need information-based forecasts. Reliable reports help in planning system evolution and upgrades. The German Satellite Data Archive (D-SDA) at the German Aerospace Centre DLR has established a comprehensive reporting system to respond to a variety of reporting demands – fully integrated with the data and information management system as well as the IT infrastructure. Running in a state-of-the-art software environment, the D-SDA Reporting System is accessible by an easy-to-use interface and offers a high-degree of flexibility regarding input and output formats. This paper will outline the challenges for reporting at DLR-EOC and introduce the reporting system architecture embedded in the D-SDA multi-mission data management environment. It will highlight current developments and future challenges.

INTRODUCTION

The DLR's Earth Observation Centre (EOC) operates the German Satellite Data Archive (D-SDA), which provides archiving, and access services to national and international space-borne and airborne earth observation missions, campaigns, and scientific projects. The technological basis of the D-SDA is the Data and Information Management System (DIMS), developed in-house in cooperation with an industrial partner. Making use of flexible, generic components in a multi-mission environment, DIMS manages ingestion, cataloguing, archiving, and delivery of the earth observation data and products. The current data volume held in the D-SDA exceeds 5 petabytes, with an annual increase of nearly one petabyte.

Data discovery and access are provided via standardized metadata, data protocols, interfaces, and user-friendly data portals, such as the EOWEB Geoportal [1]. Figure 1 shows the overall architecture of earth observation data management and access at DLR-EOC.

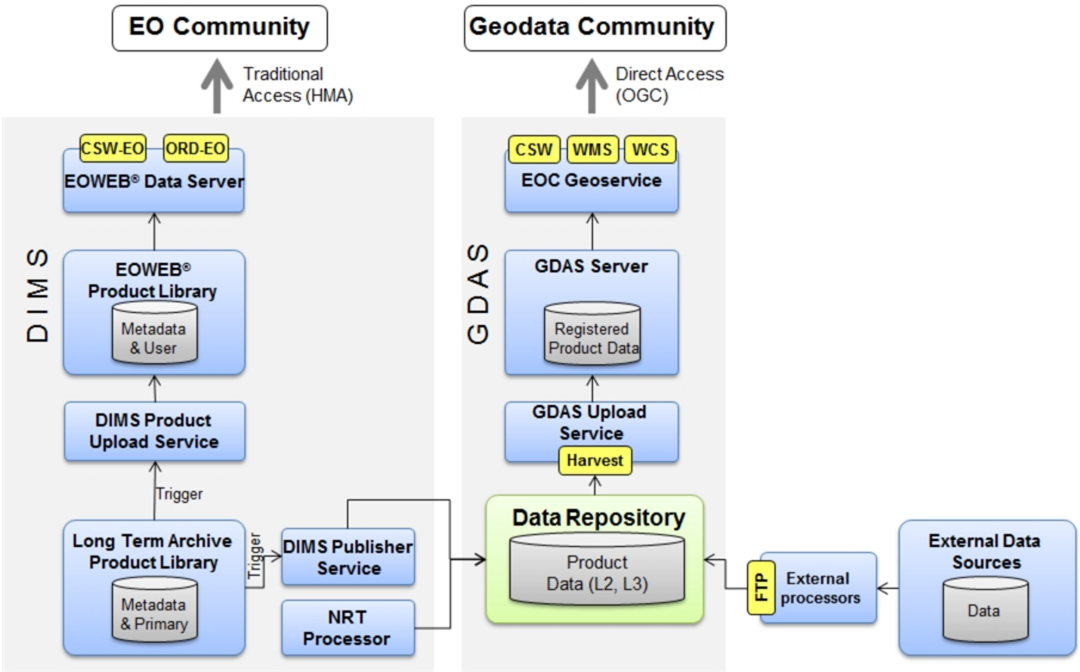

Figure 1: Earth observation data management and access at DLR-EOC.

Each part of this infrastructure carefully tracks data management and archive operations. This information is a valuable resource for generating statistics and reports. In order to respond to the variety of reporting needs, it is of critical importance to collect, prepare, present, and preserve this information within a carefully designed and managed, dedicated infrastructure.

