

# Data-Warehouse als Basis für Verwaltungsprozesse

Ulrike Lucke

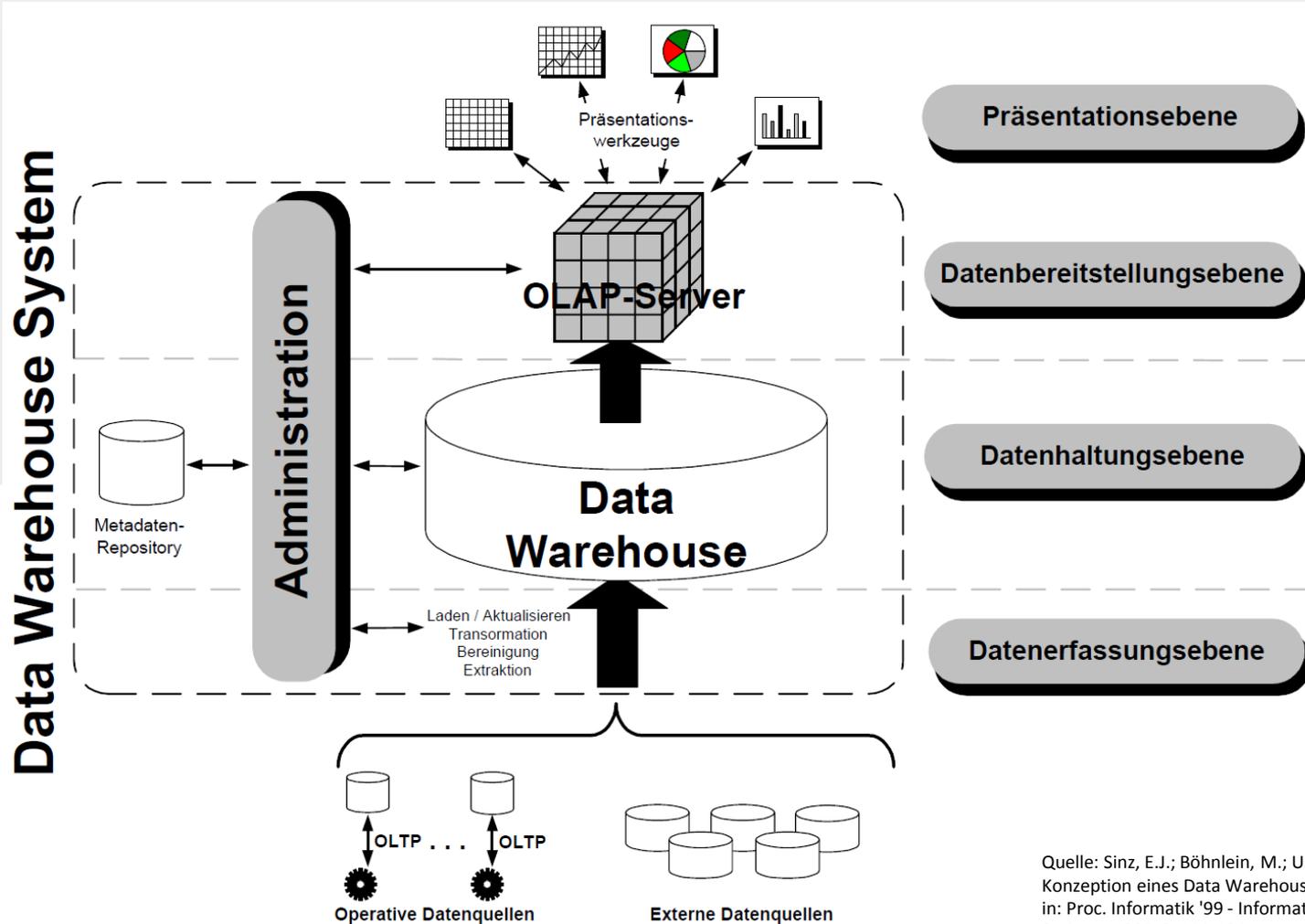
Thomas Jankowski

# Agenda

1. Data-Warehouse
2. Ausgangssituation
  - a) heterogene Systemlandschaft
  - b) Stakeholder
3. Ziele
4. Umsetzung
  - a) Systemauswahl
  - b) Systemüberblick (funktionale Sicht)
  - c) Datenerfassung (ETL-Tool)
  - d) Datenhaltung (Datenbank und Metadatenverwaltung)
  - e) Präsentationsebene (Analyse-Tool)
5. Unterstützung von Verwaltungsprozessen
6. Erfahrungen
7. Aufwände
8. Fazit

