

Überlegungen zu einer Spatial Big Data Architektur im BigGIS Projekt



Andreas Abecker¹, Torsten Brauer¹, Johannes Kutterer¹,
Karsten Schnitter¹, Jens Nimis², Patrick Wiener²

¹ Disy Informationssysteme GmbH

² Hochschule Karlsruhe - Technik und Wirtschaft

Optimal handeln mit Spatial Analytics

Unsere Mission: Aus Daten mit geografischem Bezug Wissen erschließen, anschaulich visualisieren und in Entscheidungsprozesse integrieren!

Full-Service Dienstleister mit eigener **Software made in Gemany**
Premium Fach- und Informatik-Knowhow für Ihre Projekt



- **Führendes Softwarehaus für Lösungen zum raumbezogenen Berichtswesen**
- Gegründet 1997, Standort Karlsruhe, ca. 65 Mitarbeiter
- Spezialisiert auf anspruchsvolle Aufgaben der **Datenanalyse und des Datenmanagements raumbezogener Daten** in großen Organisationen
- Anwendungsgebiete überwiegend im **Umweltdatenmanagement** und **eGovernment**.
- Lösungen auf unserem **Kernprodukt: Spatial Analytics benötigt mehr als GIS!**



Lärm: In welchem Stadtteil sind wie viele Personen von Straßen-, Flugzeug- oder Schienenlärm betroffen? Wie ändert sich das, wenn Lärmschutzmaßnahmen getroffen werden?



Hochwasser- und Katastrophenschutz: Wie viele Personen sind von einem drohenden Hochwasser betroffen? Gibt es in dem Bereich produzierendes Gewerbe mit Gefahrstoffen?

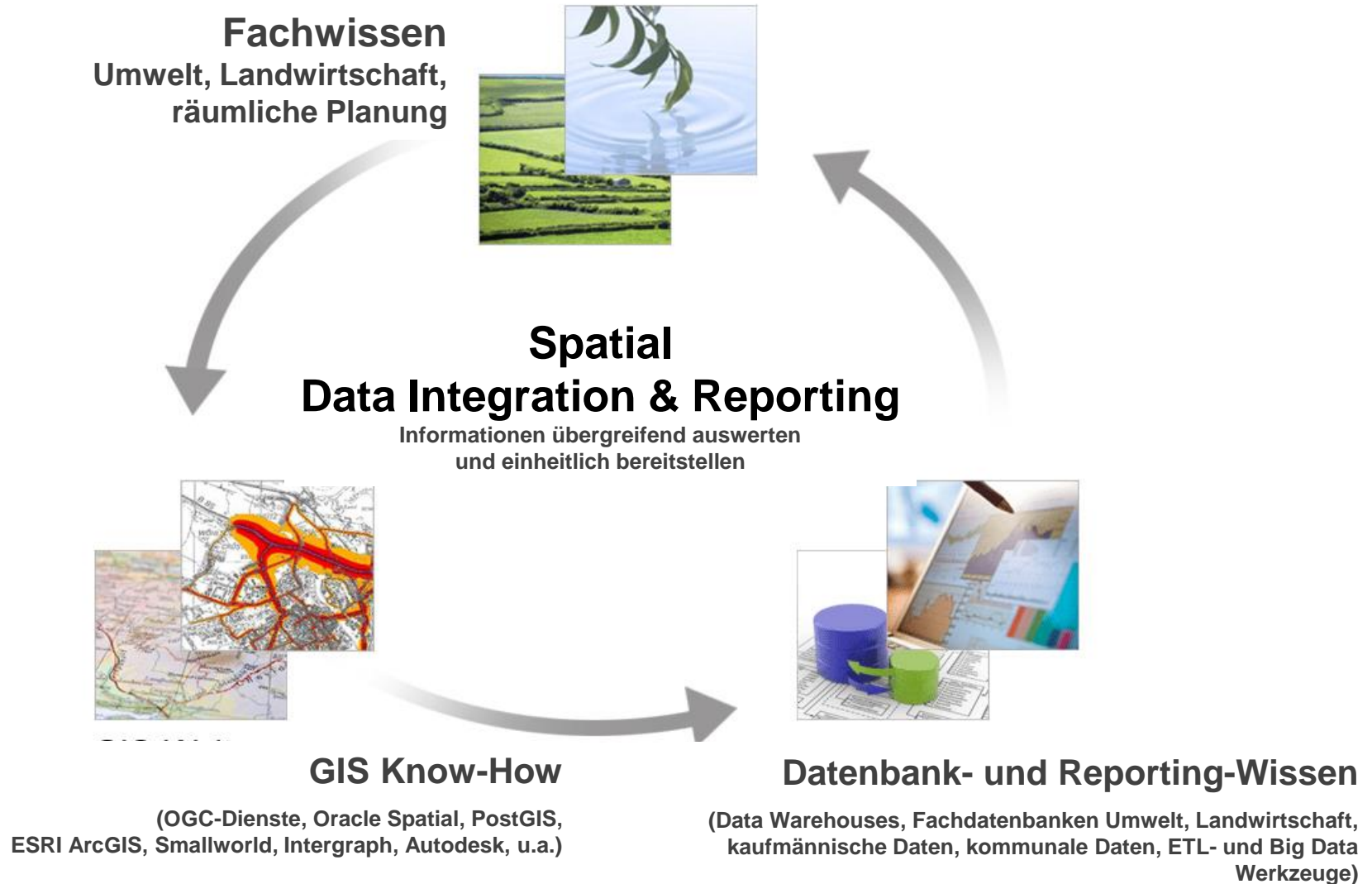


Umwelt, Wasserqualität: Wie gut ist die Gewässergüte in einem Flussabschnitt? Welche Betriebe nutzen das Wasser flussaufwärts?