Within the DLR-EOC each mission has its own reporting requirements and their implementation is based on distributed sources. Thus - from a reporting perspective - achieving the necessary level of harmonization in terms of data structure is of paramount importance. On the other hand, the demand for statistical information in such a big system covers a wide range of reports for various purposes: pure technical information, system performance reports, usage statistics, prediction reports, and comparisons between systems or missions. Some of these are key factors in management decisions. The D-SDA Reporting System has been developed at DLR-EOC in order to provide answers to these challenges. With its modular architecture, the D-SDA Reporting System collects data from different sources and integrates them into a harmonized data warehouse which is compatible with all missions in the multi-mission environment [2]. It features a high degree of flexibility in data analysis and aggregation, opening the way to a large variety of statistics and reports. These are presented to the customer via a modern user-friendly interface which is not only the starting point for generation, dissemination, and archiving of reports, but also ensures the security and confidentiality of the information.

CHALLENGES FOR REPORTING AND STATISTICS

The data that are relevant as input for the D-SDA Reporting System are spread over a variety of different servers and systems, based on UNIX, Linux, or Windows operation systems. The general communication between systems is facilitated by standard interfaces and queueing systems. There is a wide structural diversity of the data to be collected as input, reaching from plain-text log-files and XML journal-files to information extracted from databases. The variety of input data sources and formats is one of the critical challenges to be addressed by the D-SDA Reporting System.

The timeliness of the input data is another challenging factor. Some input data are collected in real-time from processing systems; others are read offline from logging information of events already completed some time ago. The respective time scales vary between days, hours, and minutes.

The temporal and structural variety of the input data calls for a harmonization layer embedded into the reporting system. The D-SDA Reporting System solves this issue by ingesting most input data into a reporting data warehouse, which holds the input data in a harmonized form and constitutes the basis for report generation.

The creation of reports has to be flexible because of the variety of user needs. Some reports are periodic, routine reports showing information in the same manner at regular time intervals. The generation of this type of report is usually timer driven but can also be ad-hoc and relies on report templates. However, the system has to provide the capability to also generate on demand special interest reports, the content of which cannot reliably be predicted in advance. For generating those types of reports know-how of the reporting database and a flexible easy-to-use interface is needed.

The need for flexibility extends also to the aspect of report delivery formats. Some ‘higher level’ reports are based on formatted graphs and tables in a ‘fixed’ data format such as PDF. Other reports are very flexible, data-only reports, delivered e.g. as comma separated values without any or with a minimum set of formatting requirements. Figure 2 shows some examples of different reports generated by the D-SDA Reporting System.

Figure 2: Examples of different reports generated by the D-SDA Reporting System.

The system must also account for variability in the degree of aggregation of report data. The requirements reach from highly aggregated reports, representing only few statements (e.g. management reports, archive statistics), to delivering large amounts of data without any aggregation, providing therefore the possibility for further handling and processing the information (e.g. custom aggregation and formatting).

Additionally, different delivery channels are requested. Currently the D-SDA Reporting System supports delivery via e-mail and server access (SFTP).

The demand for flexibility extends to the aspect of security and confidentiality of information on different usage levels. This subject, however, is beyond the scope of this paper.

While currently the report creation is based on less flexible, preformatted templates, in future, to cope with the variety of requirements, customizable report generation and delivery via an easy-to-use graphic user interface is being envisaged.

SYSTEM ARCHITECTURE

The D-SDA Reporting System at the DLR-EOC is subdivided into three major components shown in Figure 3. The Customer component at the top of the system architecture specifies the major user groups and their access to the report production via a user interface. The Reporting Production component manages the report assembly, dissemination and archiving processes. Below, the Reporting Data Services contains all components managing the data handling, data warehouse, and data collection from the various data sources. As a subcomponent the ETL (Extract Transform and Load) component provides the input data from data sources in order to be ingested into the reporting data warehouse. This data is then used by the Reporting Production for further processing and analysis.