# Data-Warehouse-System (DWHS)

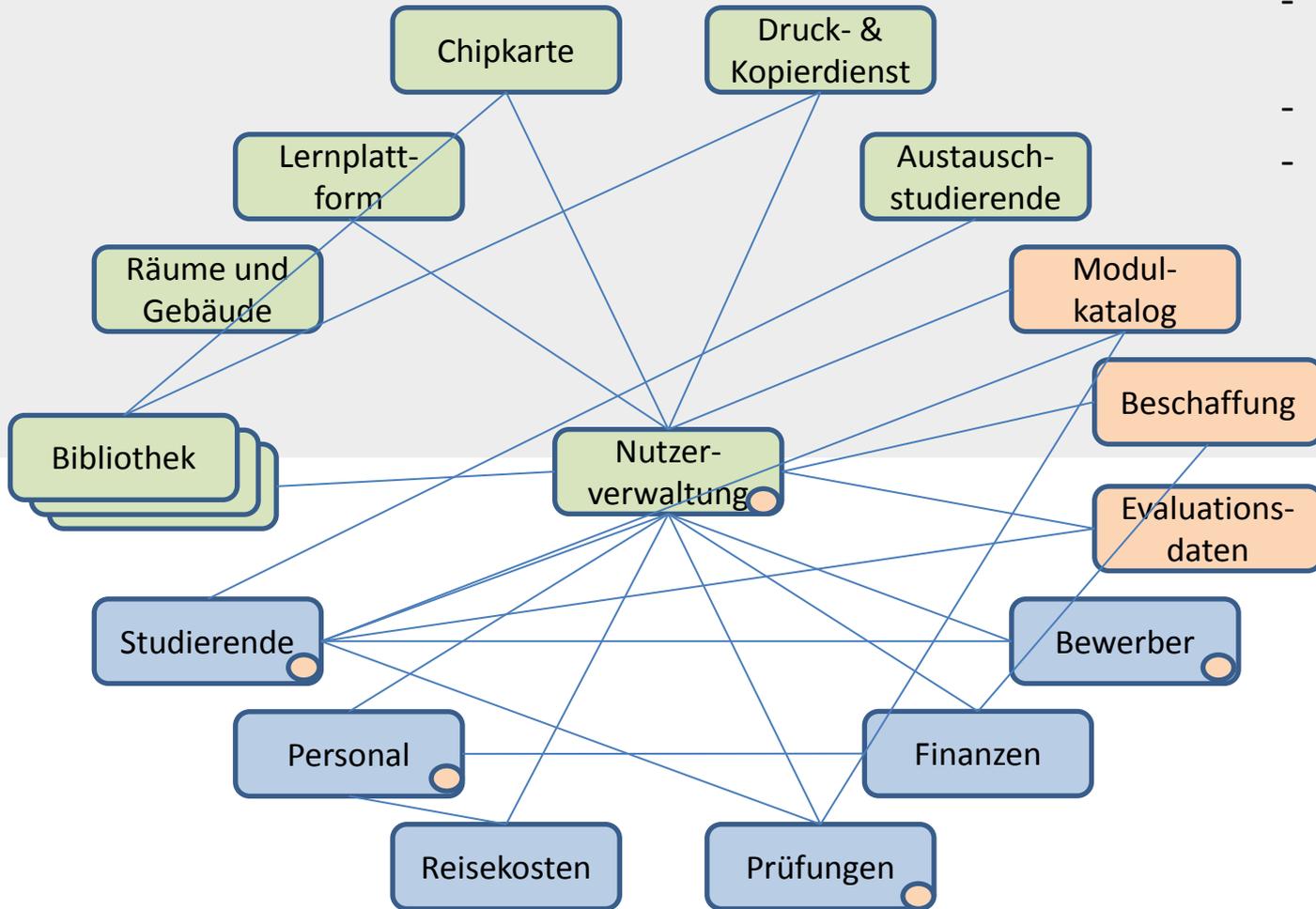
Kurzeinführung



Quelle: Sinz, E.J.; Böhnlein, M.; Ulbrich-vom Ende, A.:  
Konzeption eines Data Warehouse-Systems für Hochschulen,  
in: Proc. Informatik '99 - Informatik überwindet Grenzen,  
Springer 1999, S. 111-124

