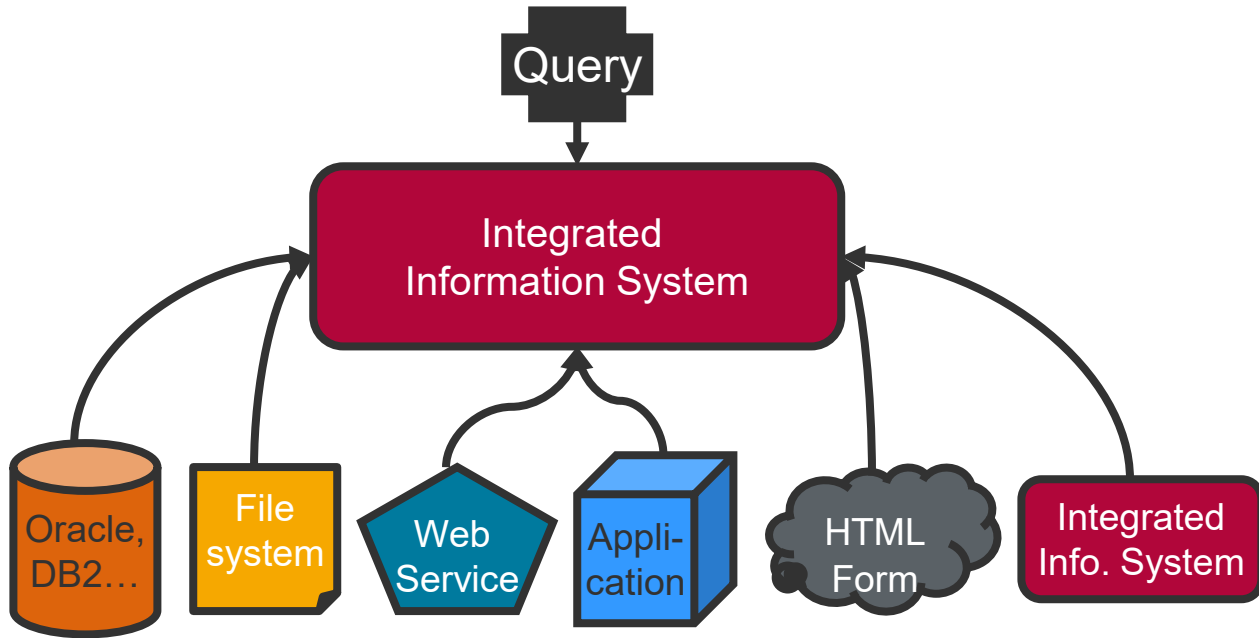




Information Integration Introduction

21.10.2019
Felix Naumann

Integrated Information Systems



Felix Naumann
Information Integration
Winter 2019/20

- 1. Introduction of research group**
2. Organisation
3. Information systems
4. Information integration by example
5. Overview of semester



Felix Naumann
Information Integration
Winter 2019/20

Information Systems Team



Dr. Thorsten Papenbrock



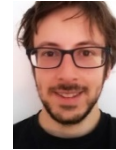
Diana Stephan



Prof. Felix Naumann



Dr. Ralf Krestel



Leon Bornemann



Hazar Harmouch



Phillip Wenig



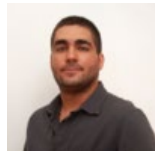
Julian Risch



Konstantina Lazaridou



Michael Loster



John Koumarelas



Tobias Bleifuß



Nitisha Jain



Lan Jiang



Gerardo Vitagliano



Tim Repke

project **DuDe**
Data Change **Data Fusion** **Duplicate Detection**
Data Profiling **Information Integration** **Entity Search**
 project **DataChEx** **Data Scrubbing** **project Stratosphere** **project DataKnoller** **Web Science**
Information Quality **Data Cleansing** **Data as a Service** **Text Mining**
Web Data **Dependency Detection** **Linked Open Data** **RDF Data Mining**
Service-Oriented Systems **ETL Management** **project Janus**
Entity Recognition **Opinion Mining** **Data Preparation**
Change Exploration

Other courses in this semester

Bachelor

- Datenbanksysteme II
- Bachelorprojekt: Data Analytics – Museumserlebnisse mit Datenanalyse opt

Bachelor & Master

- Data Engineering in Practice - Ringvorlesung (VL)
- Paint it Black: Ethical Data Analytics (S)

Master

- Information Integration (VL)
- Distributed Data Management (VL)
- Machine Learning for Data Streams (S)
- Masterproject: Genealogy of Natural Language



Extending the Database Relational Model to Capture More Meaning

E. F. CODD
IBM Research Laboratory

During the last three or four years several investigations have been exploring "semantic models" for formatted databases. The intent is to capture (in a more or less formal way) more of the meaning of the data so that database design can become more systematic and the database system itself can behave more intelligently. Two major themes are clear:

- (1) the search for meaningful units that are as small as possible—atomic semantics;
- (2) the search for meaningful units that are larger than the usual n -ary relation—molecular semantics.

In this paper we propose extensions to the relational model to support certain atomic and molecular semantics. These extensions represent a synthesis of many ideas from the published work in semantic modeling plus the introduction of new rules for insertion, update, and deletion, as well as new algebraic operations.

Key Words and Phrases: relation, relational database, relational model, relational schema, database, basic, conceptual model, conceptual schema, semantic model, knowledge representation, knowledge, CK Categories: 3.70, 3.73, 4.22, 4.25, 4.33, 4.34, 4.39

1. INTRODUCTION

The relational model for formatted databases [5] was conceived ten years ago, primarily as a tool to free users from the frustrations of having to deal with the clutter of storage representation details. This implementation independence coupled with the power of the algebraic operators on n -ary relations and the open questions concerning dependencies (functional, multivalued, and join) within and between relations have stimulated research in database management (see [30]). The relational model has also provided an architectural focus for the design of MACAIMS [13], PRTV [38], RDMS(GM) [41], MAGNUM [19], INGRES [37], QBE [46], and System R [2].

During the last few years numerous investigations have been aimed at capturing Permission to copy without fee all or part of this material is granted provided that the copiers are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

A version of this work was presented at the 1979 International Conference on Management of Data (SIGMOD), Boston, Mass., May 30–June 1, 1979.

Author's address: IBM Research Laboratory K01/282, 5600 Cottle Road, San Jose, CA 95183.
© 1979 ACM 0360-0915/79/1280-0297 \$00.75

ACM Transactions on Database Systems, Vol. 4, No. 4, December 1979, Pages 397–434.

Felix Naumann
Information Integration
Winter 2019/20

Overview

1. Introduction of research group
- 2. Organisation**
3. Information systems
4. Information integration by example
5. Overview of semester



Felix Naumann
Information Integration
Winter 2019/20

Dates and Examination

- Lectures
 - Mondays 9:15 – 10:45 in HS2
 - Thursdays 9:15 – 10:45 in HS1
- Exercises
 - Throughout semester
- First lecture
 - 21.10.2019
- Last lecture
 - 6.2.2020
- Holidays
 - 31.10. Reformationstag
 - Christmas break
- Exan
 - Written exam, 2 hours
 - Soon after teaching period
- Prerequisites
 - For participation
 - Database knowledge (e.g. DBS I)
 - For exam
 - Attendance of lectures
 - Active participation in exercises
 - Short presentation(s)
 - „Passing“ exercises

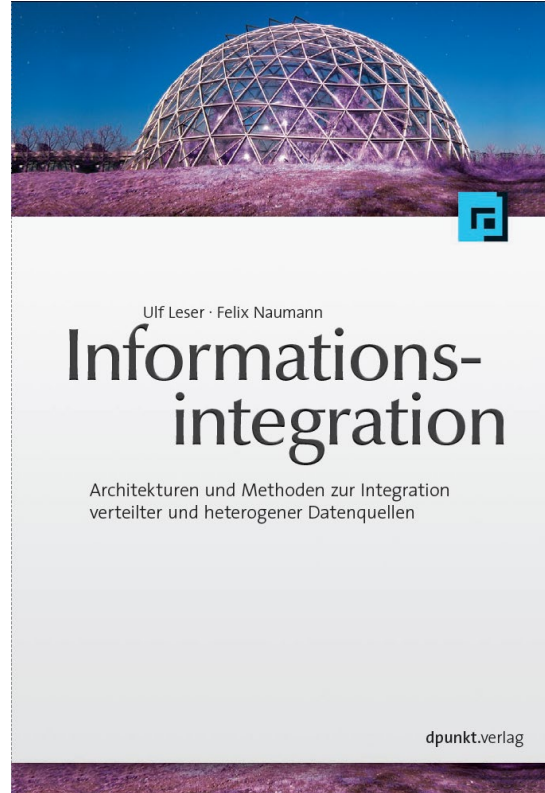
Feedback

- Evaluation at end of semester
- Questions anytime!
 - During lecture
 - Email: felix.naumann@hpi.de
 - Open door, building F, 2nd floor
- Suggestions for improvement:
 - Slides and material
 - Infos on webpage
 - ...
 - Or via email: felix.naumann@hpi.de

Felix Naumann
Information Integration
Winter 2019/20

Course books

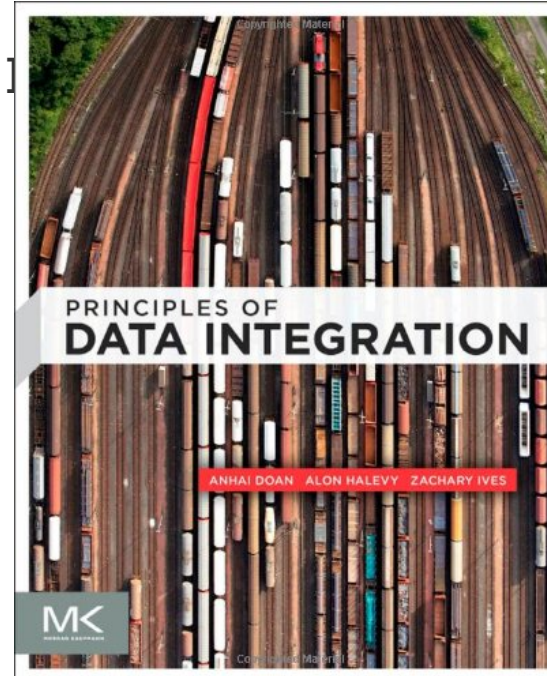
- Informationsintegration
- Ulf Leser und Felix Naumann
 - dpunkt Verlag, 2006
 - Amazon or free [pdf](#)
- n -mal in Bibliothek



Felix Naumann
Information Integration
Winter 2019/20

Course books

- Principles of Data Integration
- AnHai Doan, Alon Halevy, Zachary Ives
- Morgan Kaufmann, 2012
- € 62,-



Felix Naumann
Information Integration
Winter 2019/20

Further literature

- Topics from
 - Föderierte Datenbanksysteme. Konzepte der Datenintegration
Stefan Conrad, 1997, ISBN: 3540631763
 - Principles of Distributed Database Systems
M. Tamer Özsu, Patrick Valduriez
ISBN: 0136597076
- Further links to literature during lectures
- All mentioned articles can be obtained via email from me or:
 - Materialien-directory
 - Google Scholar: <http://scholar.google.com/>
 - DBLP: <http://dblp.uni-trier.de/db/index.html>
 - CiteSeer: <http://citeseer.ist.psu.edu/>
 - ACM Digital Library: www.acm.org/dl/
 - Homepages of authors

Felix Naumann
Information Integration
Winter 2019/20

TBD

Felix Naumann
Information Integration
Winter 2019/20

Your introduction

- Which year / semester?
- HPI or IfI or others?
- Erasmus or other exchange?
 - English?
- Database background / experience?
 - Data Profiling and Data Cleansing?
 - Other relevant courses?
- Your motivation?
 - Have you ever integrated data?
 - DWH?