Landwirtschaft: Auswertung und Controlling von EU-Fördergeldern: Welcher Landwirt bekommt aufgrund welcher Felder Förderung? Ist das stimmig?

Wie groß ist der erwartete Ertrag in einem Anbaugebiet? Welche Erträge sind zu erwarten, wenn ein Unwetter- oder Hochwasserereignis eintritt?



- **Große Fachverwaltungen mit Raumbezug**

Umwelt

Landwirtschaft/ Verbraucherschutz

Innere Sicherheit

Weitere: Infrastruktur/ Verkehr/ Immobilien

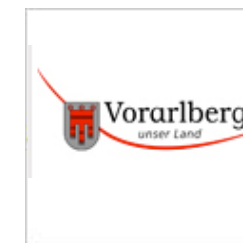
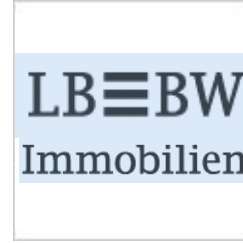
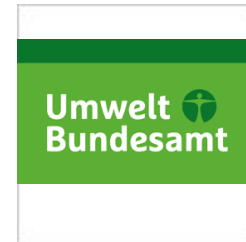
- **Unternehmen mit raumbezogenen Geschäftsmodell**

Agrarindustrie

Infrastruktur/ Bahn



Kunden (Auszug)



 **ORACLE** Locator/Spatial

 **SAP HANA**

 PostgreSQL / PostGIS

 SQLite / Spatialite

 **OGC**[®] WMS
Open Geospatial Consortium, Inc.

 **OGC**[®] WMTS
Open Geospatial Consortium, Inc.

 **OGC**[®] WFS(-T)
Open Geospatial Consortium, Inc.

 **ArcGIS** Server (Rest)
ESRI

 Shapefile

 **Informix**[®] SOFTWARE

 Access

 **IBM**[®] DB2[®]

 Microsoft[®] SQL Server[™]

INGRES[™]

 OpenStreetMap

 **HSQL**[®] database engine

Apache Derby 

 Firebird

Daneben: DXF, CSV, XLS, DBF, WPS, MrSID, Rasterbildkataloge...

disy Forschung



- BMBF-KMU_innovativ: **PartSense** – Participatory Sensing für mobile Workforce und Bürgerbeteiligung (Jan 2012 – Dez 2013)
- BMBF-IWRM: **IWRM-Südafrika-II** – Integriertes Wasserressourcenmanagement (Juni 2012 – Dez 2015) (disy: Data Warehousing, Web-GIS)
- EU-FP7-STREP: **WatERP** – Integriertes Wasserressourcenmanagement (Okt 2012 – Sep 2015) (disy: Water Data Warehousing, Sensordatenmgt.)
- BMBF-KMU_innovativ: **RichWPS** – Benutzerfreundliche Geodatendienste (Jan 2013 – Mar 2015) (disy: Konsortialführung, WPS-Client/-Server-Software)
- EU-CIP-PSP: **eENVplus** – eEnvironmental Services for Advanced Application within INSPIRE (Jan 2013 – Dez 2015) (disy: Multi-Thesaurus-Anbindung)
- BMWi ZIM **Automatisiertes Deponiemanagement** (Dez 2013 – Mai 2016) (disy: intelligente telemetrische Sensordatenverarbeitung)
- BMWi **SmartRegionalStrategy** (Januar 2015 – Juni 2018) – Strategische Entscheidungsunterstützung mit Big Geo Data für Energieversorger
- BMBF FONA-Programm: **SMART_MOVE** (Februar 2015 – Dezember 2017) – Integriertes Wasserressourcenmanagement am Jordan
- BMBF **BigGIS** (April 2015 – März 2018) – Big Data für Geodaten in Umweltmonitoring, Katastrophenschutz und Smart Cities
- BMBF FONA-Programm: **WaterMiner** (August 2016 – Juli 2019) – Wasserressourcenmanagement in Bergbaufolgelandschaften in Vietnam



Die Verfügbarkeit räumlicher und georeferenzierter Daten wird in den kommenden Jahren explosionsartig ansteigen!

Variety

- Billigerer und einfacherer Zugang zu (immer detaillierteren) Satellitendaten (inkl. Mikro- & Nano)
- Zunehmende Anwendungsideen für unbemannte Flugobjekte (UAV)
- Billigere und leistungsfähigere in-situ Sensoren mit (Nah-)Echtzeit-Datenfernübertragung
- Billigere und leistungsfähigere mobile Sensoren mit (Nah-)Echtzeit-DFÜ, auf Fahrzeugen, an Smartphones gekoppelt, etc.
- Fußgänger-Tracking, Fahrzeug-Tracking, ...
- Internet-of-things, Industry 4.0 usw.
- Volunteered Geographic Information
- Georeferenzierte Social-Media Inhalte