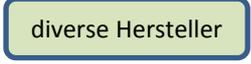
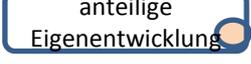
# Ausgangssituation

heterogene IT-Systemlandschaft



- sehr heterogene IT-Landschaft
- diverse Schnittstellen
- umfangreiche Eigenentwicklungen

**Legende:**

-  HIS eG
-  Eigenentwicklung
-  diverse Hersteller
-  anteilige Eigenentwicklung

# Ausgangssituation

## Stakeholder

Die nachfolgende Sortierung gibt gleichzeitig auch an, welche Bereiche einen höheren Bedarf sehen als andere.

- Dezernat für Planung und Steuerung -> detaillierte Fachdaten
- Mitarbeiter der Verwaltung  
(da operative Systeme teilweise Statistikbedarf für tägliche Aufgaben nicht decken -> detaillierte Fachdaten)
- Fakultätsleitungen (Managementberichte und detaillierte Fachdaten)
- Universitätsleitung (Managementberichte und detaillierte Fachdaten)
- Externe Adressaten (detaillierte Fachdaten)
- Universitäts-Öffentlichkeit

# Ziele

## Anforderungen an das DWHS

- Import der Daten in das Data-Warehouse (ETL-Prozess)
  - automatisierbar
  - hochschuleigene Plausibilitäten
  - Datenanpassung
  - Reaktion auf Änderung in Datenquellen muss möglich sein
- Verwaltung der Daten und Berichte
  - Historisierung
  - Metadaten
  - Skalierbar (Performance)
  - (zentrale) Rechteverwaltung
  - automatischer Berichtsversand
- Berichte
  - Web-Anwendung
  - Nutzer kann Filter und Sortierung beeinflussen
  - Komplexe Berichte  
(Kombination mehrerer Berichte, Diagramme und Freitexte, ergänzt um Inhaltsverzeichnis)
  - Datenexport (PDF, Word, Excel)
  - dezentrale Selbstbedienung bei gleichzeitig zentraler Qualitätskontrolle



# Umsetzung

Systemauswahl (Frühjahr 2014)



Kumulierte Darstellung der Testergebnisse von kostenfreien Open-Source-Produkten

Kriteriengruppe	Gewicht	Pentaho	JasperSoft	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%
<b>Gesamtbewertung</b>		<b>73,25%</b>	<b>72,48%</b>	<b>63,16%</b>

