

Effizientes und bedarfsgerechtes Berichtswesen auf Basis eines modularen Systems aus Open-Source-Produkten

Thomas Jankowski¹, Ulrike Lucke²

Abstract: Data-Warehouse-Systeme bzw. Business-Intelligence-Lösungen finden mittlerweile auch in der Hochschullandschaft eine immer weitere Verbreitung. Dabei kann auf fertige Lösungen zum Beispiel von Anbietern der Campus-Management-Systeme gesetzt werden, aber auch auf unabhängige Lösungen. Die Universität Potsdam hat sich für eine eigenständige Lösung auf Basis von flexiblen Open-Source-Produkten entschieden, um damit die vorhandenen Verwaltungsprozesse bestmöglich zu unterstützen. Erste Bereiche (Bewerber-, Studierenden- und Finanzverwaltung) wurden integriert. Der Beitrag gibt einen Überblick über die entwickelte Lösung und weist auf positive Erfahrungen, aber auch auf zu meisternde Herausforderungen hin.

Keywords: Business Intelligence, Data Warehouse, Berichtswesen, Hochschulstatistik, Systemarchitektur, Open Source

1 Einleitung und Begriffsklärung

Unter dem Schlagwort *Business Intelligence* (BI) haben rechnergestützte Formen der zielgerichteten Auswertung von Datenquellen eine zunehmende Bedeutung erlangt. An Hochschulen geht dies i.d.R. aus den gesetzlich vorgeschriebenen Pflichten (z.B. [BStatG87]) des *Berichtswesen* bzw. der *Statistik* hervor und betrifft v.a. die Bereiche Studium/Lehre und Forschung. Neben der hier seit langem bestehenden Herausforderung der Verarbeitung mit einer großen Vielzahl und Vielfalt von Datenquellen zeichnet sich Business Intelligence darüber hinaus durch das interaktive Arbeiten mit großen, komplexen Datenmengen aus und wird zunehmend als ein wichtiges Instrument der Hochschulsteuerung verstanden [Le04] (auch im Kontext der Systemakkreditierung).

In diesem Zusammenhang kommt dem Konzept eines sog. *Data Warehouse* (DWH, eine leistungsfähige Datenbank) eine zentrale Rolle zu. Hier werden Daten aus verschiedenen Quellen zusammengeführt. Der Begriff *Data Warehouse System* bezeichnet darüber hinaus die Gesamtheit aller Komponenten der unteren drei Ebenen (Datenerfassung, Datenhaltung und Datenbereitstellung). Dies wird in Abbildung 1 veranschaulicht. Auf dieser Basis bietet ein BI-System einen „IT-gestützten Zugriff auf Informationen sowie die IT-gestützte Analyse und Aufbereitung dieser Informationen. Ziel dieses Prozesses

¹ Universität Potsdam, Zentrale Einrichtung für Informationsverarbeitung und Kommunikation,
Am Neuen Palais 10, 14469 Potsdam, thomas.jankowski@uni-potsdam.de

² Universität Potsdam, Institut für Informatik, A.-Bebel-Str. 89, 14482 Potsdam, ulrike.lucke@uni-potsdam.de

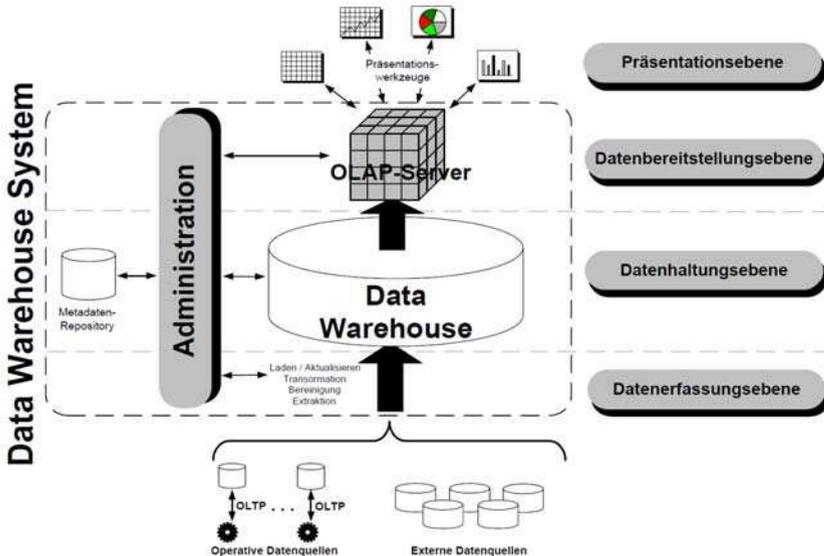


Abbildung 1: Komponenten eines Data Warehouse Systems nach [SBU99]

ist es, aus dem im Unternehmen vorhandenen Wissen neues Wissen zu generieren“ [SG15]. Über die systematische Datenaufbereitung auf den unteren drei Ebenen hinaus gelangen insbesondere die anschauliche Präsentation von bzw. die gezielte Interaktion mit Daten in den Fokus. Zudem wird nicht nur von Daten, sondern von den daraus gewonnenen Informationen gesprochen. Weiterhin wird die Datenerfassung als Prozess verstanden (sog. ETL-Prozess: extract, transform, load), für den es in der Praxis zusätzlicher organisatorischer Regelungen bedarf.