Volume

Velocity

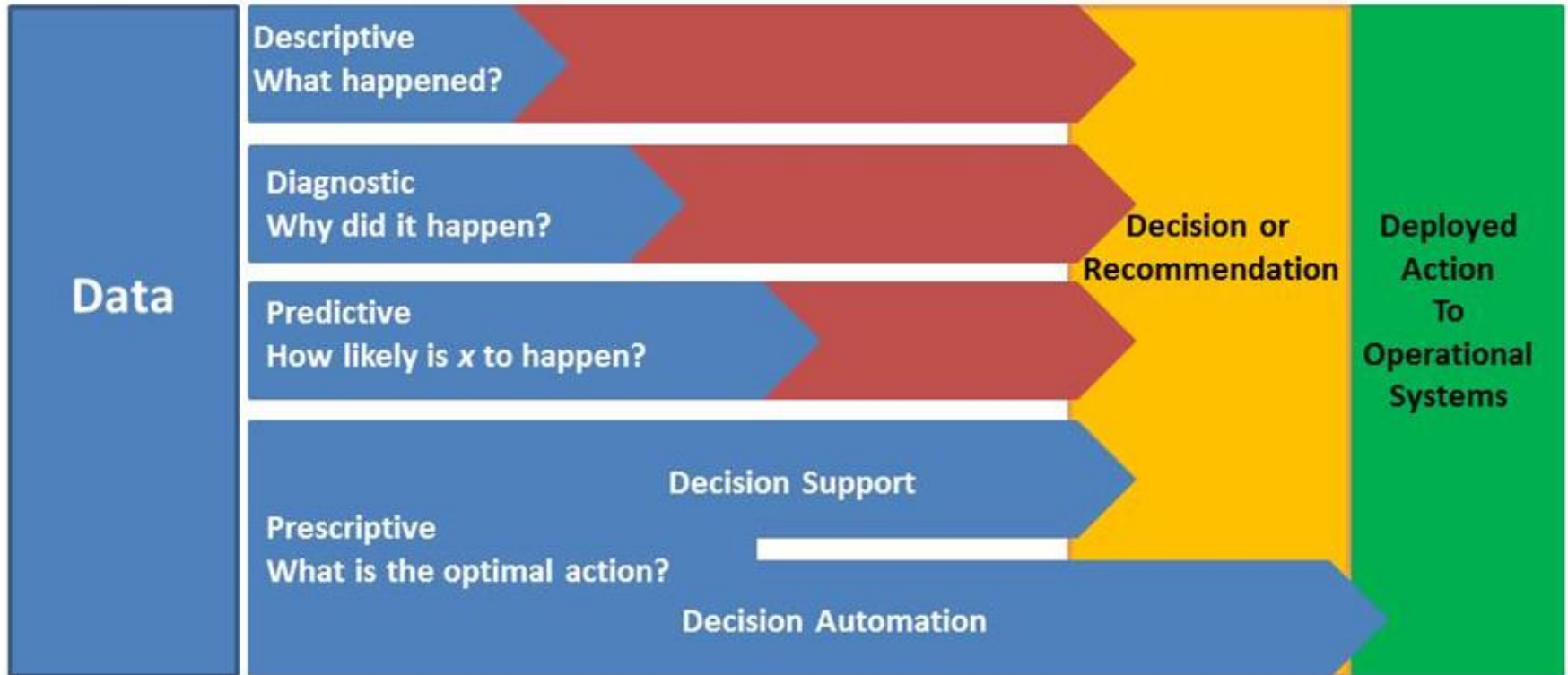
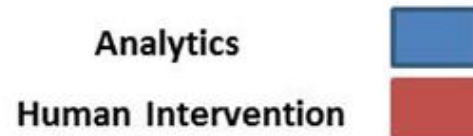
Veracity

Vielversprechende Anwendungsszenarien basieren auf der Analyse großer, heterogener und hochfrequenter räumlicher Daten

- Präzisionslandwirtschaft
- Smart City – Überwachung & Steuerung
- Verkehrsmanagement
- Katastrophenschutz und -rettung
- Smart Energy
- Kontext-spezifische Marketing- und Informationsdienste
- Klimawandel
- ...



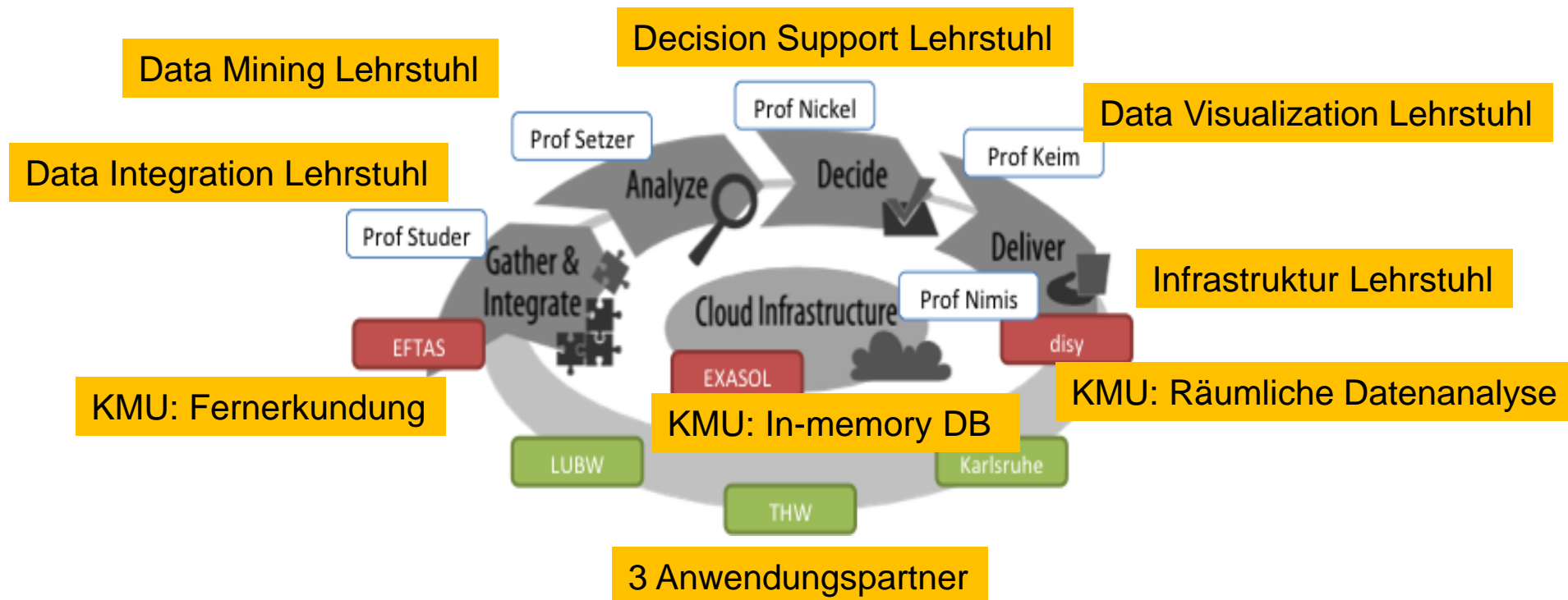
Analytics from Description to Prescription...