Entscheidung: Mischform

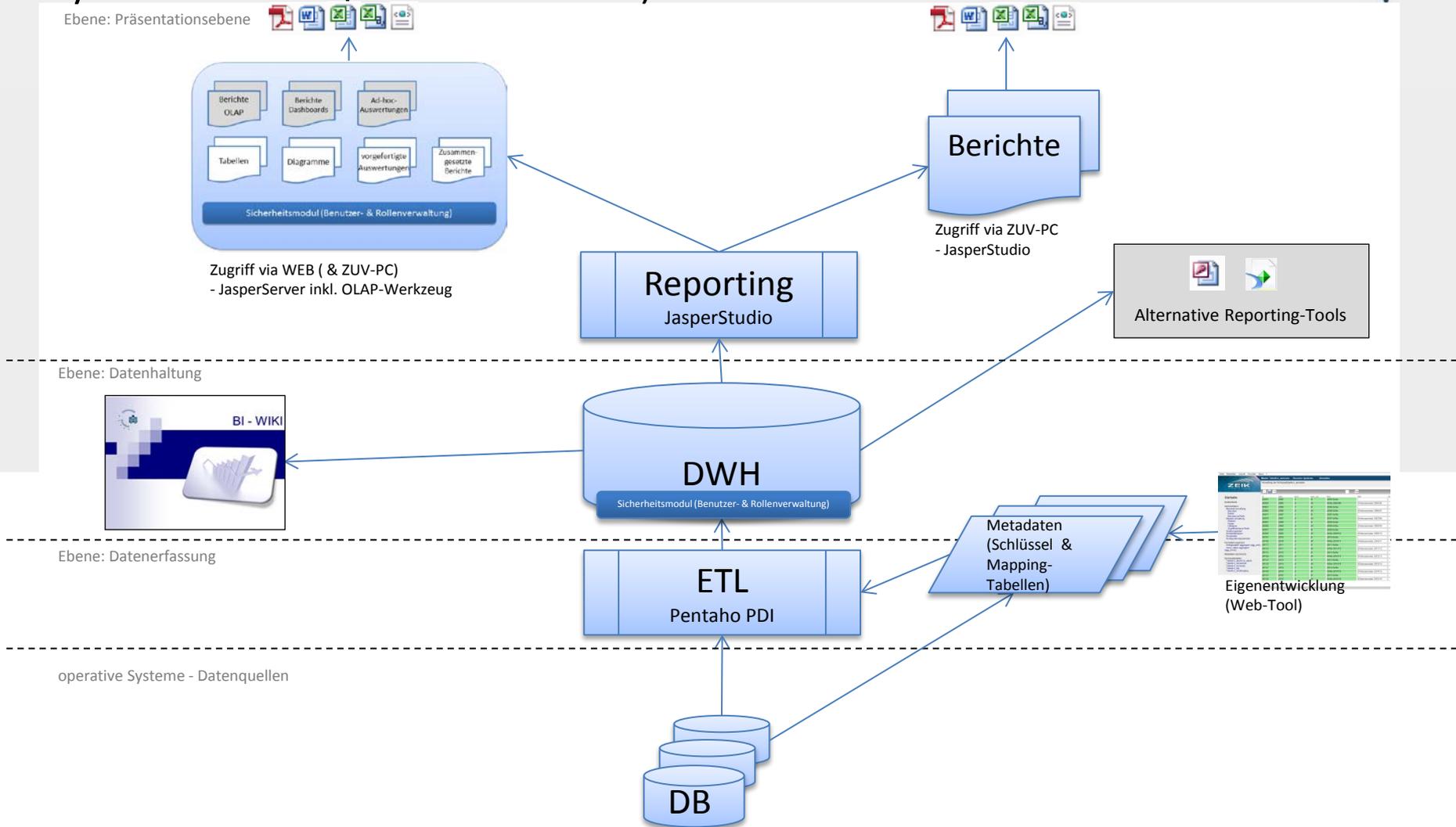
Pentaho -> Datenerfassungsebene (in der Datenbereitstellungsebene traten technische Probleme auf, die sich