Mit Blick auf die IT-Infrastrukturen und Organisationsstrukturen sind bei der Einführung eines BI-Systems an Hochschulen zudem bestehende Rahmenbedingungen zu beachten. Dazu zählen v.a. die heterogenen IT-Landschaften mit oft mangelnden Schnittstellen und verteilten Verantwortlichkeiten. Das ergibt folgende Anforderungen an ein BI- System:

- flexible Schnittstellen zu vorhandenen Datenquellen (verschiedene Plattformen, Strukturen und Formate bis hin zu teilweise analoger Datenübermittlung)
- flexible Generierung von Statistiken (für externe und interne Anforderungen) auf Basis eines anpassbaren ETL-Prozesses
- ausführliche Metadaten (sowohl für die verwendeten Quelldaten als auch für die generierten Statistiken)
- dezentrale Nutzung & Bereitstellung von Berichten (sowohl in vordefinierter Form als auch ad hoc zusammengestellt)
- differenziertes Rechtekonzept für den verteilten Zugriff (in der Hoheit der für die Daten bzw. den Prozess fachlich Zuständigen)
- Protokollierung von Aktionen; Historisierung bestimmter Zustände der Datenbasis

Oftmals besteht zudem eine gewisse Angst vor komplexen Einführungsprojekten, auch angesichts schwergängiger Produkte [Me012][Ka13]. Dieses Hemmnis wird verstärkt durch ein i.d.R. geringes IT-Budget. Vor diesem Hintergrund beschreibt der Beitrag zunächst den aktuellen Stand der Technik, wobei der Fokus auf Open-Source-Produkte gelegt wird. Daraus werden eine Systemarchitektur und ein Verarbeitungsprozess abgeleitet und mit geeigneten Werkzeugen unterlegt. Abschließend werden das daraus resultierende System und die damit gemachten Erfahrungen beschrieben.

2 Stand der Technik

Die methodische Basis von Business Intelligence bildet das Online Analytical Processing (OLAP). Hier erfolgen interaktive Anfragen auf einem mehrdimensionalen Datenwürfel, der entlang struktureller und inhaltlicher Parameter gefiltert wird [Cla98]. Als praktische Umsetzung dieses Konzeptes haben sich zahlreiche Produkte auf dem Markt etabliert. Für alle vier o.g. Ebenen finden sich kommerzielle Lösungen von Spezialanbietern (z.B. HISinOne-BI³ für Hochschulen oder qlikview⁴ für Datenanalysen allgemein), aber auch von Allroundern (wie Microsoft Power BI⁵, Oracle BI Foundation Suite⁶, IBM Cognos BI⁷ oder SAP Crystal Reports⁸). Dabei decken qlikview und Crystal Report nur die Präsentationsebene ab. Zudem ist eine Reihe von Open-Source-Werkzeugen verfügbar (wie z.B. Pentaho⁹, JasperSoft¹⁰ und RapidMiner¹¹).

Die kommerziellen Werkzeuge sind in der für den Einsatzbereich Hochschule nötigen Konfiguration i.d.R. preisintensiv (Kaufpreis und Beratertage) – mit Ausnahme von HISinOne-BI, das aber stark auf die HIS-eigenen Datenquellen ausgerichtet ist. Mit Blick auf diese enge Integration in Standard-Software besteht hier aber noch einiges Potential. Die verfügbaren Open-Source-Werkzeuge halten jedoch einem Abgleich mit den Anforderungen der verschiedenen Nutzergruppen und den vorhandenen Datengrundlagen stand, auch weil sie kaum noch Unterschiede zu kommerziellen Lösungen bzgl. Wartungsaufwand zeigen. Tabelle 1 zeigt eine Bewertung der drei o.g. Werkzeuge anhand der in Abschnitt 1 identifizierten Anforderungen.

Dafür wurden die genannten Kriterien weiter ausdifferenziert¹² und jeweils die Eignung der Werkzeuge (Erfüllungsgrad in Stufen von 0/30/60/100%) geprüft. Gemäß den definierten Gewichten pro Kriterium/Gruppe entsteht die dargestellte Gesamtbewertung.