After Gartner (September 2013)

Das Projekt BigGIS

- Projekt: BigGIS: Prädiktive und präskriptive Geoinformationssysteme basierend auf hochdimensionalen geotemporalen Datenstrukturen
- Laufzeit: April 2015 – März 2018
- Förderung: Bundesministerium für Bildung und Forschung (BMBF)

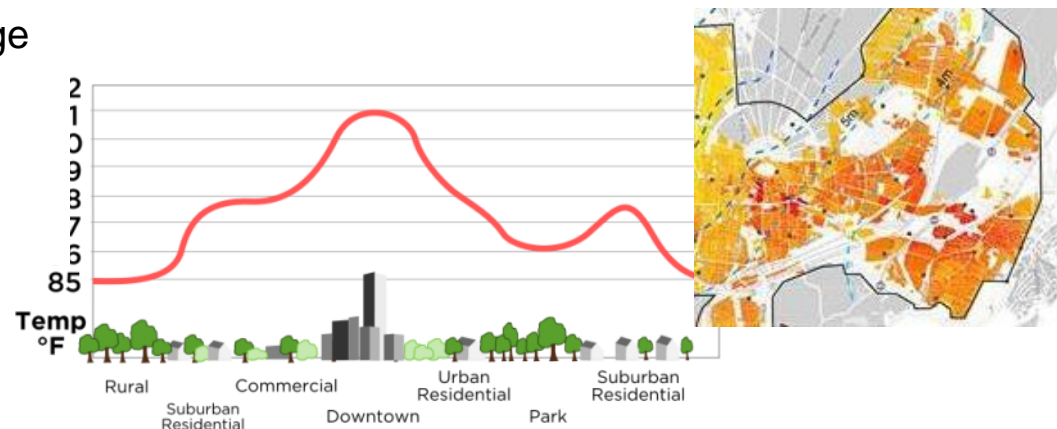


BigGIS Application Scenario 1: Urban Heat Islands

- Context: Urban micro-climate depends on weather, pollution, land-use, urban green, architecture, ...
- Goal: More exact assessment of actual situation and short-term, fine-grained prediction of urban micro-climate (temperature, ozone, PM10, ...)
- Approach: new measurements plus inter-/extrapolation of measurement data
- Applications:
 - Routing people with minimum heat exposure (cp. OpenSense project)
 - Targeted warnings for high-risk groups
 - Warnings for kindergarten, old-age homes, etc.
- Predominant big data characteristics: variety, veracity

Data sources:

- Official topographic and cadastral data
- Thermography aerial survey Karlsruhe
- Normalized Difference Vegetation Index (EnviSAT, Landsat)
- Level-of-detail 2: 3D model Karlsruhe
- Sensors of meteorological service and environment agency (DWD, LUBW)
- Climate data of Karlsruhe University
- Planned: Mobile sensors
- Planned: Participatory sensing
- Planned: Radar data
- Planned: Social media analysis

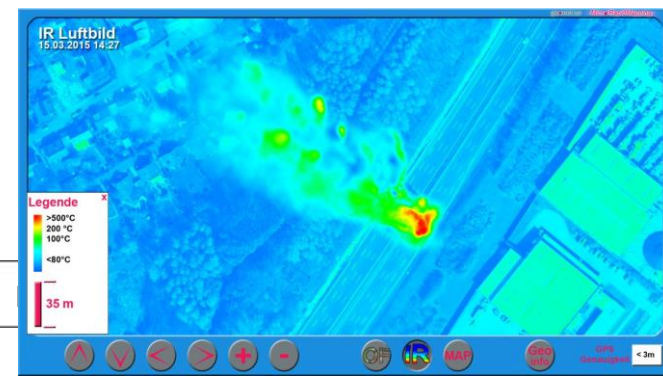


BigGIS Application Scenario 2: Disaster/Emergency Management

- Context: Disaster Management (floods, (wild)fires, chemical accidents, terrorist attacks, ...)
- Goal: Within 15min after the event, have an emergency map for fire brigades – plus continuous updates – plus predictions about further evolution (e.g., movement of cloud of poisonous gas)
- Approach: Combine UAV remote sensing data with background knowledge and in-situ observations; data focus and dimension reduction is key
- Predominant big data characteristics: volume, variety, veracity, (velocity)

Data sources:

- Micro Rapid Mapping: micro flight robot (AiD MC8 Octocopter) with sensors such as RGB camera (Sony Smart Shot IL CE QX1), thermal camera (FLIR Quark 2), Hyperspectral (Cubert UHD 185 Firefly), RTK GPS
- Official topographic + cadastral data: critical infrastructures, endangered population, protected sites, ...
- Crowdmapping + social media content



Micro Rapid Mapping | Need of Fire Brigades | Example for Data Volume

Expected data volume:

Example of a 10 minutes flight...