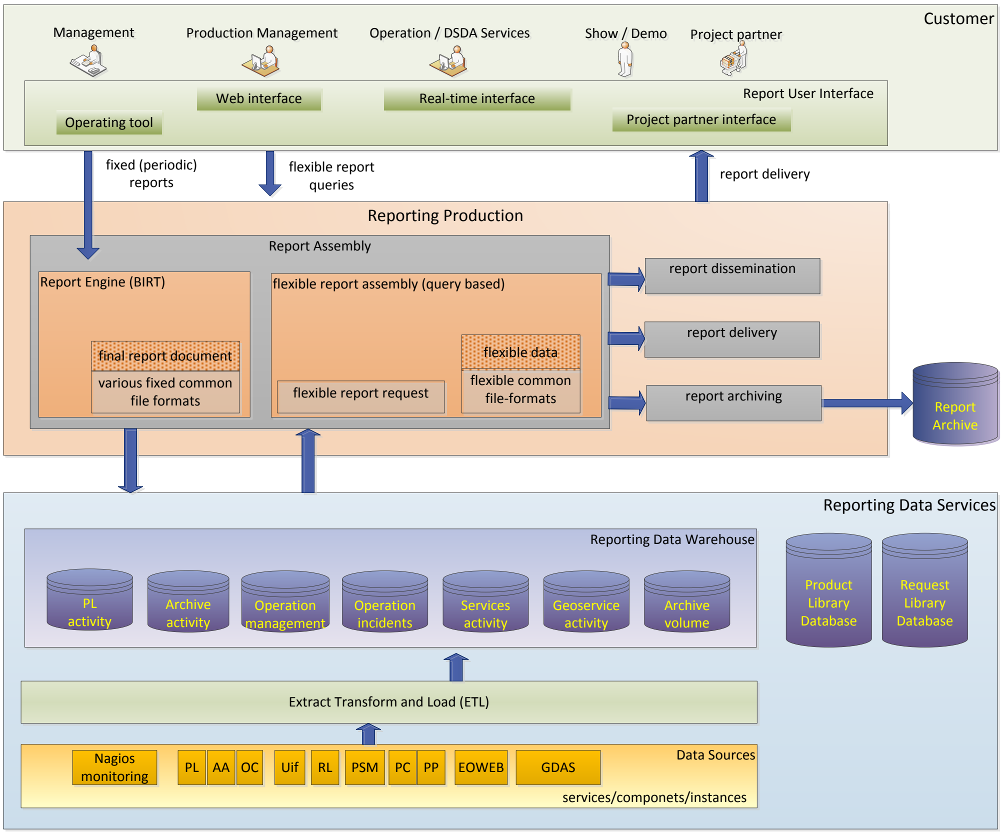

Figure 3: Architecture of the D-SDA Reporting System.

DATA COLLECTION

The data collection for reporting in the D-SDA Reporting System includes the process of data extraction from the different systems, transformation of the input data, and load into the data warehouse (Figure 4). All three tasks are accomplished by the Extract Transform and Load (ETL) component which takes both an active and a passive role. In its active role the ETL component initiates the data extraction, while in its passive role the ETL component merely passes on messages from other systems. The extraction tasks of the ETL component can be classified into three main categories:

- Extraction of data from other databases,
- Message driven extraction from other systems via a queue server,
- Data extraction from files of different formats and from various locations.

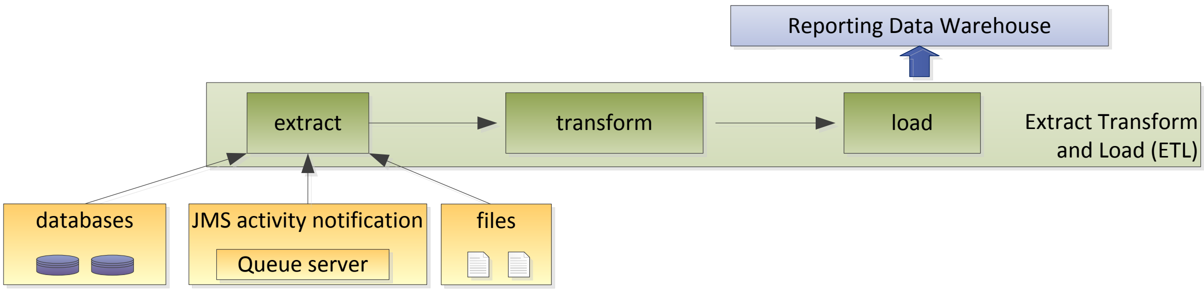

Figure 4: Data collection of the D-SDA Reporting System.

Each of these three main categories is described by four characteristic parameters, i.e. the physical extraction process, the status of the ETL component, the use/need of temporary data, and the expansion possibilities.