Felix Naumann
Information Integration
Winter 2019/20

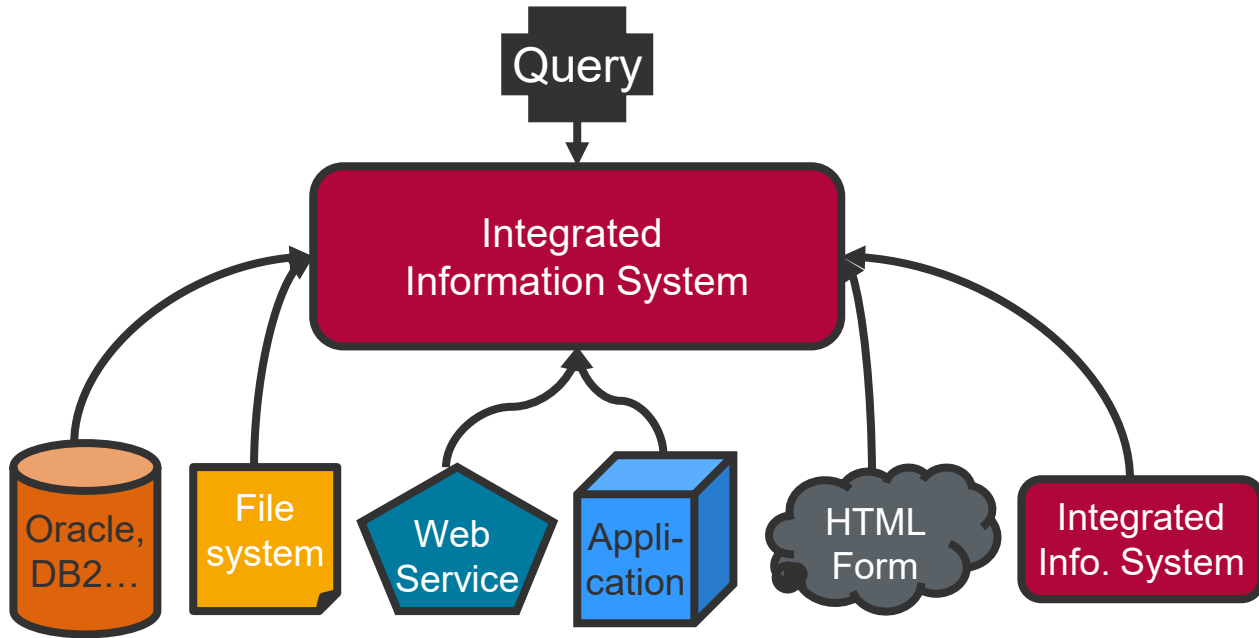
Overview

1. Introduction of research group
2. Organisation
- 3. Information systems**
4. Information integration by example
5. Overview of semester

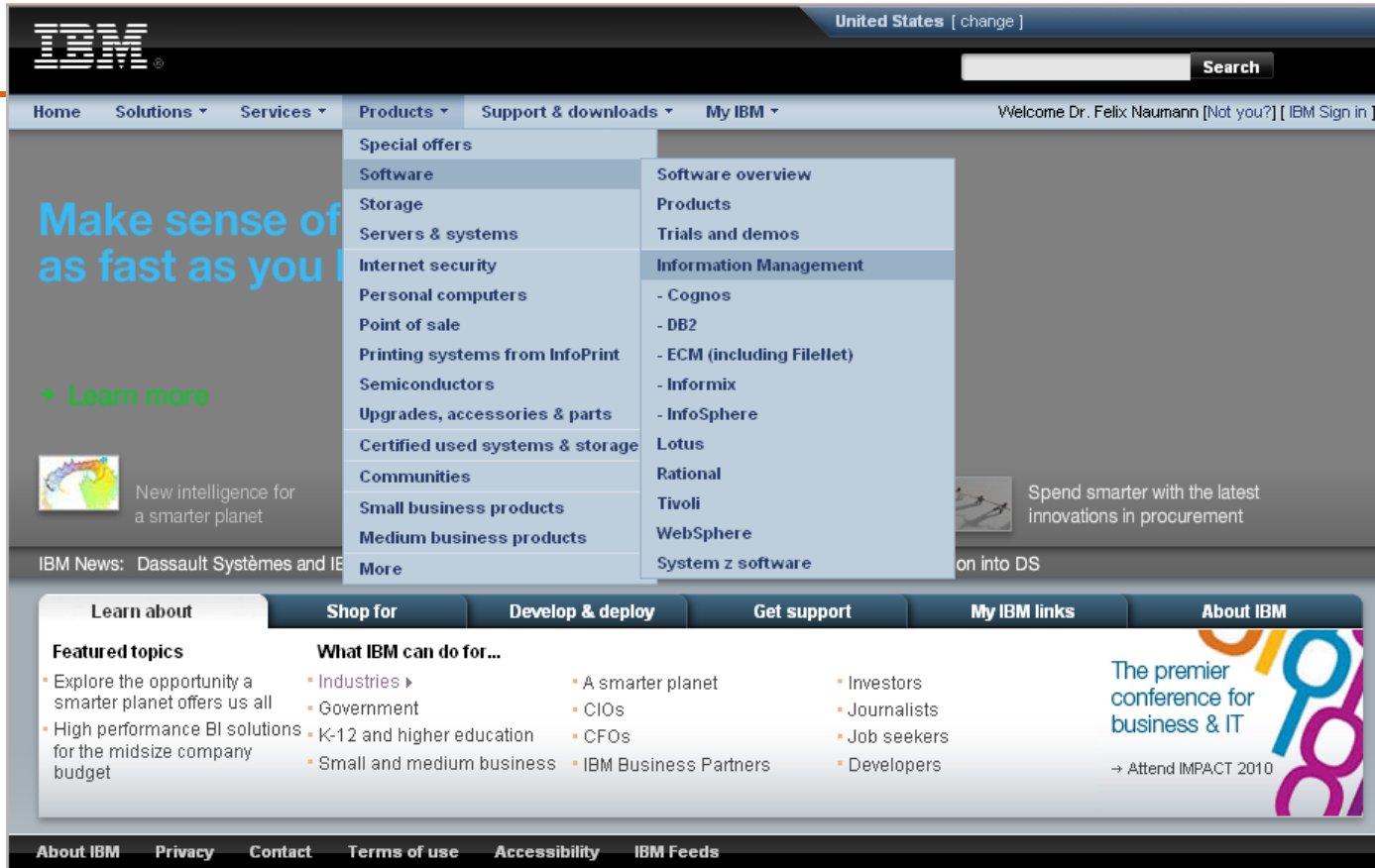


Felix Naumann
Information Integration
Winter 2019/20

Integrated Information Systems



Felix Naumann
Information Integration
Winter 2019/20



The screenshot shows the IBM website interface. At the top left is the IBM logo. To its right, the text "United States [change]" is visible. A search bar with a "Search" button is located in the top right. Below the logo, a navigation bar contains links for "Home", "Solutions", "Services", "Products", "Support & downloads", and "My IBM". A user greeting "Welcome Dr. Felix Naumann [Not you?] [IBM Sign in]" is displayed on the right side of the navigation bar.

The "Products" menu is open, showing a list of categories: "Special offers", "Software", "Storage", "Servers & systems", "Internet security", "Personal computers", "Point of sale", "Printing systems from InfoPrint", "Semiconductors", "Upgrades, accessories & parts", "Certified used systems & storage", "Communities", "Small business products", "Medium business products", and "More". A sub-menu for "Software" is also open, listing: "Software overview", "Products", "Trials and demos", "Information Management" (which is highlighted), "Lotus", "Rational", "Tivoli", "WebSphere", and "System z software". Under "Information Management", there are sub-items: "- Cognos", "- DB2", "- ECM (including FileNet)", "- Informix", and "- InfoSphere".

On the left side of the page, there is a large blue heading "Make sense of as fast as you" with a green arrow pointing to "Learn more". Below this is a small globe icon and the text "New intelligence for a smarter planet".

At the bottom of the page, there is a navigation bar with tabs: "Learn about", "Shop for", "Develop & deploy", "Get support", "My IBM links", and "About IBM". Below these tabs, there are sections for "Featured topics", "What IBM can do for...", and a promotional banner for "The premier conference for business & IT" with the text "Attend IMPACT 2010".

Felix Naumann
Information Integration
Winter 2019/20

Information Management

Products

Services

Downloads

Library

Case studies

News

Training and certification

Events

Support

Communities

- IBM Business Partners
- Information Management
- Data Management
- Enterprise Content Management
- InfoSphere
- ISVs
- DeveloperWorks
- IBM Information Management User Groups
- Information Management Early Software Programs

Software > Information Management >

Information Management



Optimizing With Information Governance

[→ Learn more](#)


Information Agenda Catalog

A new directory of key solutions

[→ Learn more](#)


Smart Archive Strategy

[→ Learn more](#)

Why Information Management Solutions

IBM Information Management solutions deliver trusted information throughout your information supply chain and help you analyze your information to gain insights, identify breakdowns, and make better decisions that will optimize your business. Find [Information Management products](#) by category, by A-Z, or by keyword search.

What we offer

[Data Management](#)

Manage business data and support operational and analytic applications with powerful data servers.

[Cognos Business Intelligence and Performance Management](#)

Create and administer reports, scorecards, plans & forecasts with data from multiple sources.

[Express Middleware](#)

Easy to install, fast to deploy products for small and medium businesses.

[Enterprise Content Management](#)

Helps organizations managing content, optimize business processes and enable compliance.

[Information Integration, Master Data Mgmt & Data Warehousing](#)

Real-time access to business information and an enterprise-wide view of their critical business data.

Information & Analytics



EMEA Conference, 19 - 21 May 2010, Rome

[Register today](#)
[→ Sneak Preview of IOD EMEA Conference 2010 Agenda](#)

IBM Cognos Finance Forum 2010



Begins in February and travels to 75+ cities around the world.

[→ Learn more and register today!](#)

Felix Naumann
Information Integration
Winter 2019/20

InfoSphere Software

- Data Warehousing & Analytics
- Information Integration
- Master Data Management
- Data & Process Models for your Industry
- Cloud Computing
- Information Infrastructure
- Industries
- Products
- Services
- Success stories
- Library
- Events

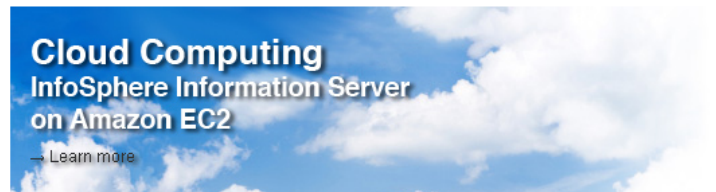
Related links

- InfoSphere Foundation Tools
- IBM Business Partners
- InfoSphere Videos
- Information On Demand
- End of Service Products

Software > Information Management >

Information Integration, Data Warehouse and Master Data Management

InfoSphere software for trusted information



→ Learn more

Why IBM for Trusted Information

IBM delivers market-leading solutions to critical information intensive business problems allowing customers to achieve new levels of innovation through best of breed information integration, data warehouse and MDM capabilities.