³ <https://www.his.de/produkte/hisinone/management/business-intelligence.html>

⁴ <http://www.qlik.com/>

⁵ <https://www.powerbi.com/>

⁶ <http://www.oracle.com/us/solutions/business-analytics/>

⁷ <http://www.ibm.com/software/products/de/business-intelligence/>

⁸ <http://www.crystalreports.com/>

⁹ <http://www.pentaho.de/>

¹⁰ <https://www.jaspersoft.com/>

¹¹ <https://rapidminer.com/>

¹² verkürzte Darstellung der ca. 50 einzelnen Kriterien; Details werden auf Anfrage bereitgestellt

Kriteriengruppe	Gewicht	Pentaho	JasperReports	RapidMiner
Systemschnittstellen	20,00%	100,00%	100,00%	100,00%
Einbindung vorhandener Datenquellen	15,00%	30,00%	30,00%	30,00%
ETL- und Statistik-Tools/ Benutzerfreundlichkeit	25,00%	75,00%	77,50%	69,50%
Kosten und Aufwand	15,00%	46,67%	46,67%	60,00%
Rollen und Rechte	15,00%	100,00%	100,00%	28,57%
Sonstiges (z.B. Support)	10,00%	80,00%	66,00%	80,00%
Gesamtbewertung		73,25%	72,48%	63,16%

Tabelle 1: Vergleich von Open-Source-Tools mit Anforderungen von Hochschulen

Im Ergebnis liegt Pentaho leicht vor JasperSoft, aber auch Rapidminer wäre zumindest eingeschränkt geeignet. Den Einsatz von Pentaho und JasperSoft im Rahmen eines modularen Systems beschreibt Abschnitt 3.

3 Technische Lösung

3.1 Systemkomponenten

Ausgehend von der vorgestellten Architektur eines DWH-Systems wurde mit Blick auf die Stärken und Schwächen der einsetzbaren Werkzeuge eine verteilte Lösung aus unterschiedlichen Komponenten konzipiert. Dabei werden die Datenbereitstellungs- und die Präsentationsschicht verschmolzen, da anstelle einer komplexen OLAP-Software die Abbildung der nötigen Datenstrukturen in einer relationalen Datenbank auf einfachere Weise möglich ist (sog. Relational OLAP [CMR02]). Die resultierende Architektur ist in Abbildung 2 dargestellt. Die einzelnen Ebenen werden nachfolgend kurz erläutert.

Operative Systeme

Hierzu zählen alle in der Hochschule eingesetzten Legacy-Tools für die Verwaltung zur Hochschulsteuerung relevanter Daten, wie für Studienbewerber, Studierende, Lehrveranstaltungen, Prüfungsleistungen, Personal, Finanzen, Drittmittel, Publikationen usw.

Datenerfassungsebene

Hier erfolgen die Extraktion, Transformation und das Laden der Daten in das DWH. Dabei werden komplexe ETL-Prozesse in kleine, handhabbare Teilprozesse zerlegt. Es kann auf beliebige Datenquellen zugegriffen werden (bspw. relationale Datenbanken, Excel-Dateien, strukturierte Textdateien). Für jede Datenquelle werden ein eigener ETL-Prozess und Importrhythmus definiert, die automatisch ausgeführt werden. In Abschnitt 3.2 sind die Möglichkeiten des ETL-Prozesses näher erläutert.

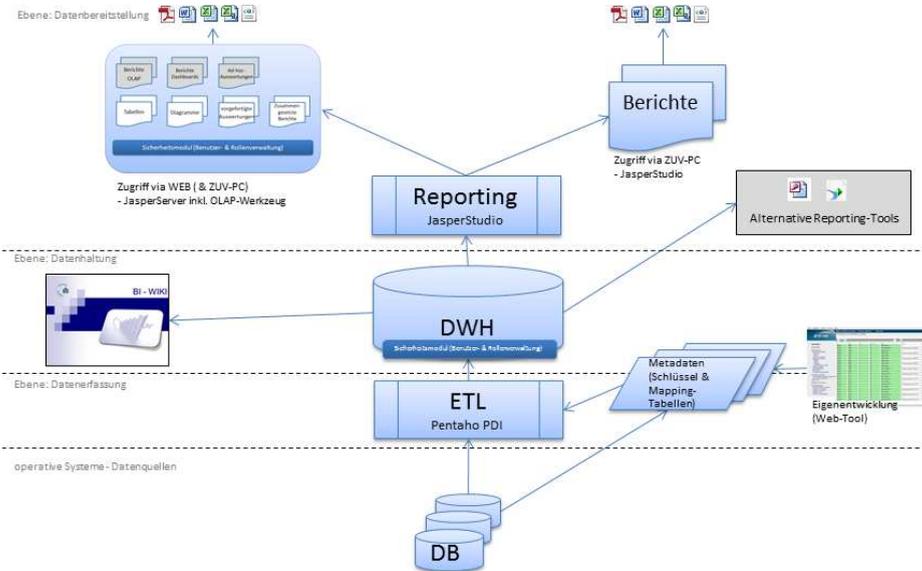


Abbildung 2: Architektur des modularen BI-Systems

Datenhaltungsebene

Zur Verwaltung der angebotenen Datenquellen werden zwei Tools eingesetzt.