Data extraction by querying other databases is usually triggered by schedulers. In this case the ETL component actively initiates the requests. This method is used to extract the processing information of systems which generate only temporary data, such as near real-time data. So using temporary databases is an essential skill. Subsequently data spread over several database tables are aggregated in a single data source. This is an important issue for optimizing the performance of the reports.

Message-driven data extraction makes use of the Java Message Service (JMS), which retrieves information about the activity of other systems. It has a broad and flexible range of applications. Each system's queue server creates a message sequence about its own activity, for instance the start, progress, status, and end of a processing event. Other examples are dissemination or archiving events. The ETL component passively waits for messages. Upon arrival, it picks up the notifications from the message queue and integrates these into specific data sets inside the reporting data warehouse. This technique also offers the possibility to route messages to other systems, such as forwarding alarm signals to monitoring systems.

Data extraction from files can handle various file categories, for example heterogeneous log-files or XML journal-files. Here the ETL component takes an active role and parses the different files, extracts the information needed, and inserts it into the reporting data warehouse. With this technology, data provided by older applications can also be integrated into the reporting data warehouse.

REPORT CREATION WORKFLOW

For the internal creation of fixed periodic reports, the D-SDA Reporting System supports the use of templates. For report creation open source software is used, i.e. the Business Intelligence and Reporting Tools (BIRT) which "provides reporting and business intelligence capabilities for rich client and web application […] based on Java" [3]. BIRT has two main components, a visual report designer for setting up reports and a report engine for generating runtime reports. It also includes a charting engine and supports a variety of final report formats. The Reporting System can also access a number of different data sources such as databases, web services, scripting objects, and XML systems.

For implementing such a report the initial step consists of defining a report template file, i.e. the 'report design'. It is generated in XML format and describes the data sources, their properties, as well as layout and charting of the specific report. As a second task during this step, the rules for data pick-up, data transformation, business logic, and data presentation are implemented.

In the actual implementation of the reporting workflow, the creation of reports based on fixed templates is possible. Report generation is then activated via a user interface, where required input parameters such as start/end time, aggregation level, dissemination options, archiving options, etc. are inserted as shown in Figure 5.