- Hundreds of frames out of the RGB video stream, the infrared sensor and information cubes of the UHD Firefly:



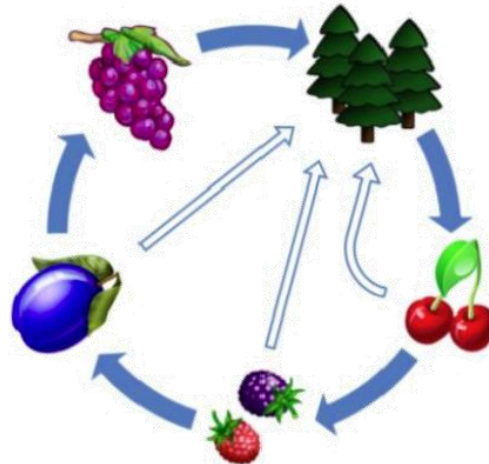
picture credits: Freiwillige Feuerwehr Werne

	Number Pixel X	Number Pixel Y	Color Depth (Bit)	Frames/s or Cubes/s	Number of spectral Channels	Recording time (min)	GByte
HD-Video							
Full-HD-24	1.920,00	1.080,00	24,00	24,00	1,00	10,00	83,43
Full-HD-50	1.920,00	1.080,00	24,00	50,00	1,00	10,00	173,81
hyperspectral							
UHD Firefly	1.000,00	1.000,00	12,00	1,00	125,00	10,00	104,77
UHD Firefly	1.000,00	1.000,00	12,00	5,00	125,00	10,00	523,87

BigGIS Application Scenario 3: Invasive Species



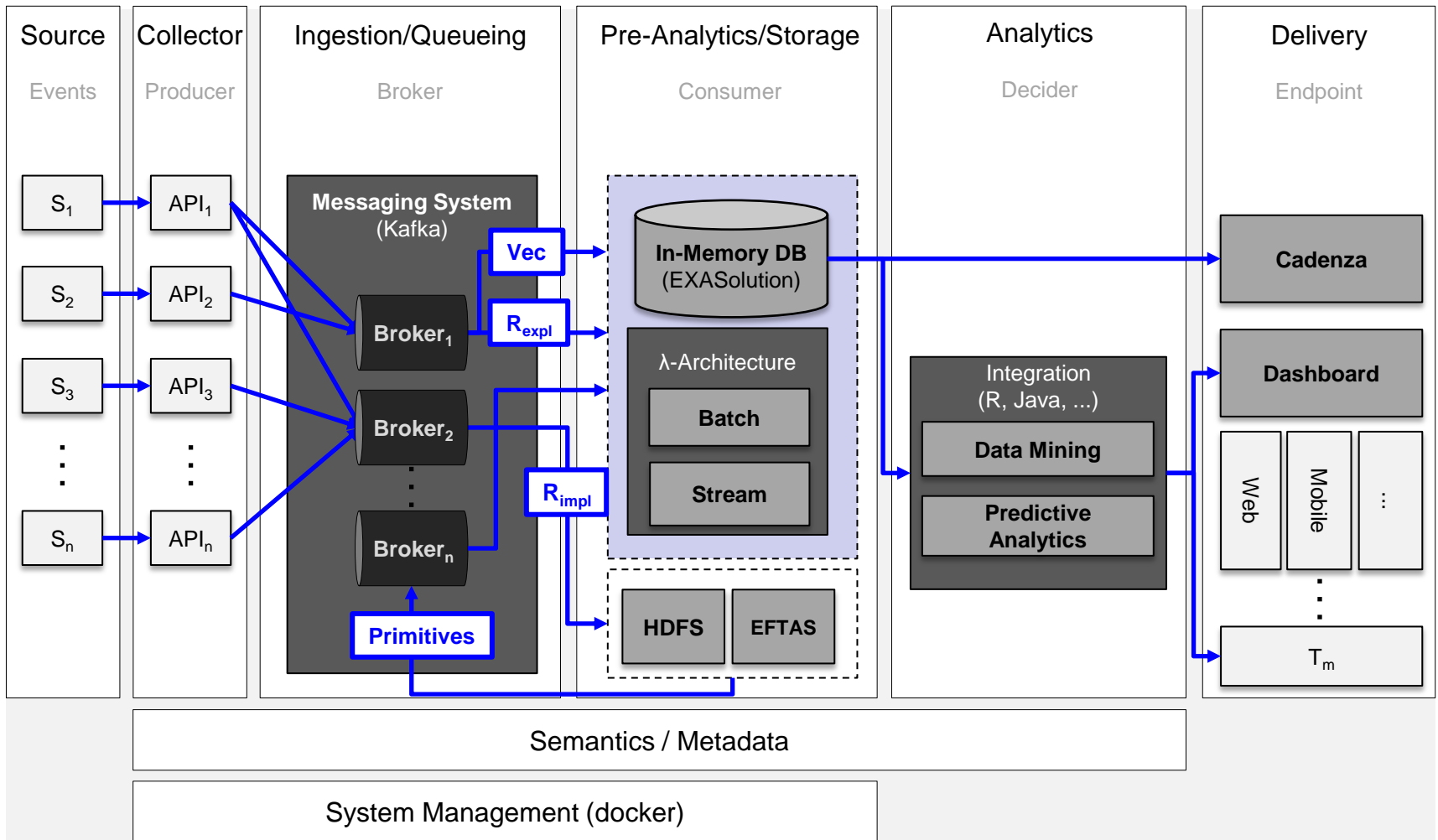
- Context: Invasive species may create serious economic damages or health problems
- Example: *Drosophila suzukii*
- Goal: understand and predict the distribution patterns and dynamics of invasive species depending on vegetation, weather etc.
- Approach: learn distribution mechanisms from historic data
- Predominant big data characteristics: variety, veracity