What we offer

Data Warehousing & Analytics

- IBM InfoSphere Warehouse
- IBM Smart Analytics System
- IBM InfoSphere Identity Insight Solutions
- IBM InfoSphere Streams

→ View all

Information Integration

- IBM InfoSphere Information Server
- IBM InfoSphere DataStage
- IBM InfoSphere Change Data Capture
- IBM InfoSphere Replication Server
- IBM InfoSphere Global Name Recognition

→ View all

Master Data Management

- IBM InfoSphere Master Data Management Server
- IBM InfoSphere Master Data Management Server for Product Information Management

→ View all

Data & Process Models for Your Industry

- IBM Banking Data Warehouse
- IBM Financial Markets Data Warehouse
- IBM Health Plan Data Model

→ View all

White papers

- IBM Master Data Management strategy: Leveraging critical data to accomplish strategic business objectives
- IBM InfoSphere Information Server: Simplify integration with unified metadata
- Trusted Information with Performance Management

Information & Analytics



EMEA Conference, 19 - 21 May 2010, Rome

Register today

→ Sneak Preview of IOD EMEA Conference 2010 Agenda

Videos



This self-guided video tour allows you to explore the unique components of InfoSphere Software

→ Start your tour

Enterprise Master Data

Felix Naumann
Information Integration
Winter 2019/20

Related links

- InfoSphere Foundation Tools
- IBM Business Partners
- InfoSphere Videos
- Information On Demand
- End of Service Products

IBM Information Integration is the leader in data integration. Our platform integrates and transforms data improving productivity, flexibility and performance, so you have the right information for your business.

What we offer

Capabilities

Information integration is the process of integrating and transforming data and content to deliver authoritative, consistent, timely and complete information, and governing its quality throughout its life cycle.

- Connect to information
- Understand information
- Cleanse information
- Transform information
- Deliver information
- Replicate data
- View all capabilities

Products

- IBM InfoSphere Information Server
- IBM InfoSphere DataStage
- IBM InfoSphere Change Data Capture
- IBM InfoSphere Federation Server
- IBM InfoSphere QualityStage
- IBM InfoSphere Data Architect
- IBM InfoSphere Replication Server
- IBM InfoSphere Information Analyzer

All Information Integration products

- Select one
- Select one
 - IBM InfoSphere Business Glossary
 - IBM InfoSphere Change Data Capture
 - IBM InfoSphere Change Data Capture for z/OS
 - IBM InfoSphere Change Data Capture for Oracle Replication
 - IBM InfoSphere Classic Connector for z/OS
 - IBM InfoSphere Classic Data Event Publisher for z/OS
 - IBM InfoSphere Classic Federation Server for z/OS
 - IBM InfoSphere Classic Replication Server for z/OS
 - IBM InfoSphere Data Architect
 - IBM InfoSphere Data Event Publisher
 - IBM InfoSphere Data Event Publisher for z/OS
 - IBM InfoSphere DataStage
 - IBM InfoSphere Federation Server
 - IBM InfoSphere Global Name
 - IBM InfoSphere Information Analyzer
 - IBM InfoSphere Information Server
 - IBM InfoSphere Information Server for System z
 - IBM InfoSphere Information Server Blade
 - IBM InfoSphere Information Services Director

→ Forrester Research: Total Economic Impact of IBM InfoSphere Foundation Tools

Information & Analytics

IBM Information ON Demand 2010
EMEA Conference, 19 - 21 May 2010, Rome

Register today

→ Sneak Preview of IOD EMEA Conference 2010 Agenda

Videos



This self-guided video tour allows you to explore the unique components of InfoSphere Software

→ Start your tour

Datasheets

- IBM InfoSphere Information Server: Understanding and...
- IBM InfoSphere Information Server: Access to your enterprise data...
- IBM InfoSphere Information Server: Use business...
- IBM InfoSphere Information Server: Transforming and...


- InfoSphere Foundation Tools
- IBM Business Partners
- InfoSphere Videos
- Information On Demand
- End of Service Products

What we offer	
<p>IBM InfoSphere Metadata Workbench</p> <p>Provides end-to-end metadata management, depicting the relationships between sources and consumers.</p>	<p>IBM InfoSphere DataStage</p> <p>Extracts, transforms, and loads data between multiple sources and targets</p>
<p>IBM InfoSphere Change Data Capture</p> <p>Real-time Change Data Capture (CDC) and replication solution across heterogeneous environments</p>	<p>IBM InfoSphere DataStage MVS Edition</p> <p>Provides native data integration capabilities for the mainframe.</p>
<p>IBM InfoSphere Change Data Capture for Oracle Replication</p> <p>Real-time data distribution and high availability/disaster recovery solution for Oracle environments</p>	<p>IBM InfoSphere Federation Server</p> <p>Defines integrated views across diverse and distributed information sources, including cost-based query optimization and integrated caching.</p>
<p>IBM InfoSphere Information Analyzer</p> <p>Profiles and establishes an understanding of source systems and monitors data rules.</p>	<p>IBM InfoSphere Information Services Director</p> <p>Allows information access and integration processes to be published as reusable services in a service oriented architecture.</p>
<p>IBM InfoSphere Business Glossary</p> <p>Creates, manages, and searches metadata definitions.</p>	<p>IBM InfoSphere Information Server FastTrack</p> <p>Simplifies and streamlines communication between the business analyst and developer by capturing business requirements and automatically translating into DataStage ETL jobs.</p>
<p>IBM InfoSphere QualityStage</p> <p>Standardizes and matches information across heterogeneous sources.</p>	<p>Connectivity Software</p> <p>Provides connectivity software for efficient and cost-effective cross-platform, high-speed, real-time, batch and change-only integration for your data sources.</p>

Demos
<p>IBM InfoSphere Information Server</p> <p>→ Take a guided tour</p>
<p>IBM Information Server Overview</p> <p>→ Watch it today</p>

Data sheets
<p>IBM Information Server</p> <p>→ Understanding and analyzing your data</p>
<p>IBM Information Server</p> <p>→ Cleansing Data and Managing Data Quality</p>
<p>IBM Information Server</p> <p>→ Transforming and delivering data</p>

White papers
<p>IBM InfoSphere Information Server</p> <p>→ Simplify integration with unified metadata</p>
<p>Driving information-led business innovation with IBM Information Server</p> <p>→ Deliver information you can trust</p>



→ Learn more

InfoSphere Information Server Resources

→ IBM InfoSphere Information Server - Simplify integration with unified metadata

→ Driving information-led business innovation with IBM Information Server - Deliver information you can trust

Highlights

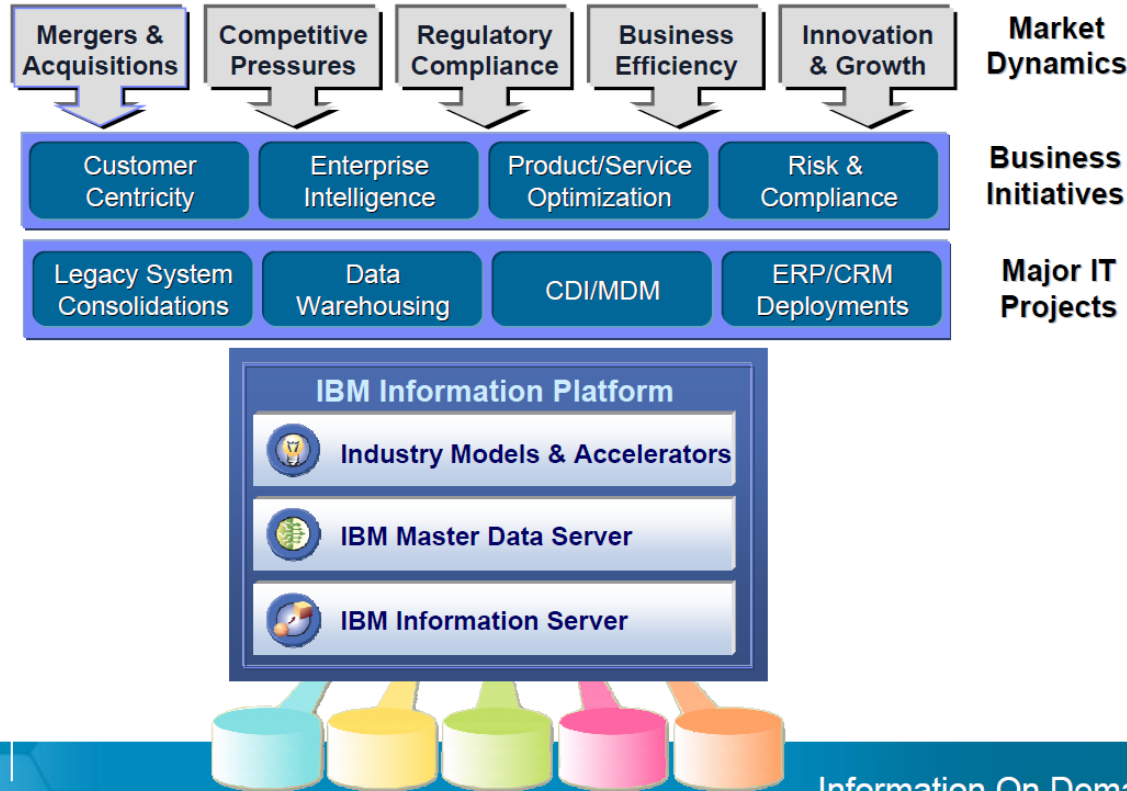
📄 Announcement letter - IBM InfoSphere Information Server V8.1 (195kB)

→ IBM Leverages DataMirror Acquisition for Real-Time Change Data Capture

📄 Get Adobe® Reader®

IBM Information Platform

Trusted Information Enables Business Success



Felix Naumann
Information Integration
Winter 2019/20

Definition Information Integration

Information integration is the combination of data and content from different sources to a common information set.

Information integration is the **correct**, **complete**, and **efficient** combination of data and content from different, **heterogeneous** sources to a common and **structured** information set for **interpretation** by users and applications.

Felix Naumann
Information Integration
Winter 2019/20

Where is Information Integration applied?