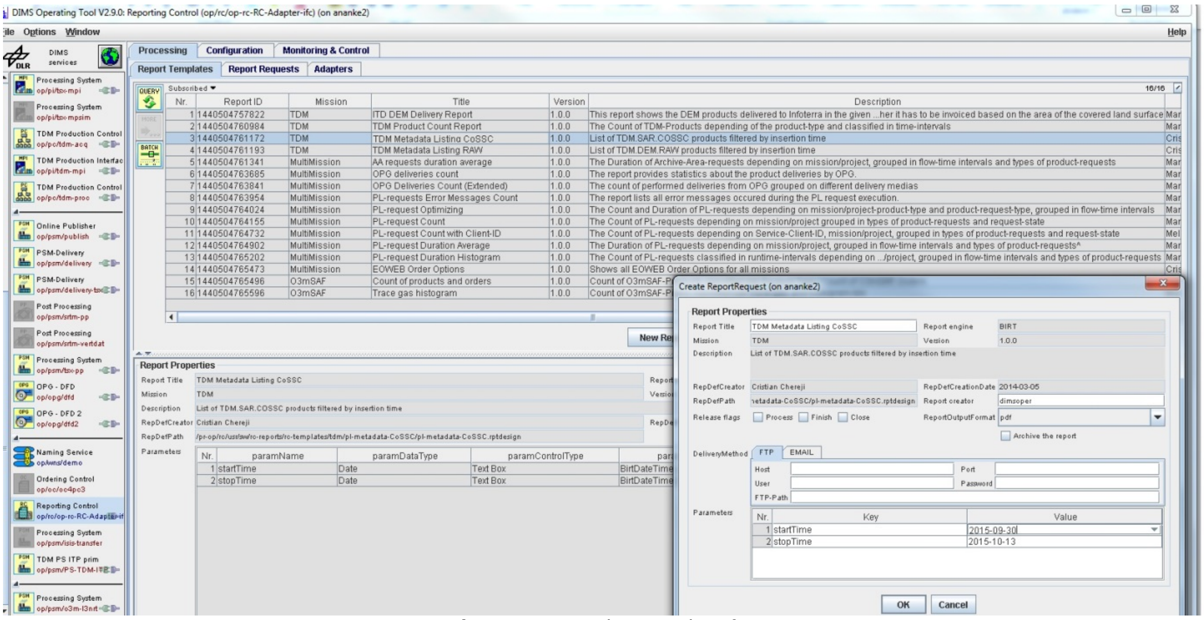

Figure 5: Reporting user interface

The entire “package” is then passed on through the reporting production finally to the report engine. This one fetches the associated template file, queries the data services, and assembles the report. The result is the final report document, which is returned to the requester via a dissemination service by email, SFTP, or any other supported means as specified.

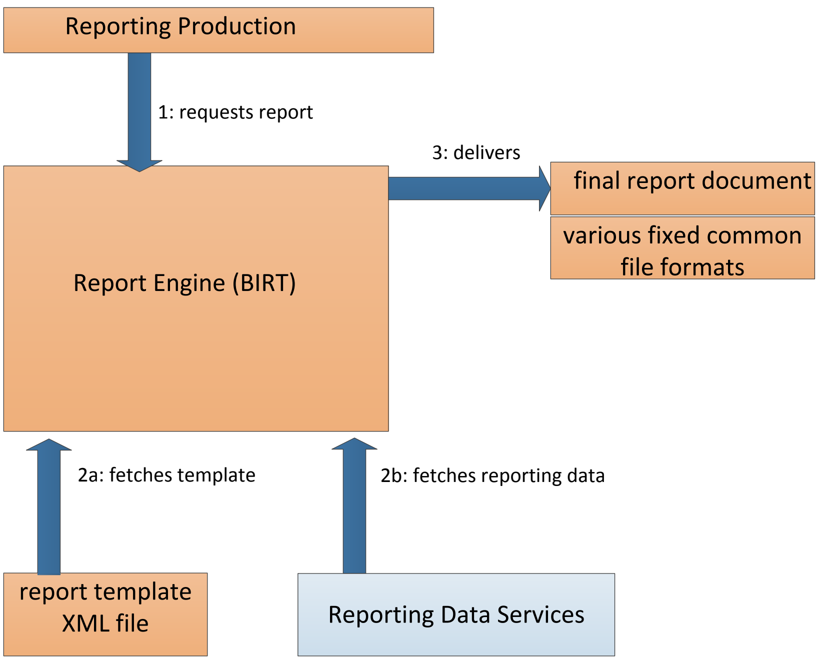

Figure 6: Report creation workflow.

The final report document can also be optionally archived and retrieved in a dedicated archiving system as illustrated in Figure 7.

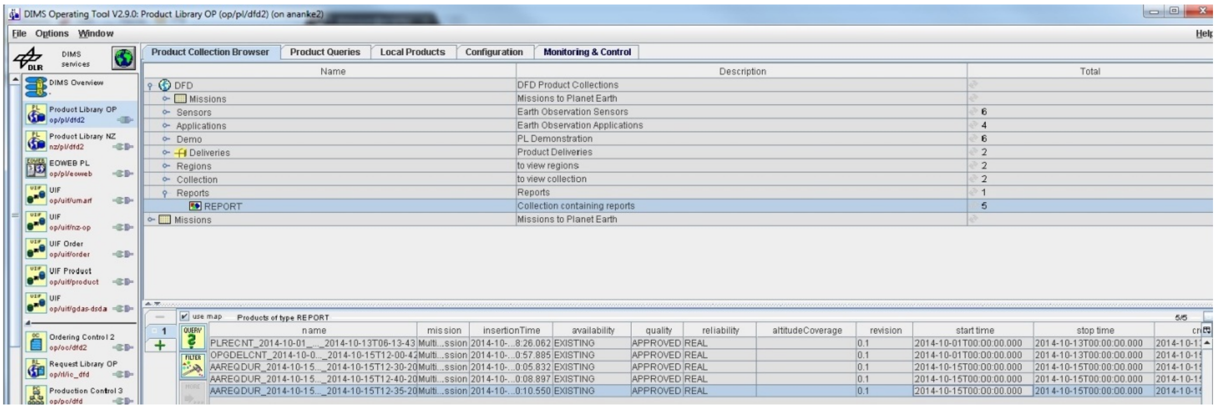

Figure 7: Report archive.