- In einem Wiki werden beliebige Informationen zum Data-Warehouse-System manuell erfasst. Das sind insbesondere Erläuterungen zu den Berichten, Beschreibungen der Datenquellen pro Datenbereich, häufig gestellte Fragen und ihre Antworten, fachliche Ansprechpartner sowie Informationen zu den ETL-Prozessen. Das Wiki bietet die Möglichkeit, dass Informationen zu den Berichten dezentral durch die fachlich zuständigen Mitarbeiter gepflegt werden können.
- Zudem wird eine Web-Anwendung zur Datenverwaltung eingesetzt, die ebenfalls dezentral durch die fachlich zuständigen Mitarbeiter gepflegt werden kann. Das umfasst vier Arten von Daten: *Schlüsseltabellen* enthalten zusätzliche Angaben zu beliebigen Schlüsseln (bspw. Kostenstellenschlüssel), das können abweichende Gruppierungs- oder Sortierungsmerkmale oder auch überlagernde Bezeichnungen sein. *Mappingtabellen* enthalten Regeln zum Umschlüsseln von Daten, z.B. wenn in zwei operativen Systemen für die gleiche Kostenstelle unterschiedliche Schlüssel verwendet werden. *Technische Metadaten* enthalten bspw. Informationen über die erfolgten Datenimporte und erlauben den Fachanwendern die Festlegung, wann Daten als historisiert (d.h. unveränderbar) gekennzeichnet werden. Die Pflege *zusätzlicher fachlicher Daten* ist notwendig, wenn es für bestimmte Informationen kein zugrundeliegendes operatives System gibt, die Daten aber dennoch in Statistiken verwendet werden sollen. Dies kann auch sinnvoll sein, wenn sich aufgrund geringer Datenmengen die Entwicklung eines Importprozesses nicht lohnt.

Datenbereitstellungs- und Präsentationsebene

Es werden Werkzeuge eingesetzt, welche OLAP-Analysen und die Erstellung von Standardberichten auf Basis der relationalen Datenbank des Data-Warehouses ermöglichen. Die Inhalte der Standardberichte werden im Moment der Berichts-anforderung aktuell ermittelt und können durch den Anwender dynamisch über Filterkriterien eingeschränkt werden.

3.2 Prozess von der Datenerhebung bis zur Statistik

Neben der technischen Lösung bedarf es auch organisatorischer Vereinbarungen für die Bereitstellung und Verarbeitung von Daten, die wiederum technisch abgebildet werden müssen. Abbildung 3 zeigt die dafür nötigen Schritte und die beteiligten Akteure; dies sind i.d.R. die Fachabteilung und der Data-Warehouse-Experte.

Es sind acht Schritte vom Erkennen des Statistikbedarfs bis zur Veröffentlichung einer Statistik erkennbar. Zunächst muss die Fachabteilung in der Lage sein, die gewünschten Statistiken/Berichte und die anzuzeigenden Merkmale (Daten) klar zu beschreiben (Schritt 1). Anhand dieser Informationen kann der DWH-Experte erkennen, welche Daten zusätzlich benötigt werden (Schritt 2). Anschließend beginnt eine detaillierte Analyse, welche Daten tatsächlich vorhanden sind, welche zusätzlich erhoben werden müssen und welche nicht bereitgestellt werden können (Schritt 3). In der Folge kann die Datenbankstruktur im DWH entworfen und realisiert werden (Schritt 4). Dann folgt mit der Realisierung des ETL-Prozesses der aufwendigste Teil (Schritt 5). Hierbei geht es nicht nur darum, die Daten von einer Datenquelle in das DWH zu kopieren. Vielmehr können diverse Verarbeitungsschritte in dem ETL-Prozess integriert werden:

- Plausibilitätsprüfung (Sind alle Pflichtangaben vorhanden? Sind die Schlüssel gültig? Gibt es Verstöße gegen inhaltliche oder fachliche Regeln?)
- automatische Fehlerkorrekturen (soweit möglich)
- Schlüssel-Mapping (Zuordnung von Schlüsselwerten nach definierten Regeln)
- Prozesskonstanten und -variablen festlegen (z.B. Zeitstempel als Importdatum)
- dynamische Wahl des ETL-Prozess-Pfades in Abhängigkeit von den zu importierenden Daten (Beispiel: Drittmittelfinanzdaten durchlaufen einen anderen ETL-Pfad, als Haushaltsmittelfinanzdaten. Somit können andere Plausibilitäten abgebildet werden, auch wenn es sich um die gleiche Schnittstelle handelt.)
- Gruppierung und Aggregation von Daten
- Ergänzung neuer Merkmale nach festgelegten Regeln
- Fehlerbehandlung (d.h. Erzeugung und Versand von Fehlerprotokollen, ggf. Löschung der Daten des letzten Imports, Benachrichtigung des fachlich Zuständigen, Abbruch des Importvorgangs)

Im Schritt 6 liegen die notwendigen Daten bereits im DWH vor. Anschließend müssen die organisatorischen Rahmenbedingungen für die Nutzung der Daten geschaffen werden (Schritt 7), bevor die Berichte und Statistiken freigeschaltet werden (Schritt 8).