- In general
 - Business integration
 - Application integration
 - Prozess integration (workflow integration)

- In particular
 - Databases and information systems
 - Distributed
 - Autonomous
 - Heterogeneous

```

ID   RNGTHCHI   standard; RNA; ROD; 1016 BP.
XX
DT   01-AUG-1991 (Rel. 28, Created)
DT   04-MAR-2000 (Rel. 63, Last updated, Version 2)
XX
DE   Rat GTP cyclohydrolase I mRNA, complete cds.
XX
KW   GTP cyclohydrolase I.
XX
OS   Rattus norvegicus (Norway rat)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
XX
RN   [1]
RP   1-1016
RX   MEDLINE; 91093270.
RX   PUBMED; 1985963.
RA   Hatakeyama K., Inoue Y., Harada T., Kagamiyama H.;
RT   "Cloning and sequencing of cDNA encoding rat GTP cyclohydrolase I: The
RT   first enzyme of the tetrahydrobiopterin biosynthetic pathway";
RL   J. Biol. Chem. 266(2):765-769(1991).
XX
FT   CDS           128..853
FT           /codon_start=1
FT           /db_xref="GOA:P22288"
FT           /db_xref="SWISS-PROT:P22288"
FT           /EC_number="3.5.4.16"
FT           /gene="GTP cyclohydrolase I"
FT           /product="GTP cyclohydrolase I"
FT           /protein_id="AAA41299.1"
FT           /translation="MEKPRGVRC TNGFPERELPRPGASRPAEKS RPPEAKGAQPADAWK
FT           AGRPRSEEDNELNLPNLAAAYSSILRSLGEDPQRQGLLKTWPRAATAMQFFTKGYQETI
FT           SDVNLNDAIFDHDHDMVIVKIDIMFMSCEHHLVFPVGRVHIGYLPNKQVLGLSKLARIV
FT           EIYSRRLQVQERLTKQAVAITALQPAGVGVVIEATHMCMVMRQVQKMNKSTVTSTML
FT           GVFREDPKTREFFLTLIRS"
SQ   Sequence 1016 BP; 236 A; 279 C; 291 G; 210 T; 0 other;
gacttcgaac ctccattcggg gcagaactcc tgtccccggtg acagccacag gtcacggccc 60
ccgggctaagc cgagccgcgag cgcttggttag caccttaggg tgtctcggga gcaatcgcgc 120
cgggtccatg gagaagccgc ggggtgtaag gtgcaccaat ggggtccccg agcggggagct 180
...
catcaggagc tgaacttccg tgtgcgagcc cgggtttgca gacccccgct gaggccagcg 900
ttatctgtct cgattgtaca ttccagttcc agttgggtata cttgtcaact ttatttctca 960
ccatgaattg tatttaataa ttatttatag agatgtcaaa taaagggtgat caact 1016

```

Molecule type
Name

Date of creation and last update

Free text description

Keywords describing the molecule

Organism

**Article the sequence
was published in**

Structural annotation (coding sequence)

**Link to functional annotation
of resulting protein**

**Translated
protein
sequence**

**Sequence
of
bases**

Felix Naumann
Information Integration
Winter 2019/20



Erweiterte Suche Bücher

Je mehr Felder Sie ausfüllen, desto zielgerichteter können wir suchen. Es reicht jedoch aus, nur eines der Felder auszufüllen.

Autor/in:

Titel:

Schlagwörter:

ISBN: (10- oder 13-stellig, ohne Bindestriche)

Verlag:

Verfeinern Sie Ihre Suche, indem Sie nur nach bestimmten Buchformaten suchen.

Nur gebraucht:

Format:

Ordnen nach:

Erscheinungsdatum:

Suche in: Deutsche Bücher Englische Bücher

Beispiele

- ♦ Geben Sie **"fidelity"** in das Titel- und **"hornby"** in das Autoren-Feld ein, erscheint *High Fidelity* von Nick Hornby.
- ♦ Indem Sie **"grisham, john"** in das Autor-Feld und **"Taschenbücher"** im Feld **"Format"** auswählen, finden Sie nur die Taschenbuchausgaben der Romane von John Grisham.


www.xmethods.net

Recent Listings [View the [FULL LIST](#)]

Publisher	Style	Service Name	Description	Implementation
StrikeIron	DOC	Try It Gale Group Web Domain Business Intelligence	Based on a website domain get in-depth financial and corporate information	
StrikeIron	DOC	Try It Gale Group Business Intelligence	Based on company name get in-depth financial and corporate information	
StrikeIron	DOC	Try It Gale Group Business Information	Standard financial and corporate information for 440,000 U.S. and international companies	
StrikeIron	DOC	Try It IPligence Geo IP Location	Comprehensive information on IP addresses	
mharvanek	DOC	Try It DuoShare Address Quality Integrator	DuoShare Address Quality Integrator standardizes, corrects, validates (to the delivery point), and enhances U.S. addresses using a number of USPS® certified processes.	
CDYNE	DOC	Try It CDYNE Phone Notify!	Calls any phone number and speaks text or sound file to the person. Also supports advanced Dialplans, Machine Detection and Incoming calls.	
CDYNE	DOC	Try It CDYNE SMS Notify!	Send SMS Text Messages to your mobile telephone devices without the use developing your own software!	
StrikeIron	DOC	Try It The Sports Network Men's College Basketball	Comprehensive sports data for NCAA Mens Division I Basketball	
StrikeIron	DOC	Try It Wall Street Horizon Interest Rate Calendar	Access current and historical interest rate information	
EDIXMLOnline	DOC	Try It EDIXMLOnline	Translate, generate & validate EDI files using X12/EDIFACT SEF files. Translate EDI to XML format.	MS .NET
StrikeIron	DOC	Try It Wall Street Horizon Real-Time Company Earnings	Information to analyze and evaluate investments and future earning potential	
phi1281	RPC	Try It Weight Watchers Point Calculator	Given the parameters of calories, fat grams, & fiber grams, computes the Weight Watchers point value for a given serving of food.	ColdFusion
SOATrader	RPC	Try It Captcha Web Service	This Web service will create captcha	

Fertig 6.828s

StrikeIron: Data as a Service



CONTACT REGISTER LOGIN

MARKETPLACE SOLUTIONS PRODUCTS DEVELOPERS PARTNERS NEWS COMPANY SUPPORT YOUR ACCOUNT

- Business Users
- Enterprise IT Professionals
- Application Developers
- Partners

STRIKEIRON: Strikelron's Data

Services give you access to live data to use now, integrate into applications or build into Web sites.

SEARCH WEB SERVICES

Search

All Categories ▼


Browse

LATEST NEWS

Strikelron Offers Gale Data on
Salesforce.com's AppExchange

Strikelron Hits Home Run
with Major League Baseball Data from The Sports Network

Strikelron Data Fuels
Enterprise Mashup Platforms



WHAT DO YOU WANT TO DO?

CONNECT

- Find Data Services
- Browse Solutions
- Cleanse & Enhance Data

CREATE

- Build Mashups in Excel
- Create Business Solutions
- Integrate into Salesforce.com

COLLABORATE

- **New!** Developer Community
- Download Code and Clients
- Access Free Data

DATA SPOTLIGHT

NEW WEB SERVICES

- D&B Business Prospect
- Gale Business Intelligence
- Midnight Trader Financial News
- MapQuest Driving Directions
- TSN Men's College Basketball
- Wall Street Horizon Company Earnings


HOT WEB SERVICES

- US Address Verification
- Global SMS Pro
- Sales and Use Tax
- Email Verification
- Foreign Exchange Rates

PARTNERS

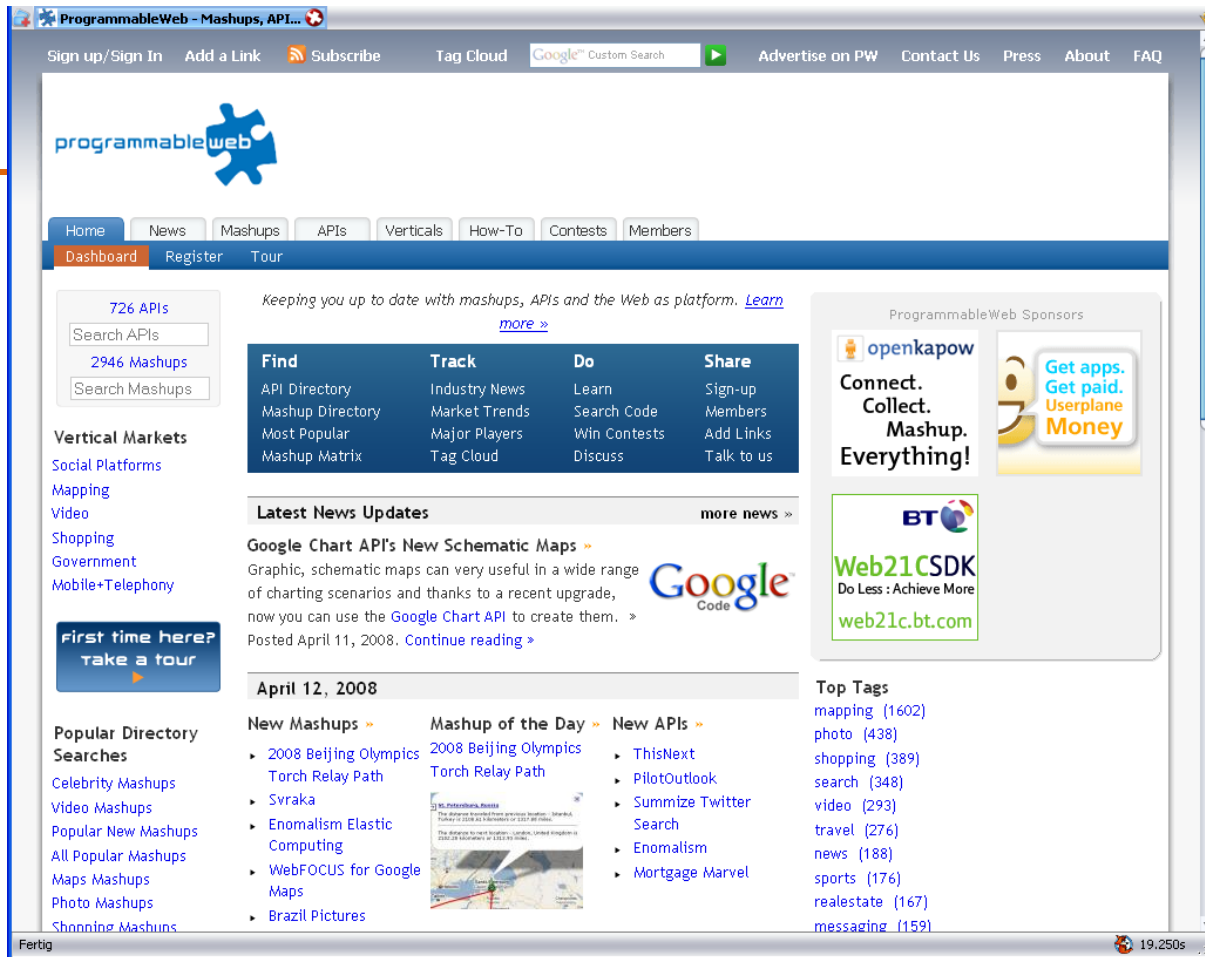
Strikelron is proud to partner with the following industry leaders:

[Blockieren](#)



© 2003-2008 Strikelron, Inc. Terms of Use | Privacy | Trademarks | Contact Us | Site Map

Felix Naumann
Information Integration
Winter 2019/20



ProgrammableWeb - Mashups, API...