ohne Support nicht lösen ließen)

Jaspersoft -> Datenpräsentationsebene



# Umsetzung

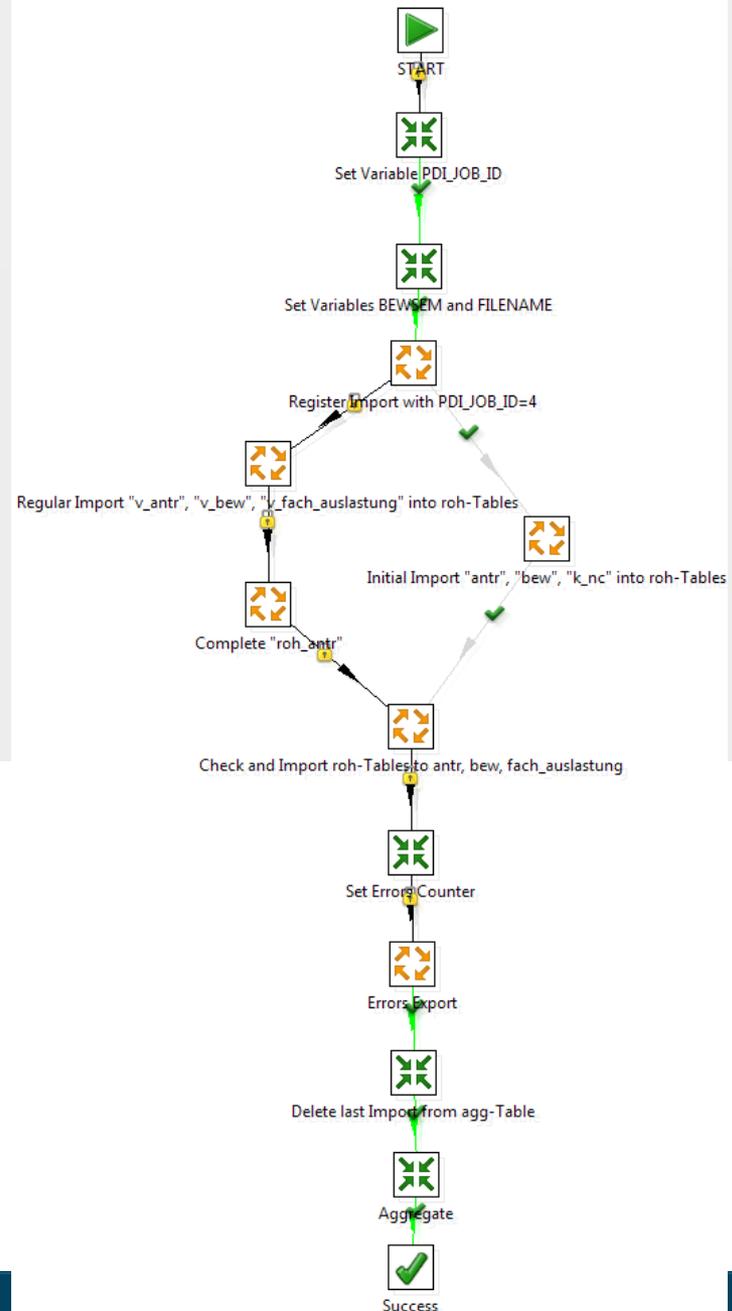
## Systemüberblick (funktionale Sicht)