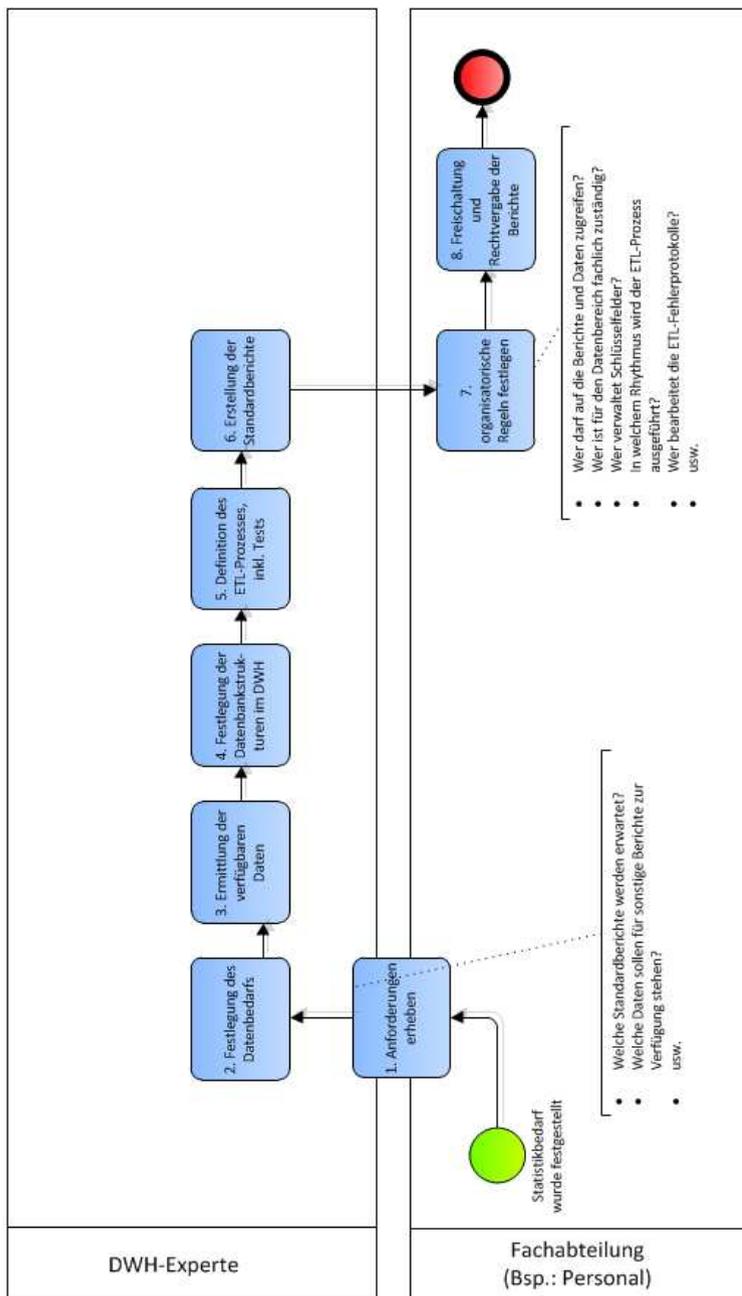


Abbildung 3: DWH-Gesamtprozess – von der Datenerhebung bis zur Statistik

3.3 Eingebundene Open-Source-Werkzeuge

In dem in Abschnitt 2 vorgestellten Produktvergleich liegt Pentaho insgesamt leicht vor JasperSoft. Hinter diesen beiden Produktnamen verbergen sich jedoch eine Sammlung von Open-Source-Tools, die jeweils für einen bestimmten Einsatzbereich (bspw. Datenerfassung oder Präsentation) gedacht sind. Diese sind teilweise aufeinander abgestimmt, können aber auch separat genutzt werden. Schon während der Produktbewertung zeigte sich, dass die Community-Version von Pentaho zwar geringfügig bessere Funktionen bietet, aber im Präsentationsbereich vereinzelt Probleme auftraten (verschwindende Felder, unerklärliche Fehlermeldungen).

Gemäß der entwickelten Systemarchitektur fiel die Entscheidung daher differenziert aus. Für die Ebene der Datenerfassung wurde zur Definition der ETL-Prozesse Pentaho gewählt. Pentaho hatte im Gesamtvergleich die höchste Bewertung erhalten, und ließ sich im ETL-Bereich auch problemlos ohne Support durch den Hersteller implementieren. Die Datenbereitstellung und die Präsentation (sowohl von OLAP-Analysen als auch von Standard-Berichten) erfolgen jedoch mit Hilfe von JasperSoft. JasperSoft schnitt im Gesamtvergleich nur geringfügig schlechter als Pentaho ab, konnte jedoch auch ohne Support fehlerfrei implementiert werden. Für die Ebene der Datenhaltung war keine neue Lösung notwendig; hier wurde der vorhandene Postgres-Datenbank-Cluster der Universität genutzt.

4 Beispielhafte Umsetzung

Nachfolgend wird die Umsetzung für Statistiken aus dem Bereich der Hochschulbewerberdaten (Studienbewerbung und Immatrikulation) dargestellt – von konkreten Datenquellen über deren Import / Verarbeitung hin zu generierten Statistiken und deren abgestufter Freigabe für verschiedene Nutzergruppen. Andere Domänen wie z.B. Forschungsinformationssysteme können auf vergleichbare Weise gestaltet bzw. angebunden werden.