Sign up / Sign In Add a Link Subscribe Tag Cloud Google Custom Search Advertise on PW Contact Us Press About FAQ

programmableweb

Home News Mashups APIs Verticals How-To Contests Members

Dashboard Register Tour

726 APIs Search APIs

2946 Mashups Search Mashups

Vertical Markets

Social Platforms

Mapping

Video

Shopping

Government

Mobile+Telephony

First time here? Take a tour

Popular Directory Searches

Celebrity Mashups

Video Mashups

Popular New Mashups

All Popular Mashups

Maps Mashups

Photo Mashups

Shinning Mashups

Keeping you up to date with mashups, APIs and the Web as platform. [Learn more >>](#)

Find	Track	Do	Share
API Directory	Industry News	Learn	Sign-up
Mashup Directory	Market Trends	Search Code	Members
Most Popular	Major Players	Win Contests	Add Links
Mashup Matrix	Tag Cloud	Discuss	Talk to us

Latest News Updates [more news >>](#)

Google Chart API's New Schematic Maps >>

Graphic, schematic maps can very useful in a wide range of charting scenarios and thanks to a recent upgrade, now you can use the [Google Chart API](#) to create them. >

Posted April 11, 2008. [Continue reading >](#)

April 12, 2008

New Mashups >>

- 2008 Beijing Olympics Torch Relay Path
- Svraka
- Enomalism Elastic Computing
- WebFOCUS for Google Maps
- Brazil Pictures

Mashup of the Day >>

2008 Beijing Olympics Torch Relay Path

New APIs >>

- ThisNext
- PilotOutlook
- Summize Twitter Search
- Enomalism
- Mortgage Marvel

Top Tags

- mapping (1602)
- photo (438)
- shopping (389)
- search (348)
- video (293)
- travel (276)
- news (188)
- sports (176)
- realestate (167)
- messaging (159)

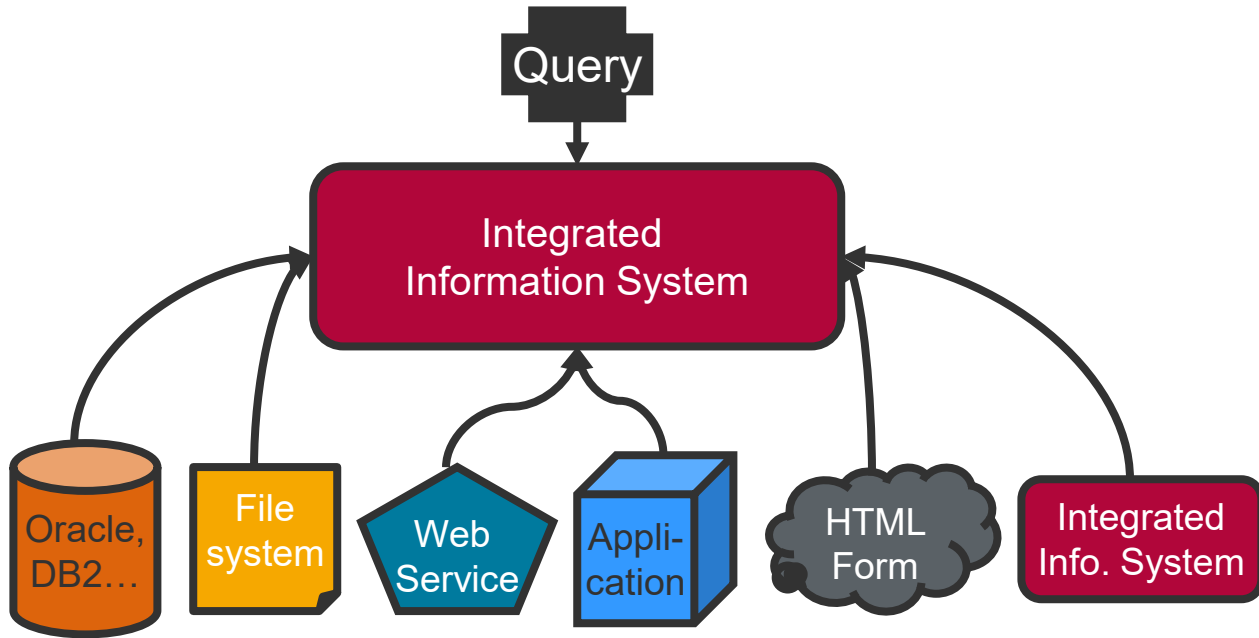
openkapow Connect. Collect. Mashup. Everything!

Get apps. Get paid. Userplane Money

BT Web21CSDK Do Less : Achieve More web21c.bt.com

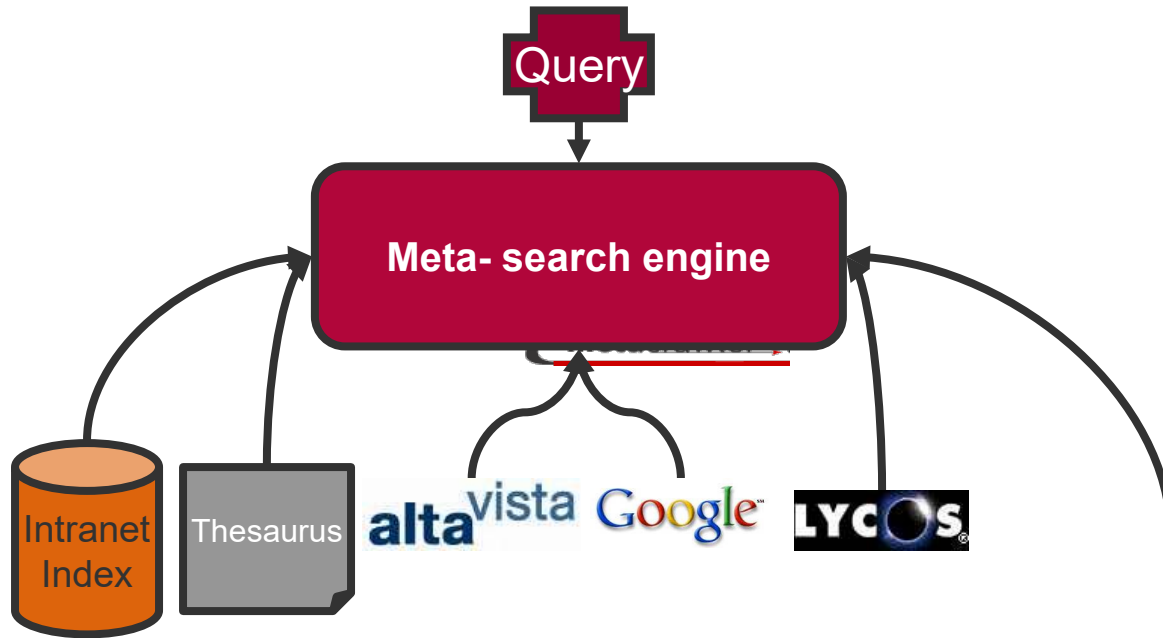
Fertig 19.250s

Integrated Information Systems



Felix Naumann
Information Integration
Winter 2019/20

Integrated Search Engine



Felix Naumann
Information Integration
Winter 2019/20

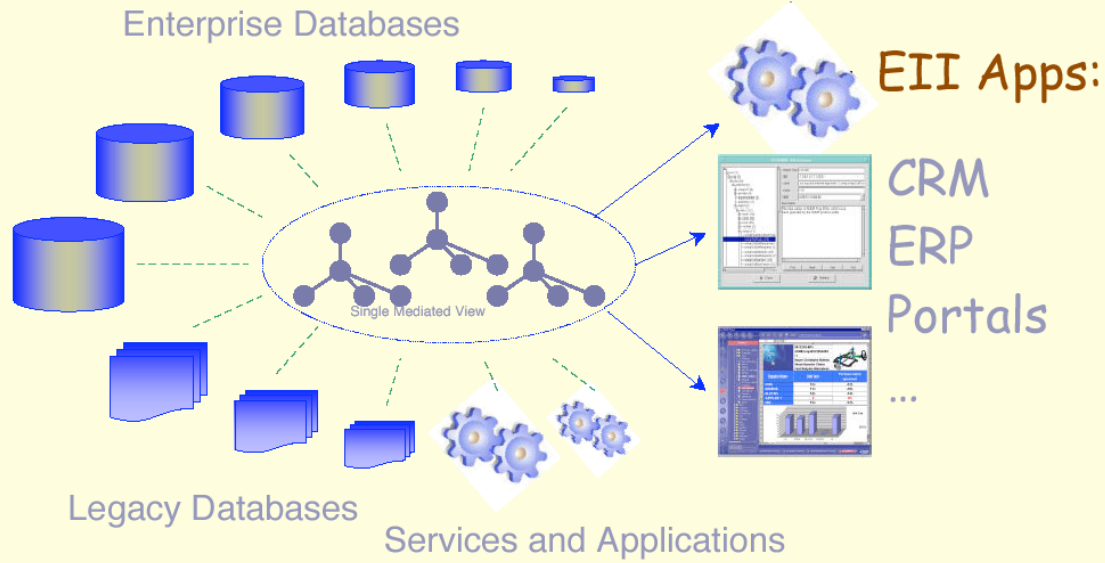
Integration = Abstraction

- Logical database design abstracts from physical database design
 - Data independence
 - Queries: Procedural vs. declarative
- Information integration abstracts from logical database design
 - Source independence (physical and logical location)
 - Data model and syntax independence
 - Independence of semantic differences (hopefully!)
- Following slides: VLDB 2004 keynote speech
 - Structures, Semantics and Statistics
 - Alon Halevy, University of Washington



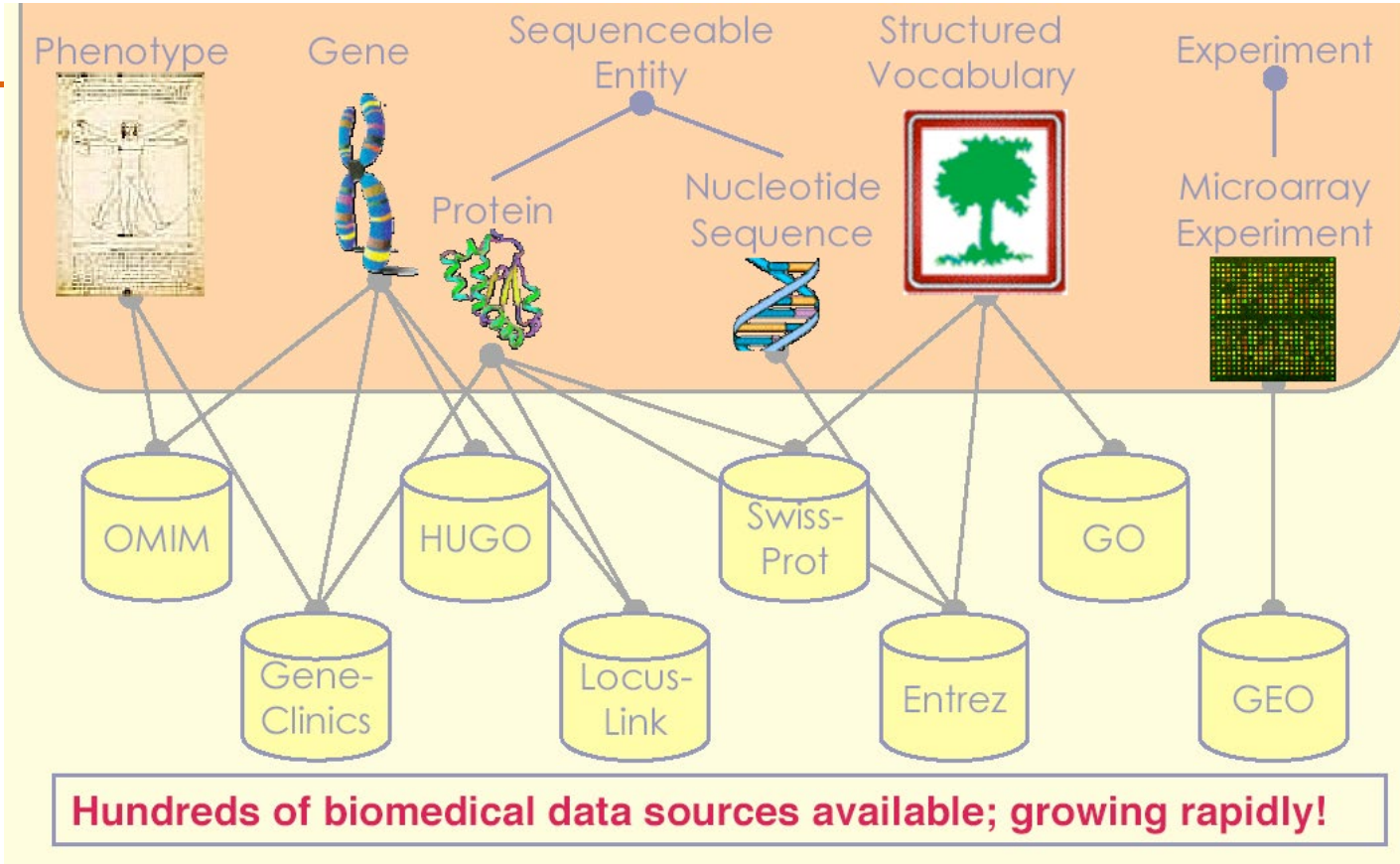
Felix Naumann
Information Integration
Winter 2019/20

Application Area 1: Business



50% of all IT \$\$\$ spent here!