# Umsetzung

## Datenerfassung (ETL-Tool)

- grafischer Editor
- prozessorientiert
- übersichtlich durch Subprozesse
- Plausibilitätsprüfung
- automatische Fehlerkorrekturen (soweit möglich)
- Schlüssel-Mapping
- Prozesskonstanten und -variablen festlegen (z.B. Zeitstempel als Importdatum)
- dynamische Wahl des ETL-Prozess-Pfades in Abhängigkeit von den zu importierenden Daten
- Gruppierung und Aggregation von Daten
- Ergänzung neuer Merkmale nach festgelegten Regeln
- Fehlerbehandlung
- automatische zeitgesteuerte Ausführung



# Umsetzung

Datenhaltung (Datenbank und Metadatenverwaltung)

## Datenbank

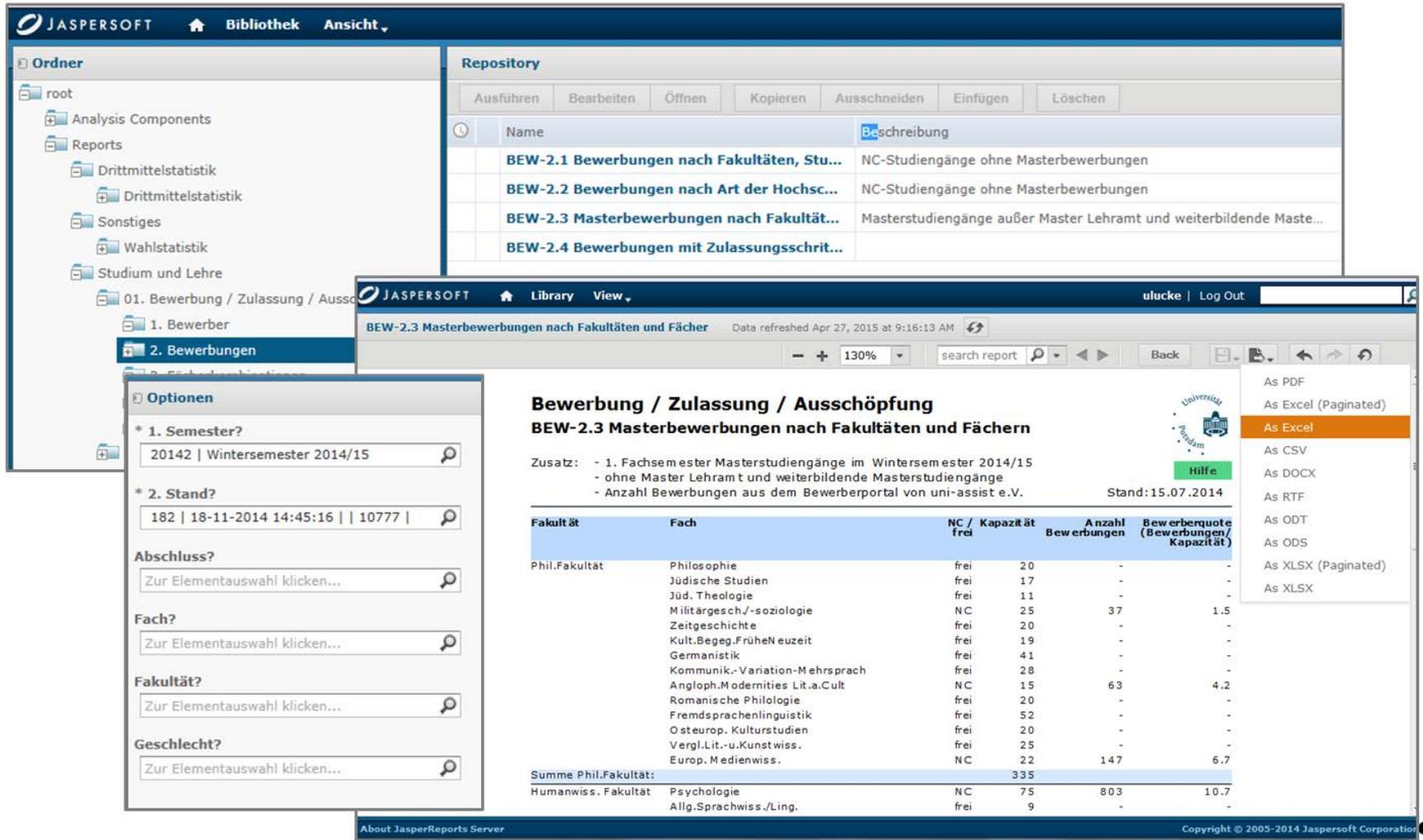
vorhandener Postgres-Cluster

## Metadatenverwaltung

- Wiki
  - Erläuterungen zu Berichten
  - Beschreibungen der Datenquellen
  - Häufig gestellte Fragen
  - Ansprechpartner
  - Informationen zu den ETL-Prozessen
- Webanwendung für Dateneditierung in einer beliebigen Datenbank
  - dezentrale Zugriffsrechte
  - Verwaltung von Schlüssel- und Mappingtabellen durch fachlich zuständiges Personal
  - abweichende Gruppierungs- und Sortierungsvorgaben
  - Technische Metadaten (Log-Daten, Fehlerprotokolle, Historisierungsschalter)
  - Verwaltung von Daten, für die der Aufbau einer eigenen Anwendung aus Kosten-Nutzen-Aspekten nicht sinnvoll erscheint.

# Umsetzung

## Präsentationsebene (Analysetool)



The screenshot displays the JasperReports web interface. The main window shows a report titled "Bewerbung / Zulassung / Ausschöpfung BEW-2.3 Masterbewerbungen nach Fakultäten und Fächern". The report includes a table with columns for Faculty (Fakultät), Subject (Fach), and various statistics (NC/Kapazität, Anzahl Bewerbungen, Bewerberquote). A context menu is open over the table, showing options like "As PDF", "As Excel (Paginated)", and "As Excel".

**Repository**

Name	Beschreibung
BEW-2.1 Bewerbungen nach Fakultäten, Stu...	NC-Studiengänge ohne Masterbewerbungen
BEW-2.2 Bewerbungen nach Art der Hochsc...	NC-Studiengänge ohne Masterbewerbungen
BEW-2.3 Masterbewerbungen nach Fakultät...	Masterstudiengänge außer Master Lehramt und weiterbildende Maste...
BEW-2.4 Bewerbungen mit Zulassungsschrit...	

**Options**

- \* 1. Semester?  
20142 | Wintersemester 2014/15
- \* 2. Stand?  
182 | 18-11-2014 14:45:16 | | 10777 |
- Abschluss?  
Zur Elementauswahl klicken...
- Fach?  
Zur Elementauswahl klicken...
- Fakultät?  
Zur Elementauswahl klicken...
- Geschlecht?  
Zur Elementauswahl klicken...

**Bewerbung / Zulassung / Ausschöpfung**  
**BEW-2.3 Masterbewerbungen nach Fakultäten und Fächern**

Zusatz: - 1. Fachsemester Masterstudiengänge im Wintersemester 2014/15  
- ohne Master Lehramt und weiterbildende Masterstudiengänge  
- Anzahl Bewerbungen aus dem Bewerberportal von uni-assist e.V. Stand: 15.07.2014