4.1 ETL-Prozess zur Datenerfassung

Die Definition des Importprozesses erfolgt in Pentaho grafisch. Dabei können ETL-Prozesse in beliebige Sub-Prozesse gegliedert und grafisch modelliert werden. Prozesse können zu Analysezwecken schrittweise durchlaufen und der Datenzustand bei jedem Prozessschritt geprüft werden. Abbildung 4 zeigt den Hauptprozess für Bewerberdaten, der als Datenquelle HISinOne (Modul APP) nutzt. Jeder Prozess hat ein Start- und ein Endereignis und gliedert sich in diverse Schritte. Dabei kann es sich um Einzelarbeitsschritte (grüne Symbole) oder um komplexe Sub-Prozesse (orange Symbole) handeln. Die Prozessschritte können nacheinander oder auch parallel ablaufen. Bei parallel ablaufenden Prozessschritten gibt es zuvor eine Weiche, welche bestimmte Bedingungen abprüft, wie in Abbildung 4 dargestellt. Die Wahl des Weges wird jedoch

in diesem Beispiel nicht automatisiert durch eine Bedingung gesteuert, sondern der Prozessverantwortliche hat hier die Möglichkeit manuell einzugreifen und den linken (regulärer, d.h. täglicher Datenimport) oder den rechten Pfad (initialer, d.h. einmaliger Datenimport) festzulegen.

Jeder Importschritt kann wiederum aus einem Sub-Prozess bestehen. Der Prozess kann schrittweise durchlaufen werden, wobei nach jedem Schritt die Daten geprüft werden können. Im Laufe des ETL-Prozesses können beliebige Protokolle geschrieben werden, entweder in die Datenbank oder in das File-System. Es ist auch möglich, in Abhängigkeit von bestimmten Bedingungen (z.B. beim Auftreten eines Fehlers) E-Mail-Benachrichtigungen zu versenden.

4.2 Berichtsdefinition und -bereitstellung

Die Berichte und Statistiken werden mit JasperStudio erstellt, wie Abbildung 5 anhand einer Auswertung von Bewerberdaten zeigt. Neben den primären Ausgabeelementen (vordefinierte Texte und Grafiken; dynamisch erzeugte Tabellen, Diagramme, Berichte, Grafiken, Verzeichnisse usw.) können auch Steuerungselemente (z.B. Selektions- und Sortierparameter) und eigene Java-Script-Bausteine spezifiziert werden.

Die Berichte können über Jasperserver im Web angeboten und mit Rechtedefinitionen versehen werden. Im Ergebnis können sich Angehörige der Hochschule, gemäß den für sie definierten Freigaben, Berichte online ansehen bzw. analysieren. Die Berichte werden in einer Baumstruktur dargestellt, wie in Abbildung 6 dargestellt ist. Zusätzlich