For the future the creation of flexible reports will be implemented, will allow customizable reporting queries and will be using a new graphic interactive user interface.

REPORTING SYSTEM ACCESS – VIA AN INTERACTIVE USER INTERFACE

For interactive report generation, user access to and direct interaction with the reporting system plays an important role. Low-level, system based access requires in-depth IT knowledge and understanding of the system, thus barring non-specialists from independently interacting with the system. Hence the next evolution step of the D-SDA Reporting System foresees the integration of a modern tool such as ‘ReportServer’. ReportServer is open source software which integrates and makes accessible many reporting tasks from within a single user interface [4].

The user interface is the entrance to the reporting system (Figure 6). On one hand it has to be a flexible tool providing integrated access to a large variety of optional, selectable possibilities. On the other hand it has to allow the user to efficiently access and analyse exactly the data she/he needs, and present them as desired. These analyses can range from print-ready, high-level evaluations to elaborate, fine-grained ad-hoc reporting.

Since a reporting data-base may contain sensitive or confidential information, protecting its content from unauthorized access and use is critical. A flexible authentication procedure using an Active Directory or single-sign-on interface provides the necessary level of security. For optimum functionality and categorizing users’ reporting demands, user groups and user teams can be defined. Multi-level, hierarchical structures currently appear to be the best approach. Thus each user can access her or his individual report domain which she/he can share with other team members in working groups.

The possibility of generating ad-hoc list-like reports via the user interface is also crucial. This allows flexible, direct access for analysis and configuration of the data. The data can be used later for individual processing, entered, for example in spreadsheet programs, converted into various output formats, or stored for further use, e.g. for generating variations of a report for different audiences. Another interesting feature for future developments, although not essential, is the ability to directly visualize the data in crosstabs or charts.

Flexible re-use and integration of the complete functionality already implemented in earlier development phases is an important aspect of the design and evolution of the D-SDA Reporting System. Variable report engines such as the BIRT engine already in use, allow to re-use existing report templates and designs without the need to re-create them in a different environment. Of equal importance is embedding an archiving functionality such as that provided by the ‘Report Archive Server.’

CONCLUSIONS

The D-SDA Reporting System has been developed and implemented in response to evolving internal and external requirements for reporting in Earth Observation archives. Those needs and demands are expected to further increase in the future as a result of a rising interest in using Earth Observation data, different mission and project requirements as well as more stringent budgetary constraints. Therefore flexible and reliable reporting systems with easy-to-use interfaces are needed – integrated into and communicating with all components of the respective archiving data management system in use. In consequence, it is of fundamental importance to carefully define system architectures and interfaces and to develop workflows which efficiently handle and present the information required.

REFERENCES

[1] K. Dengler, T. Heinen, A. Huber, K. Molch, and E. Mikusch, “The EOC Geoservice: Standardized Access to Earth Observation Data Sets and Value Added Products” in Proc. *PV 2013 – Ensuring the Long-Term Preservation and Value Adding to Scientific and Technical Data*, 4-6 November 2013, Frascati, Italy.

[2] M. Guettler, V. Manilici, K. Molch, and S. Kiemle, “Reporting in a Payload Data Ground Segment” in Proc *PV 2011 - Ensuring the Long-Term Preservation and Value Adding to Scientific and Technical Data*, 15-17 November 2011, Toulouse, France.

[3] Wikipedia, The Free Encyclopedia, “BIRT Project”, online resource, accessed at https://en.wikipedia.org/wiki/BIRT_Project on 16 September 2015.

[4] ReportServer, “ReportServer – The Business Intelligence Suite”, online resource, accessed at https://reportserver.net/en on 16 September 2015.

[5] Kai-Uwe Sattler, Stefan Conrad, “Vorlesung Data-Warehouse-Technologien: Extraktion, Transformation, Laden (ETL)”, Uni Magdeburg, online ressource, accessed at http://wwwiti.cs.uni-magdeburg.de/iti_db/lehre/dw/dw0001/dw04.pdf on 13 October 2015.

[6] IT process maps,” ITIL-Kennzahlen”, online resource, accessed at http://wiki.de.it-processmaps.com/index.php/ITIL-Kennzahlen on 13 October 2015.