Felix Naumann
Information Integration
Winter 2019/20



Felix Naumann
Information Integration
Winter 2019/20

Application Area 3: The Web

Over 450,000 web-accessible data sources!

Felix Naumann
Information Integration
Winter 2019/20

Information integration: An old problem

- On the research agenda since 50 years
- Early systems in the 70s
- Manual integration earlier
- New problems
 - Many, many sources
 - Heterogeneity
 - New types of data (XML, GIS, OO,...)
 - New types of queries (search, UDFs,...)
 - New types of results (ranking, visualization, ...)
 - New types of users (laymen, manager, admins, ...)
- Alon Halevy: „It’s plain hard!“ [Halevy04]

Felix Naumann
Information Integration
Winter 2019/20

Why is it so difficult? [Halevy04]

■ Systems reasons:

- Managing different platforms
- Query processing across multiple systems

■ Social reasons:

- Locating and capturing relevant data in the enterprise.
- Convincing people to share (data fiefdoms)
- Privacy and performance implications.

■ Logic reasons:

- Schema (and data) heterogeneity
- Challenge independent of integration architecture!

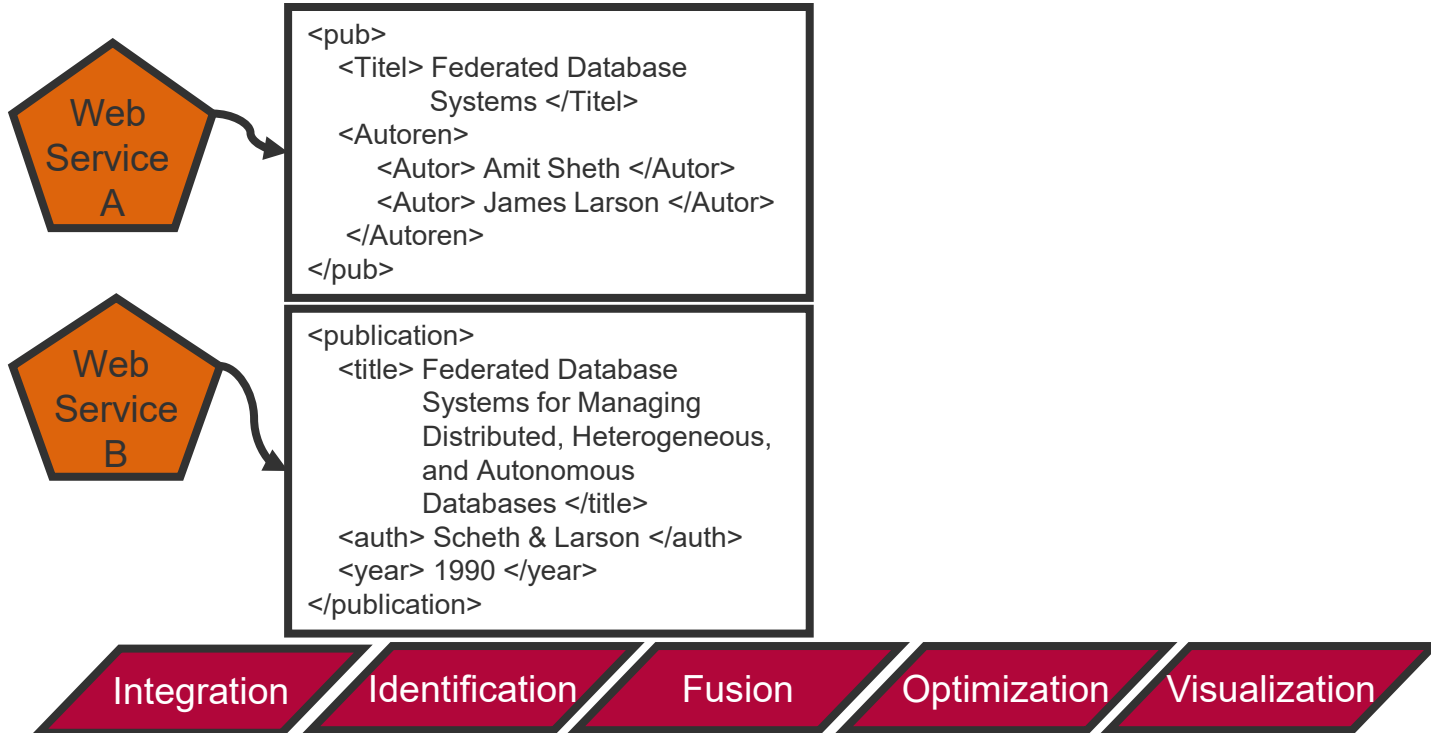
Overview

1. Introduction of research group
2. Organisation
3. Information systems
- 4. Information integration
by example**
5. Overview of semester



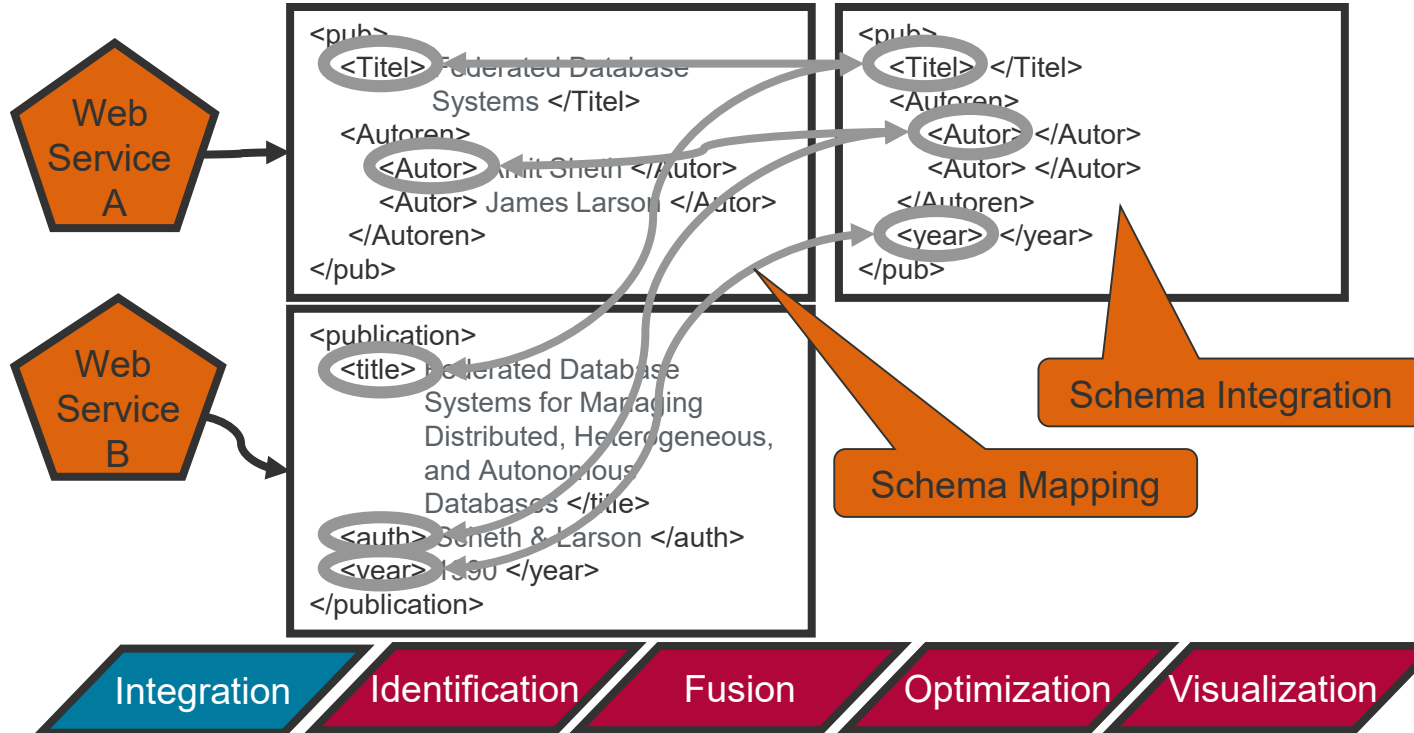
Felix Naumann
Information Integration
Winter 2019/20

Information integration



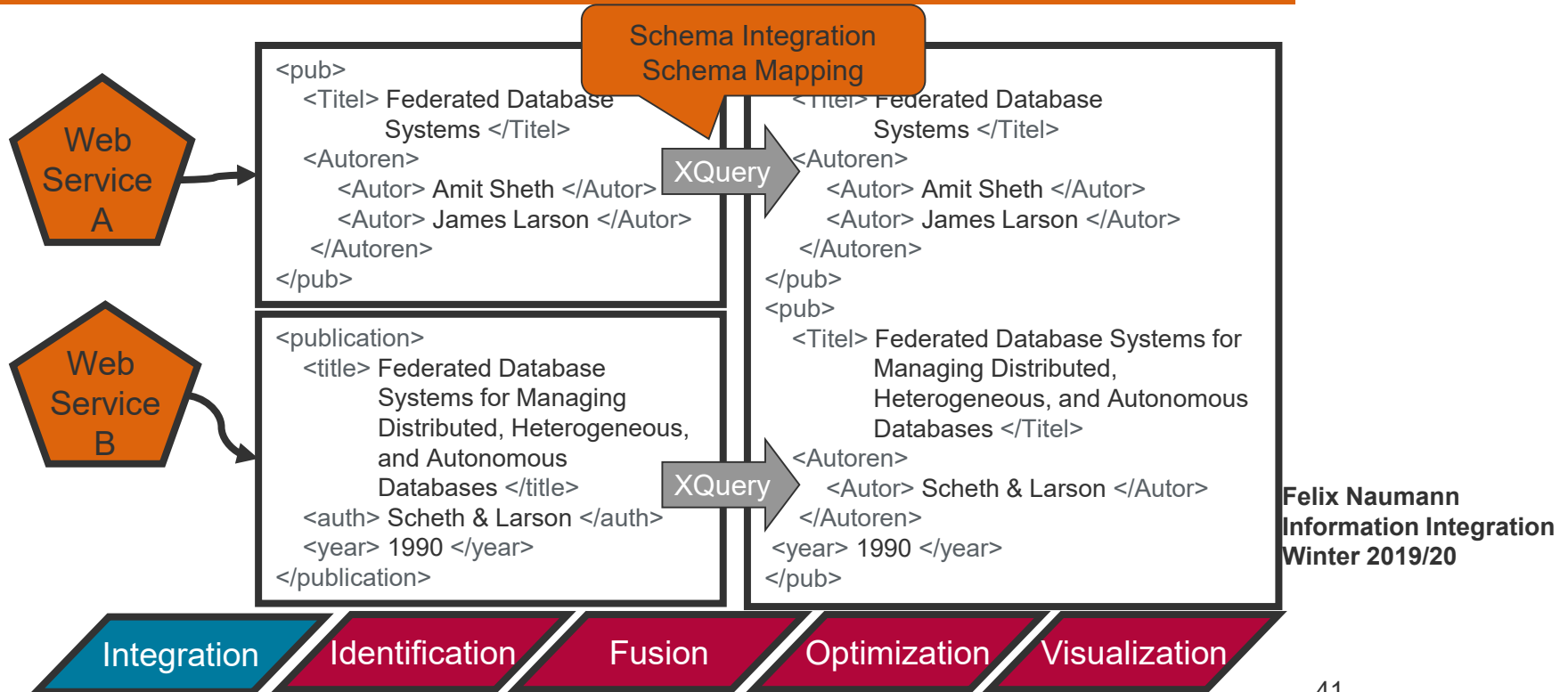
Felix Naumann
Information Integration
Winter 2019/20