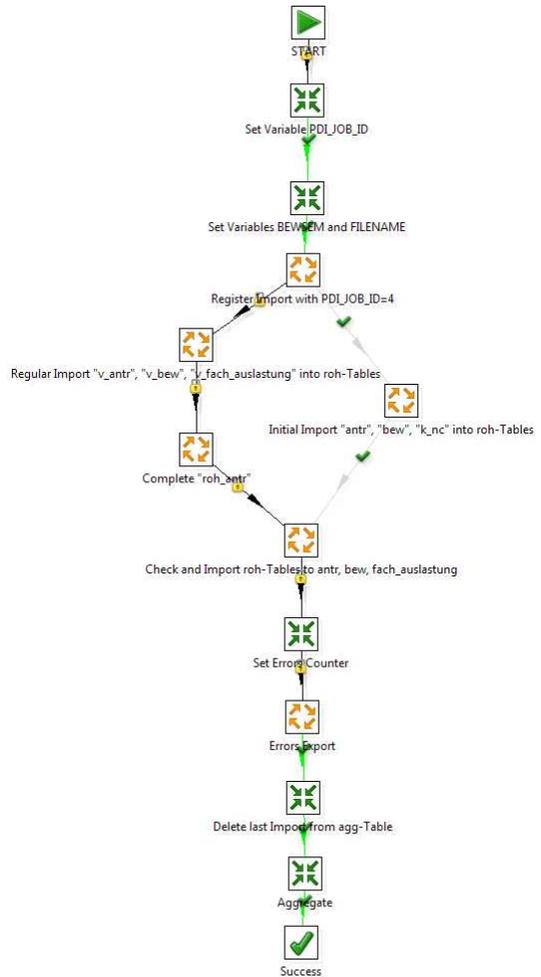


Abbildung 4: Grafische Modellierung des ETL-Prozesses in Pentaho

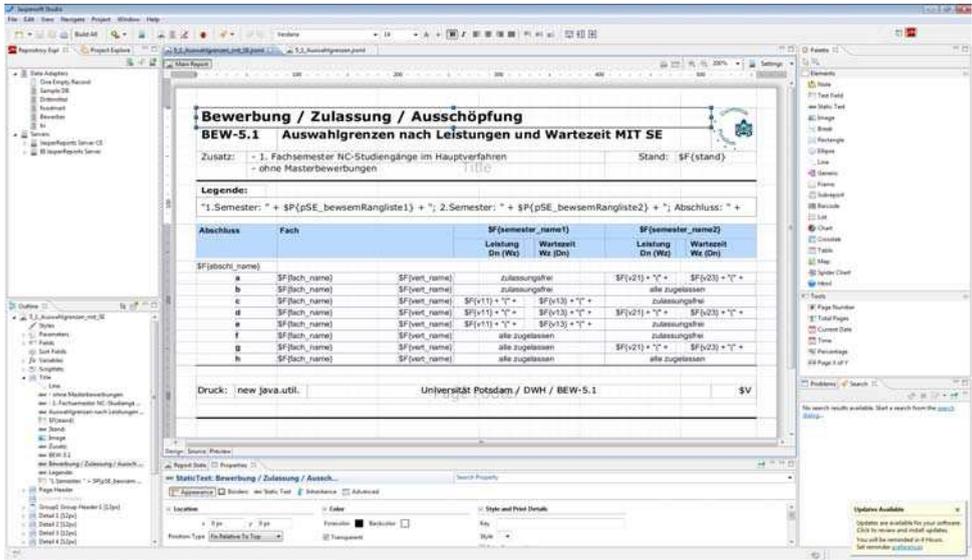


Abbildung 5: Definition einer Statistik mittels JasperStudio

kann nach Stichworten gesucht oder in Drittanwendungen (PDF, Excel usw.) exportiert werden. Vor der Ausführung eines Berichts ist es möglich, Filterkriterien (Parameter) vom Nutzer abzufragen. Zudem können Daten mit Hilfe eines OLAP-Werkzeuges analysiert werden; dies ist hier jedoch nicht dargestellt.

5 Zusammenfassung und Ausblick

In Kontrast zu den etablierten BI-Lösungen „aus einem Guss“ wird in diesem Beitrag ein modulares System für die Erstellung von Berichten und Statistiken ausgehend von den vorhandenen Werkzeugen vorgehaltenen Datenbeständen propagiert. Der beschriebene Ansatz ist an der Universität Potsdam für das Themengebiet Studierendendaten bereits umgesetzt; weitere Gebiete werden aktuell integriert. Der dargestellte Weg erfordert internes Personal, welches sich mit den Fachdaten, der Erstellung der Datenbanken, der Entwicklung von Datenbankabfragen und dem Berichtsentwurf auseinandersetzt. Dabei ist die Implementierung der ersten Schnittstelle (Datenquelle) zeitlich am aufwendigsten. Die Anbindung der ersten Datenquelle bis hin zur Entwicklung von zwölf Statistiken dauerte effektiv etwa 90 Personentage.

Bereits jetzt sind Auswirkungen auf Entscheidungsprozesse der Hochschule erkennbar. So werden bspw. Bewerbungs- und Zulassungsverfahren durch tagesaktuelle Berichte bereits während der Bewerbungsphase unterstützt, und die Hochschulplanung profitiert