Fakultät	Fach	NC / Kapazität frei	Anzahl Bewerbungen	Bewerberquote (Bewerbungen/ Kapazität)	
Phil.Fakultät	Philosophie	frei	20	-	
	Jüdische Studien	frei	17	-	
	Jüd. Theologie	frei	11	-	
	Militärgesch./-soziologie	NC	25	37	1.5
	Zeitgeschichte	frei	20	-	-
	Kult.Begeg.Frühe Neuzeit	frei	19	-	-
	Germanistik	frei	41	-	-
	Kommunik.-Variation-Mehrsprach	frei	28	-	-
	Angloph.Modernities Lit.a.Cult	NC	15	63	4.2
	Romanische Philologie	frei	20	-	-
	Fremdsprachenlinguistik	frei	52	-	-
Osteurop. Kulturstudien	frei	20	-	-	
Vergl.Lit.-u.Kunstwiss.	frei	25	-	-	
Europ. Medienwiss.	NC	22	147	6.7	
<b>Summe Phil.Fakultät:</b>			<b>335</b>		
Humanwiss. Fakultät	Psychologie	NC	75	803	10.7
	Allg.Sprachwiss./Ling.	frei	9	-	-

# Umsetzung

Präsentationsebene (Analysetool)

- keine Besonderheit im Vergleich zu anderen Analysetools (Community-Version)
- eher schlicht, aber funktional in der Community-Version
- Tabellen, Texte, Filter, Diagramme, Links, mehrere Export-Formate, OLAP
- Rechteverwaltung bis auf Berichtsebene -> kann delegiert werden
- Rechteverwaltung auf Datensatzebene -> Eigenentwicklung, muss in den SQL-Abfragen der Berichte integriert werden
- Zugriff für anonyme Nutzer
- Integration in andere Webseiten
- in der Community-Version fehlt der Adhoc-Designer

# Unterstützung von Verwaltungsprozessen

- Planung (Bsp.: Studienplatzauslastung)
- Steuerung während der Bewerbungsphase (Bsp.: Überbuchung)
- übergreifende Prozesse für Pflege von Schlüsseltabellen unterstützen (Bsp.: Kostenstellen)
- Überblick über IT-Landschaft verbessern

# Erfahrungen

- Anforderungen ändern sich im Laufe des Prozesses der Statistikbereitstellung -> es sind mehrere Iterationen notwendig
- I.d.R. ist es am Ende einfacher, als am Anfang gedacht.
- Fachdezernate erkennen schnell Vorteile für die eigene Arbeit -> Daraus folgen jedoch neue Anforderungen ;-)
- Das Schnittstellenkonzept ist pro Datenquelle der aufwendigste und schwierigste Teil
- Diese Variante (heterogene IT-Landschaft + unabhängiges DWHS mit leistungsstarken ETL-Tool) führt relativ schnell zu Erfolgen, da sukzessive vorgegangen werden kann.

# Aufwände

## IT-Kosten

- 1x virtueller Server mit 2 CPU + 8 GB RAM reicht derzeit aus
- keine Lizenzkosten, solange Community-Version ausreicht (kein Adhoc-Designer, kein animiertes management-Cockpit)
- Für Schulung und Support sollten in den ersten drei Jahren jeweils ca. 5 PT für eine externe Firma eingeplant werden.

## Personalkosten

- Einführungsphase: eine Vollzeitstelle für 2 Jahre (vertiefte SQL-Kenntnisse)
- Betriebsphase: voraussichtlich eine Halbtagsstelle (dauerhaft)

# Fazit

- Community-Version von Open-Source-Systemen sind geeignet
- sehr leistungsstark in den Ebenen Datenerhebung und Datenhaltung
- ausreichender Funktionsumfang in der Präsentationsebene
- Konzept ist modular, sodass pro Ebene auch andere Systeme eingesetzt werden können.
- Der Wechsel auf eine lizenzpflichtige Version mit mehr Funktionsumfang ist insbesondere in der Präsentationsebene möglich.

# Fragen



Prof. Dr.-Ing. habil. Ulrike Lucke

---

Universität Potsdam

Chief Information Officer

[cio@uni-potsdam.de](mailto:cio@uni-potsdam.de)

Tamara Wolowelsky

---

Universität Potsdam

ZEIK - Projekt CaMS

(Teilprojekt Data-Warehouse)

[tamara.wolowelsky@uni-potsdam.de](mailto:tamara.wolowelsky@uni-potsdam.de)

Thomas Jankowski

---

Universität Potsdam

ZEIK - Projektleiter CaMS

[thomas.jankowski@uni-potsdam.de](mailto:thomas.jankowski@uni-potsdam.de)