Data sources:

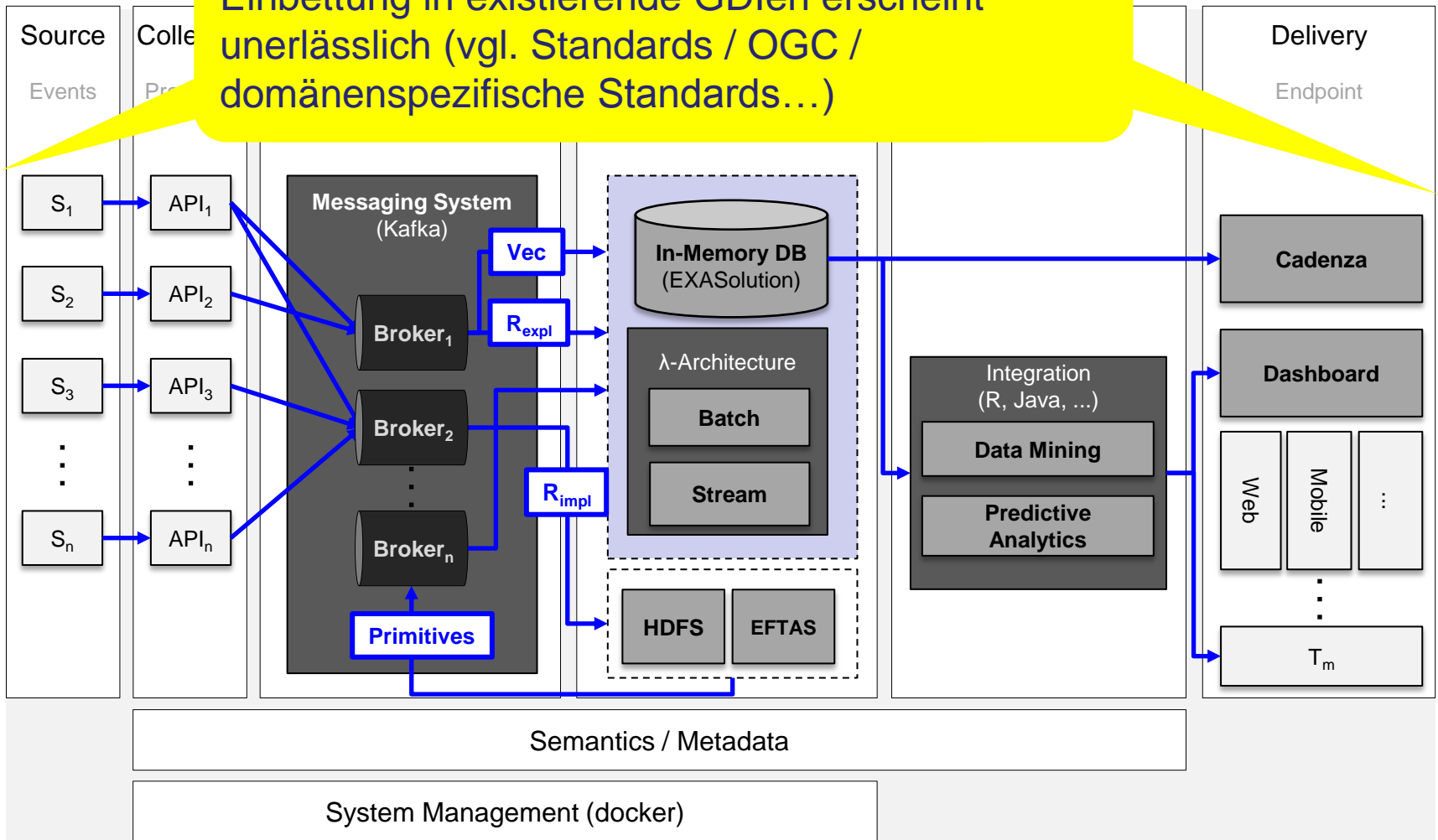
- Land-use and land-cover data (as fine-grained as possible)
- Official observation data of species (collected by environment agencies, e.g. by traps)
- Weather observations and weather forecasts (as fine-grained as possible)
- Crowdmapping for some species

Überlegungen zur Stufen-Architektur für Geodaten(ströme)