Information integration



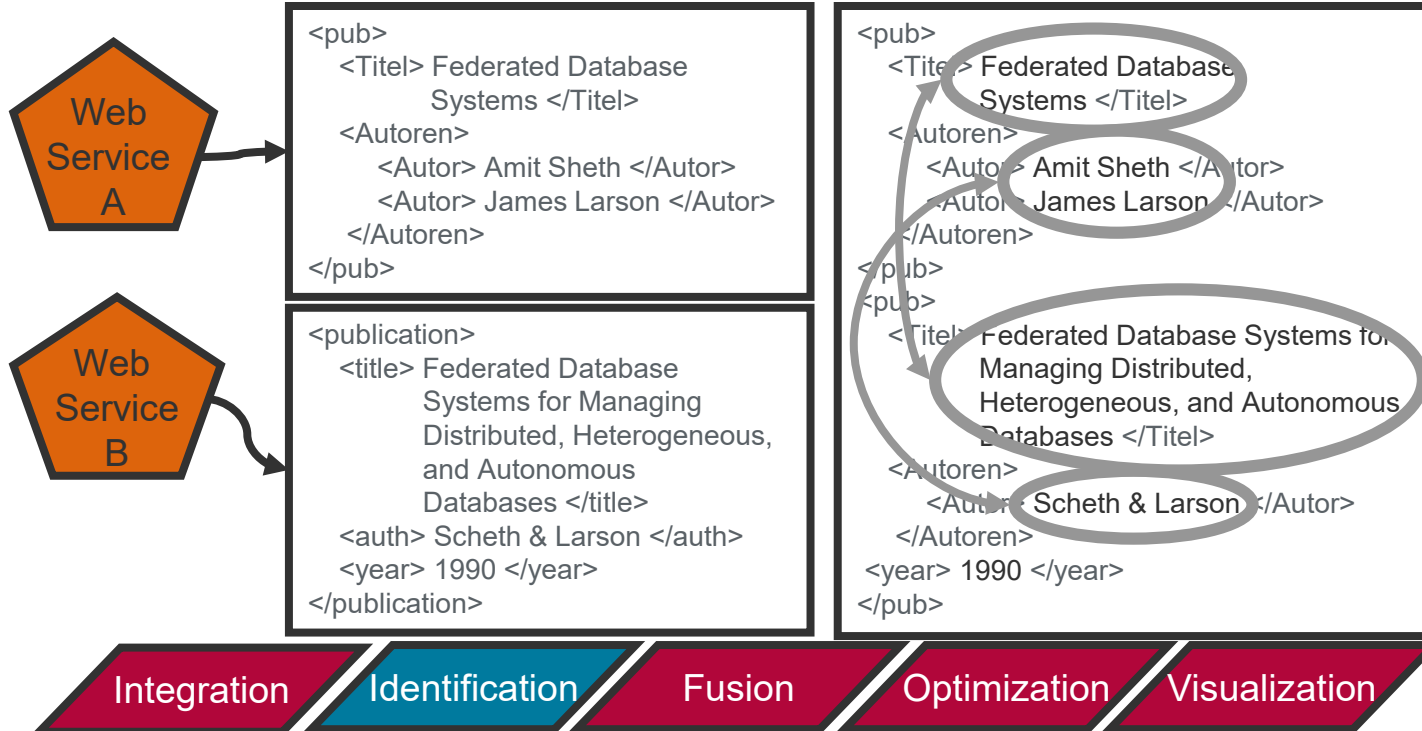
Felix Naumann
Information Integration
Winter 2019/20

Information integration



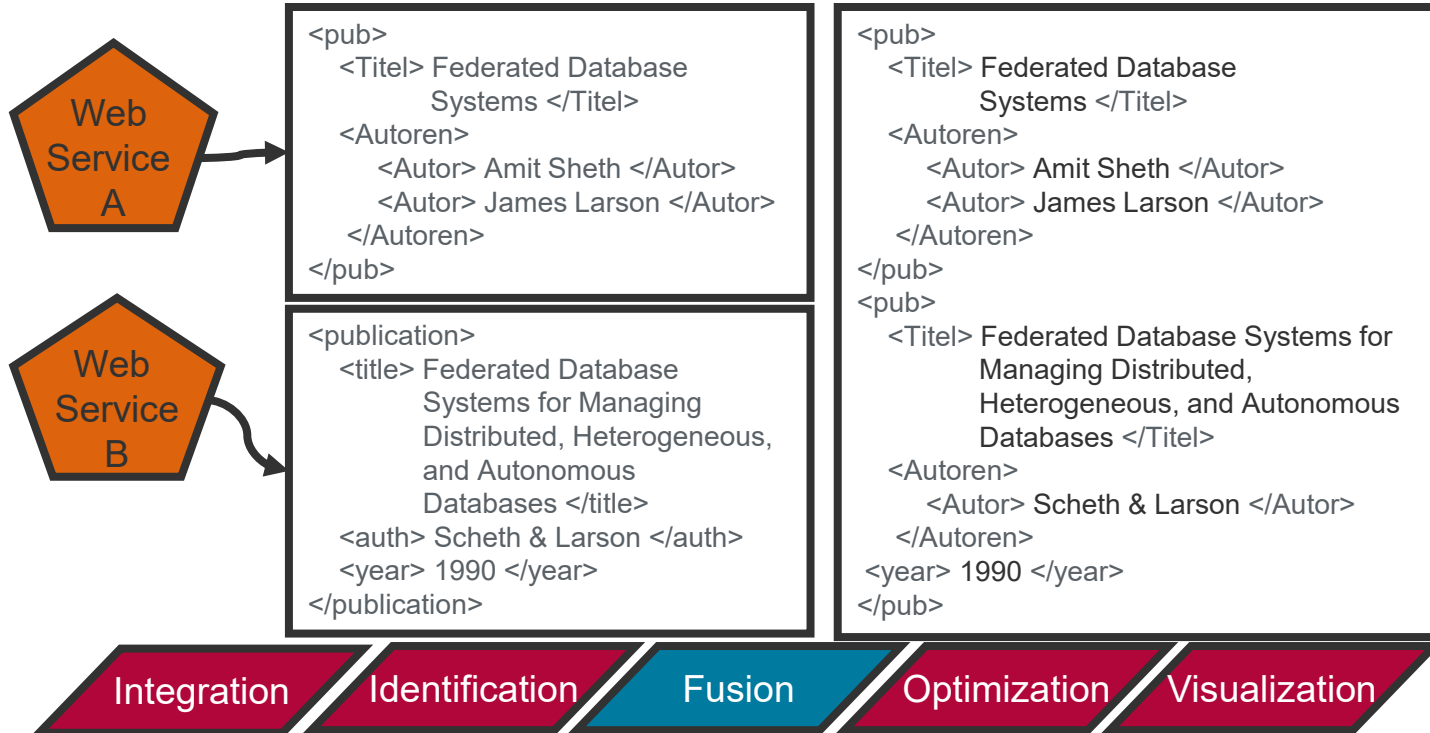
Felix Naumann
Information Integration
Winter 2019/20

Information integration



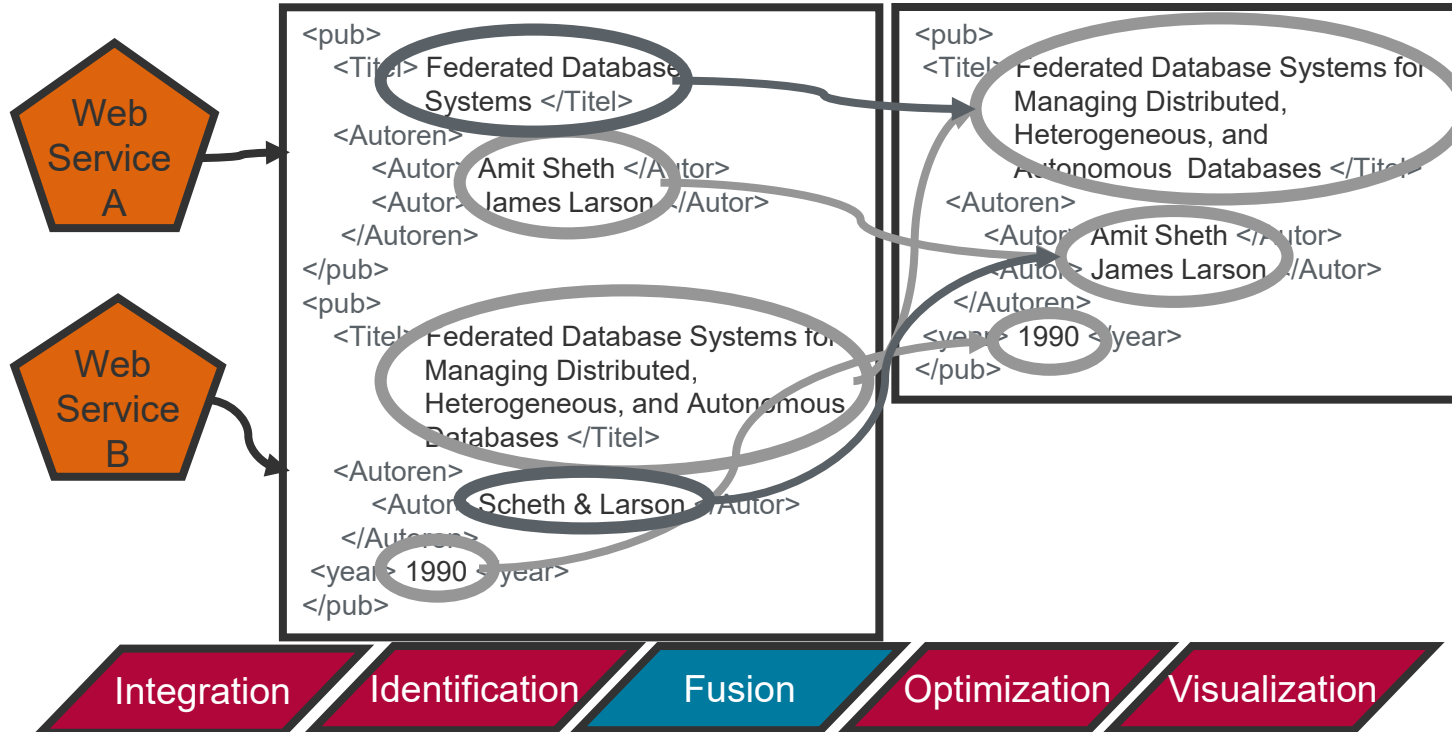
Felix Naumann
Information Integration
Winter 2019/20