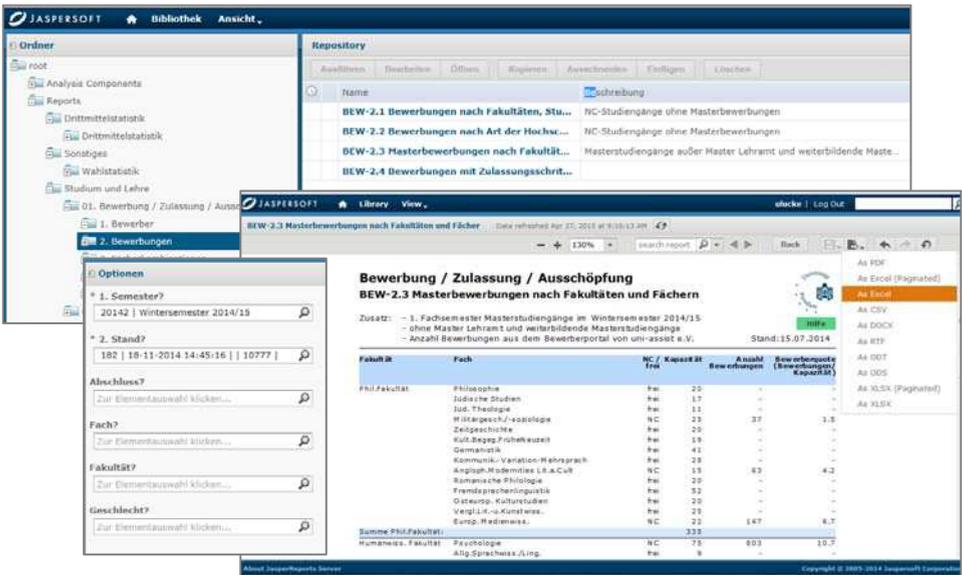


Abbildung 6: Zugriff auf die verfügbaren Berichte in JasperServer

u.a. von detaillierten Auslastungsberichten der Studiengänge. Im Vergleich zu dem früher sehr aufwändigen, langwierigen und fehleranfälligen Verfahren der Erhebung und Auswertung statistischer Daten ist hier nicht nur eine klare Verbesserung für die zentrale Verwaltung bzw. die Hochschulleitung erkennbar, sondern auch die dezentralen Prozesse in den Fakultäten und Einrichtungen der Hochschule werden vereinfacht. Das verbessert die wahrgenommene Datenqualität und so den zweckmäßigen Einsatz [Vo14]. Zugleich wird die Zugänglichkeit der Datenbestände erhöht, was die Transparenz von Entscheidungsprozessen und damit allgemeine Werte wie Identifikation, Beteiligung und Veränderungsbereitschaft signifikant fördert.

Auch aus technischer Perspektive ist die gewählte Lösung positiv zu beurteilen. Die für das Einführungsprojekt vorgesehene Mittelplanung wurde unterschritten: Zum Einen konnte der Sachkostenansatz auf die Beschaffung von Hardware reduziert werden. Darüber hinaus hat sich auch im Bereich Personalkosten eine Entlastung durch die Verlagerung von Aufwänden ergeben, die sich aber aufgrund der noch nicht reibungsfreien Abläufe bei der Datenaufbereitung derzeit nicht verlässlich quantifizieren lassen: Einerseits werden in den Fachabteilungen doppelte Arbeiten beim Sammeln und Transformieren von Daten vermieden, andererseits traten gerade in der Einführungsphase vorübergehende Mehrbelastungen auf. Die unmittelbare Implementierung der lokalen Prozesse und Datenstrukturen war jedoch weniger komplex als die Einführung eines vglw. starren Komplettsystems. Das Risiko des weiteren Betriebs der entwickelten Lösung konnte durch den Rückgriff auf standardisierte Architekturmuster und Support-Verträge aufgefangen werden. Zudem konnte durch den modularen Ansatz und die

Nutzung von open-source-Produkten die Herstellerabhängigkeit reduziert werden. Die dadurch gewonnenen Freiheitsgrade werden als wichtiger eingestuft als eine etwaige Komplettunterstützung nach den Vorgaben eines Anbieters. Als besonders empfehlenswert hat es sich erwiesen, die vorhandenen Datenquellen sukzessive anzubinden und den ETL-Prozess pro Datenquelle in Sub-Prozesse zu gliedern. So wird die Komplexität der Einführung stark reduziert, was insbesondere die Fachabteilungen (neben dem regulären Tagesgeschäft) entlastet.

Die in diesem Beitrag beschriebene Vorgehensweise ist insbesondere für Hochschulen mit einer starken Entwicklungsabteilung geeignet, sowie für Domänen die eine Separierung in einzelne Prozesse und Datenstrukturen erlauben.

Literaturverzeichnis

- [BStatG87] Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG), 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 13 des Gesetzes vom 25. Juli 2013 (BGBl. I S. 2749). http://www.gesetze-im-internet.de/bstatg_1987/
- [Cla98] Clausen, N.: OLAP – Multidimensionale Datenbanken. Addison-Wesley-Longman, Bonn 1998.
- [CMR02] Colossi, N.; Malloy, W.; Reinwald, B.: Relational extensions for OLAP, IBM Systems Journal 41(4), 2002, S. 714 - 731.
- [Ka13] Kaiser, S.; Kuhnt, E.; Lemcke, S.; Lucke, U.: Web-basierte Beschaffung. In Proc. INFORMATIK 2013, LNI P-220, Bonn : Köllen, 2013, S. 308-319.
- [KR13] Kimball, R.; Ross, M.: The Data Warehouse Toolkit – The Definitive Guide to Dimensional Modeling (3. Auflage), Wiley 2013.
- [Le04] Leszczensky, M.; Orr, D.; Schwarzenberger, A.; Weitz, B.: Staatliche Hochschulsteuerung durch Budgetierung und Qualitätssicherung, in: Hochschulplanung (167), HIS Hochschulinformationssysteme GmbH, Hannover 2004.
- [Me12] Mertens, P.: Schwierigkeiten mit IT-Projekten der öffentlichen Verwaltung. Informatik-Spektrum, 35/06, S. 433-446, 2012.
- [SG15] Springer Gabler Verlag (Herausgeber); Gabler Wirtschaftslexikon; Stichwort: Business Intelligence. <http://wirtschaftslexikon.gabler.de/Archiv/75968/business-intelligence-v10.html>
- [SBU99] Sinz, E.J.; Böhnlein, M.; Ulbrich-vom Ende, A.: Konzeption eines Data Warehouse-Systems für Hochschulen, in: Proc. Informatik '99 - Informatik überwindet Grenzen, 29. Jahrestagung der Gesellschaft für Informatik, Springer 1999, S. 111-124.
- [Vo14] Vogel, D.; Löbel, St.; Proeller, I.; Schuppan, T.: Einflussfaktoren von Führungsverhalten in der öffentlichen Verwaltung. der moderne staat, 07/02, 2014, S. 459-478.