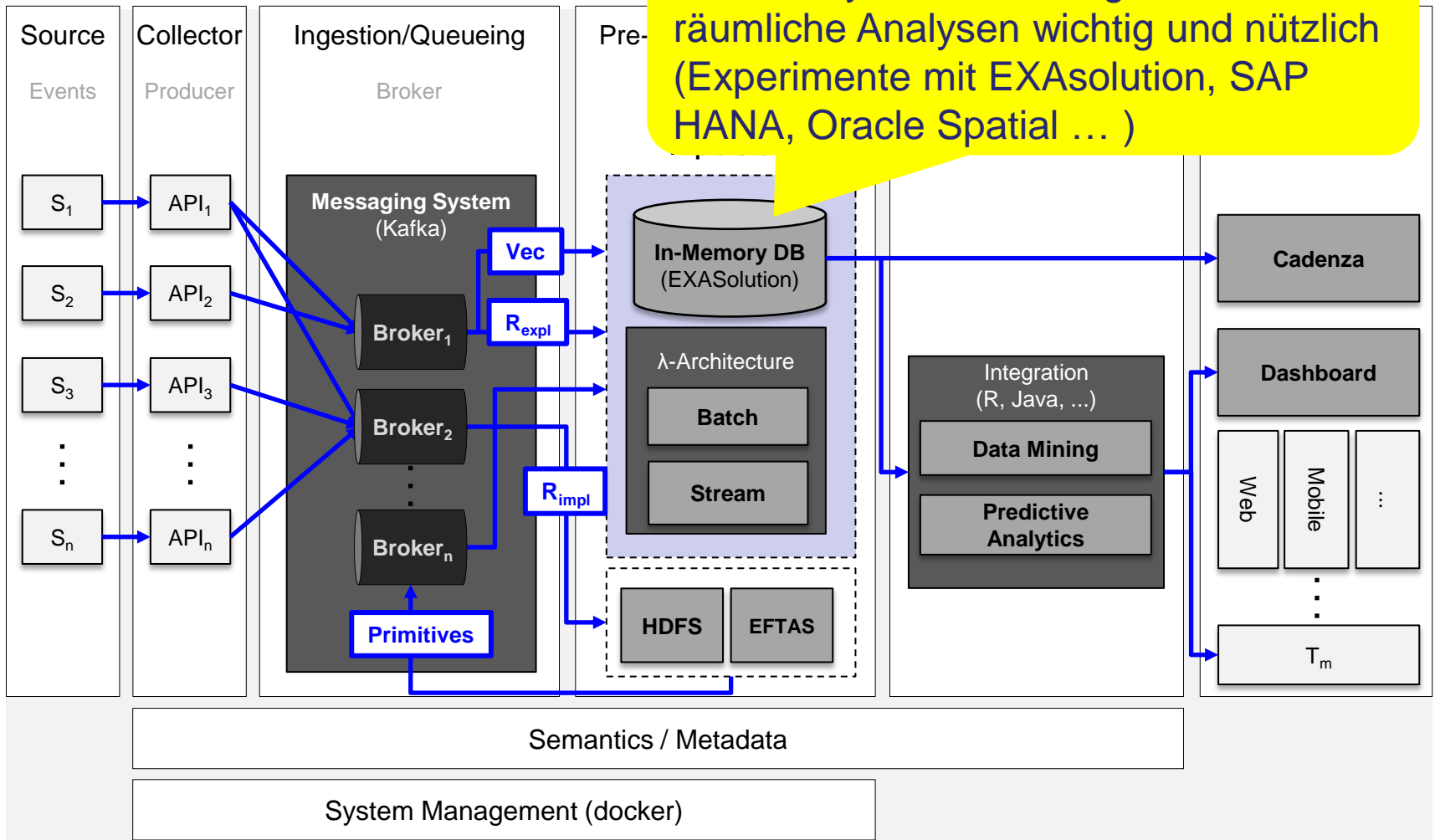
Überlegungen zur Stufen-Architektur für Geodaten(ströme)

Einbettung in existierende GDien erscheint unerlässlich (vgl. Standards / OGC / domänenspezifische Standards...)