Information integration



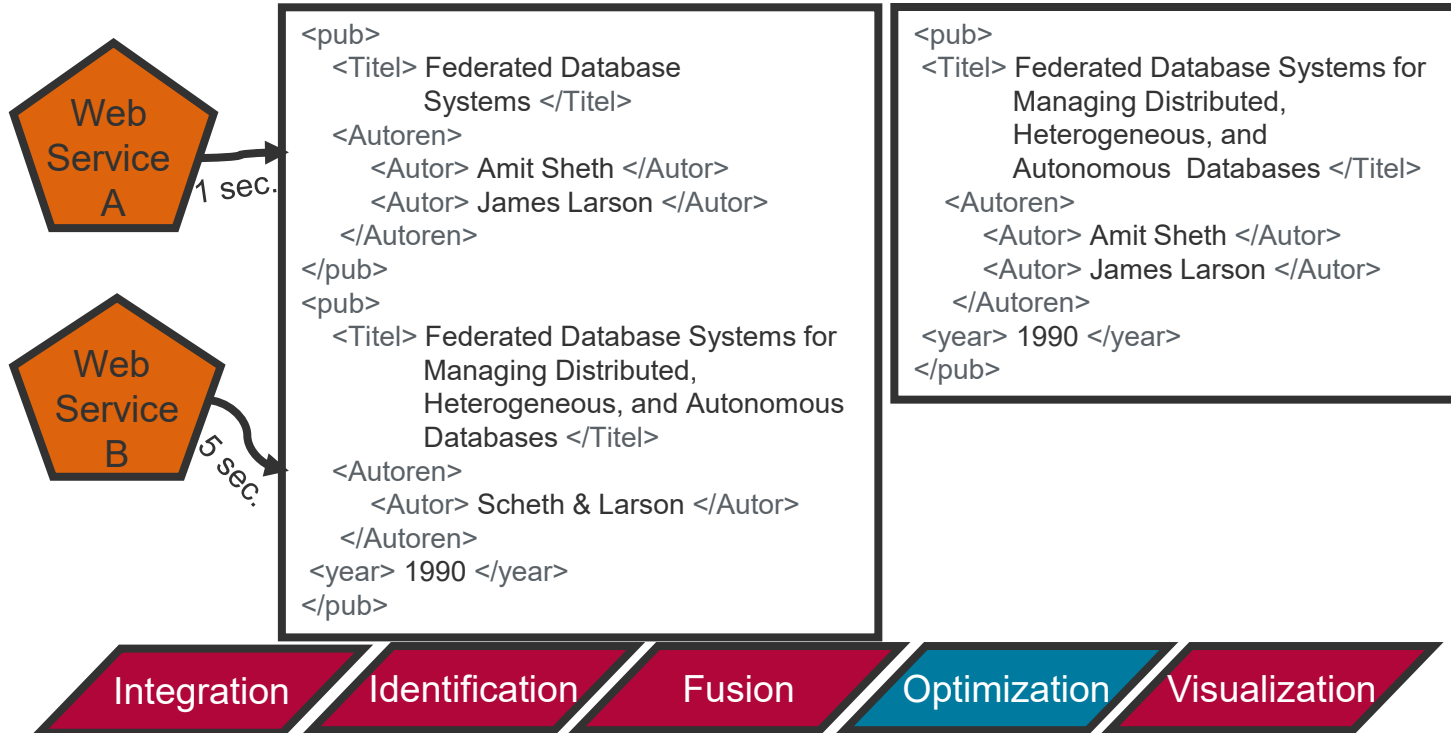
Felix Naumann
Information Integration
Winter 2019/20

Information integration



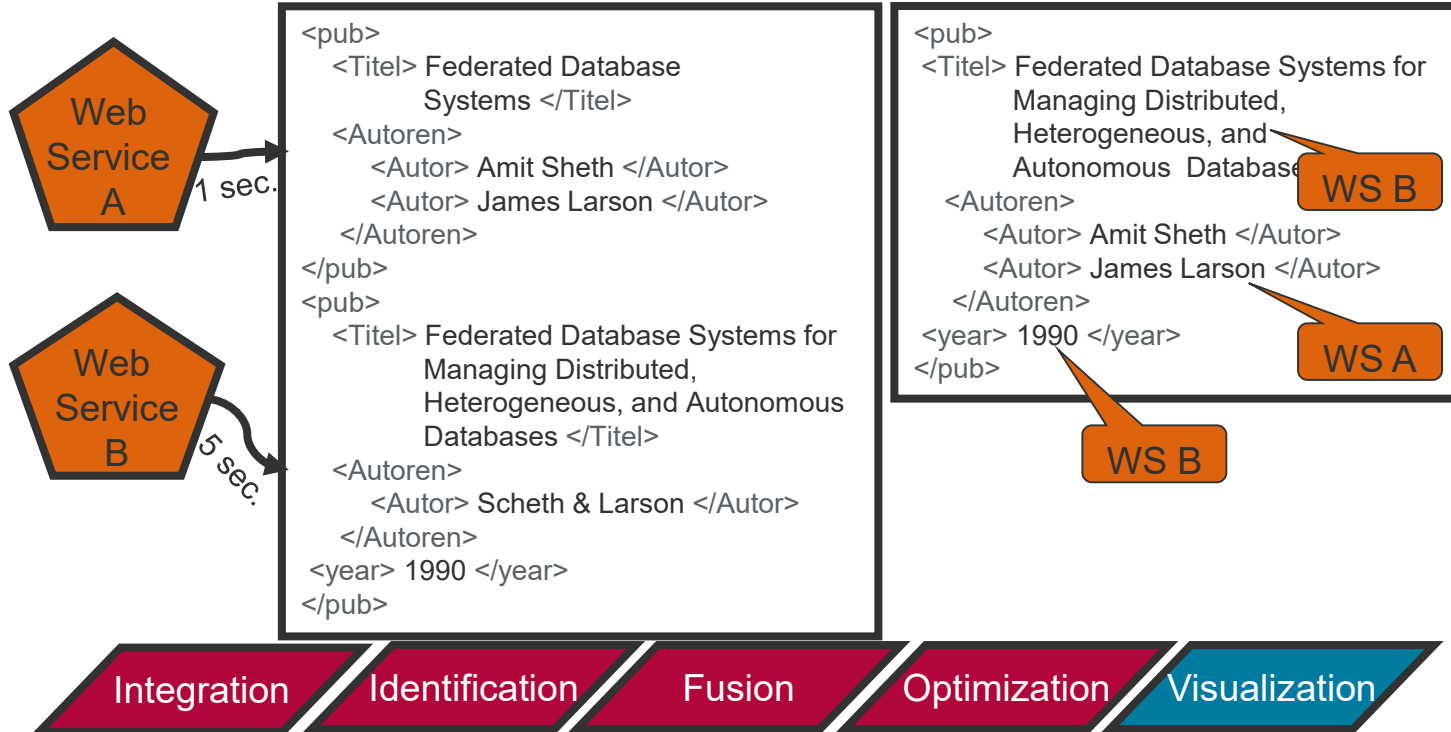
Felix Naumann
Information Integration
Winter 2019/20

Information integration



Felix Naumann
Information Integration
Winter 2019/20

Information integration



Felix Naumann
Information Integration
Winter 2019/20

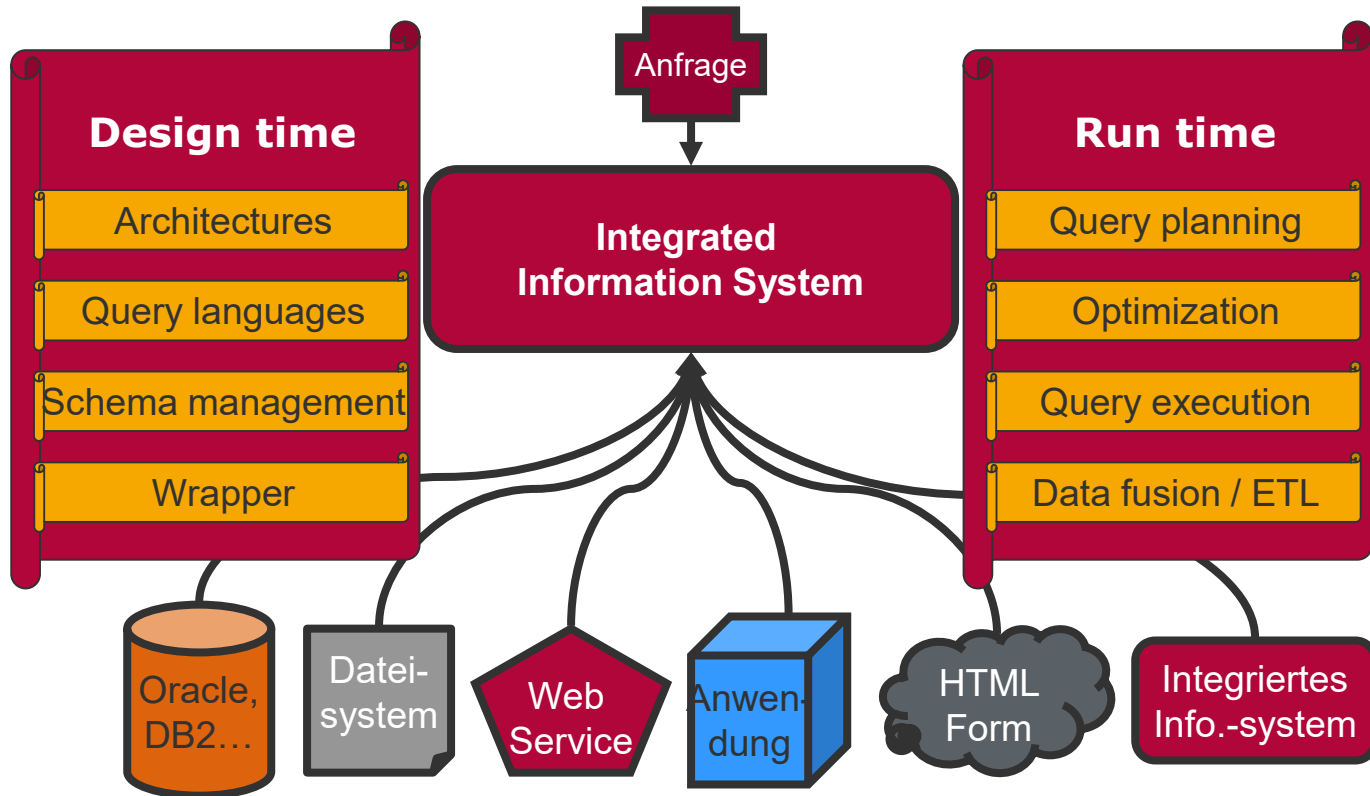
Overview

1. Introduction of research group
2. Organisation
3. Information systems
4. Information integration by example
- 5. Overview of semester**



Felix Naumann
Information Integration
Winter 2019/20

Integrated Information Systems



Felix Naumann
Information Integration
Winter 2019/20

Part I

- Einführung in die Informationsintegration
- Szenarien der Informationsintegration
- Verteilung und Autonomie
- Heterogenität

- Materialisierte und virtuelle Integration
- 5-Schichten-Architektur
- Mediator/Wrapper-Architektur / PDMS

- Schema Mapping
- Schema Matching
- SchemaSQL

- Global-as-View und Lokal-as-View Modellierung
- Global-as-View Anfragebearbeitung

Problem statement

Architectures

Mapping

Modeling

Felix Naumann
Information Integration
Winter 2019/20

Part II

- Containment & Local-as-View Anfragebearbeitung
- Bucket Algorithmus
- Verteilte Anfragebearbeitung

Queries

- Duplikaterkennung
- Datenfusion - Union & Co.
- DWH, ETL & Data Lineage
- Informationsqualität

Data integration

- Hidden Web
- Semantic Web
- ...

Applications

Felix Naumann
Information Integration
Winter 2019/20