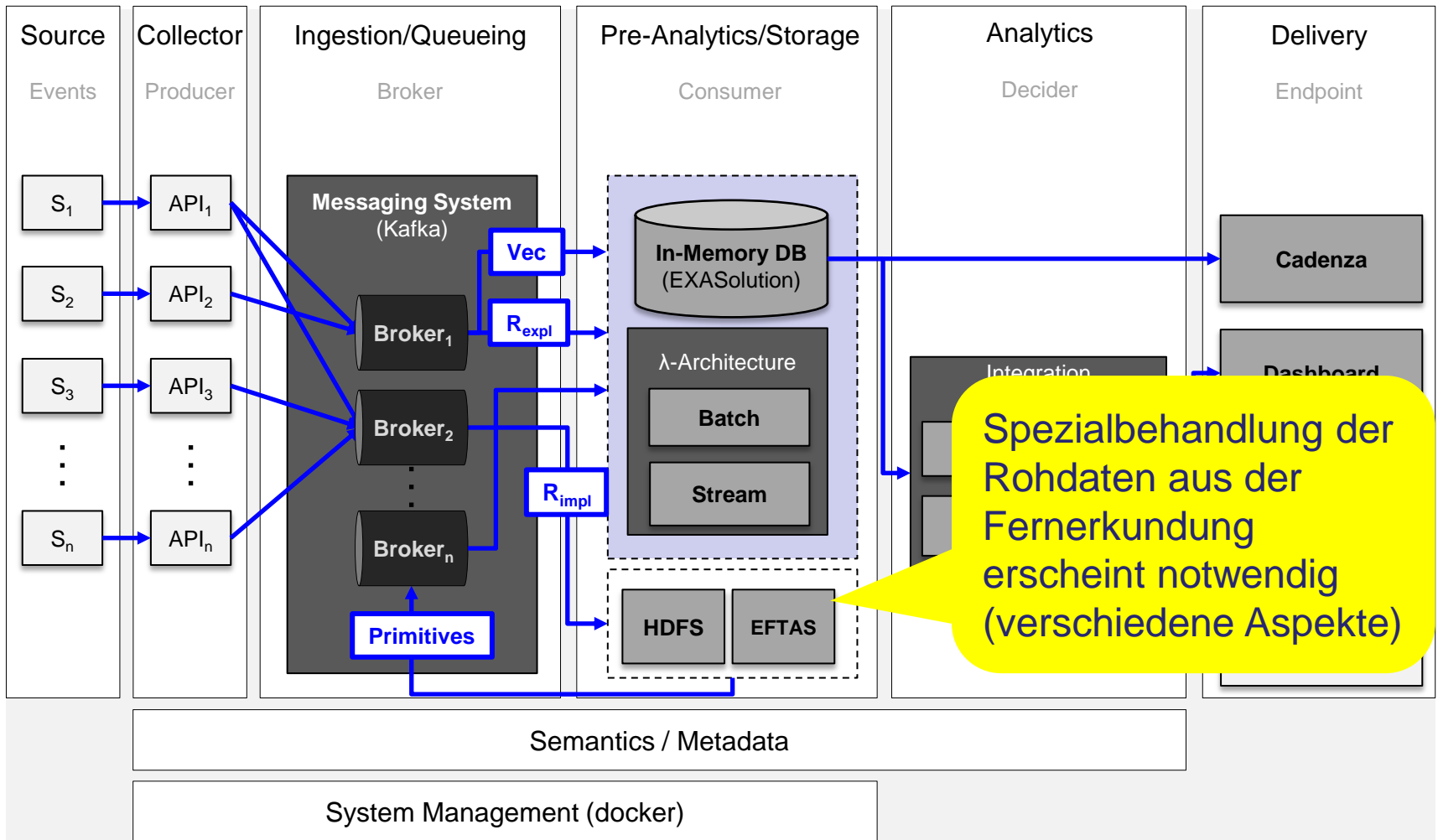


Überlegungen zur Stufen-Architektur für Geodaten(ströme)

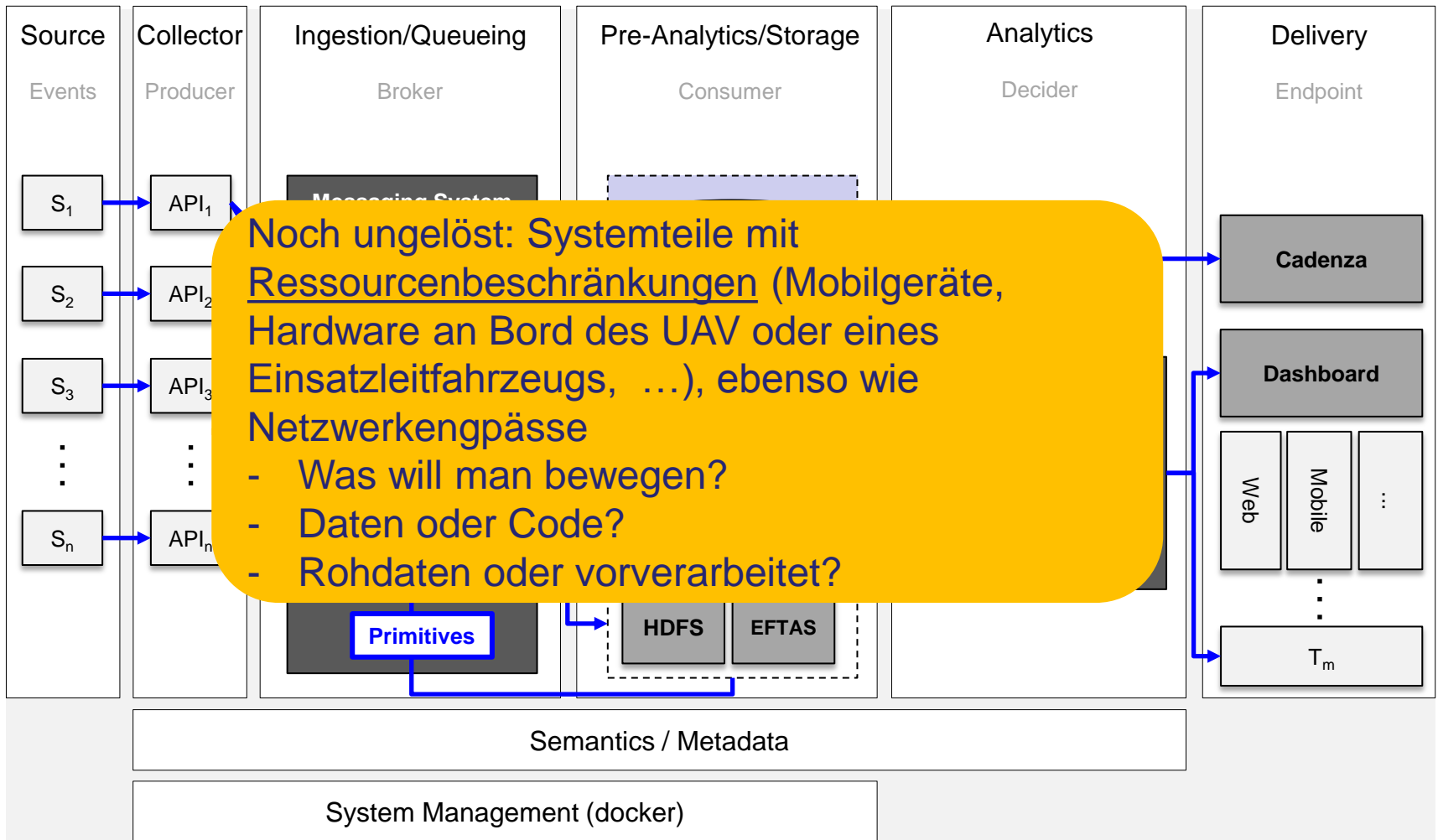
In-Memory DB Technologie für räumliche Analysen wichtig und nützlich (Experimente mit EXASolution, SAP HANA, Oracle Spatial ...)



Überlegungen zur Stufen-Architektur für Geodaten(ströme)



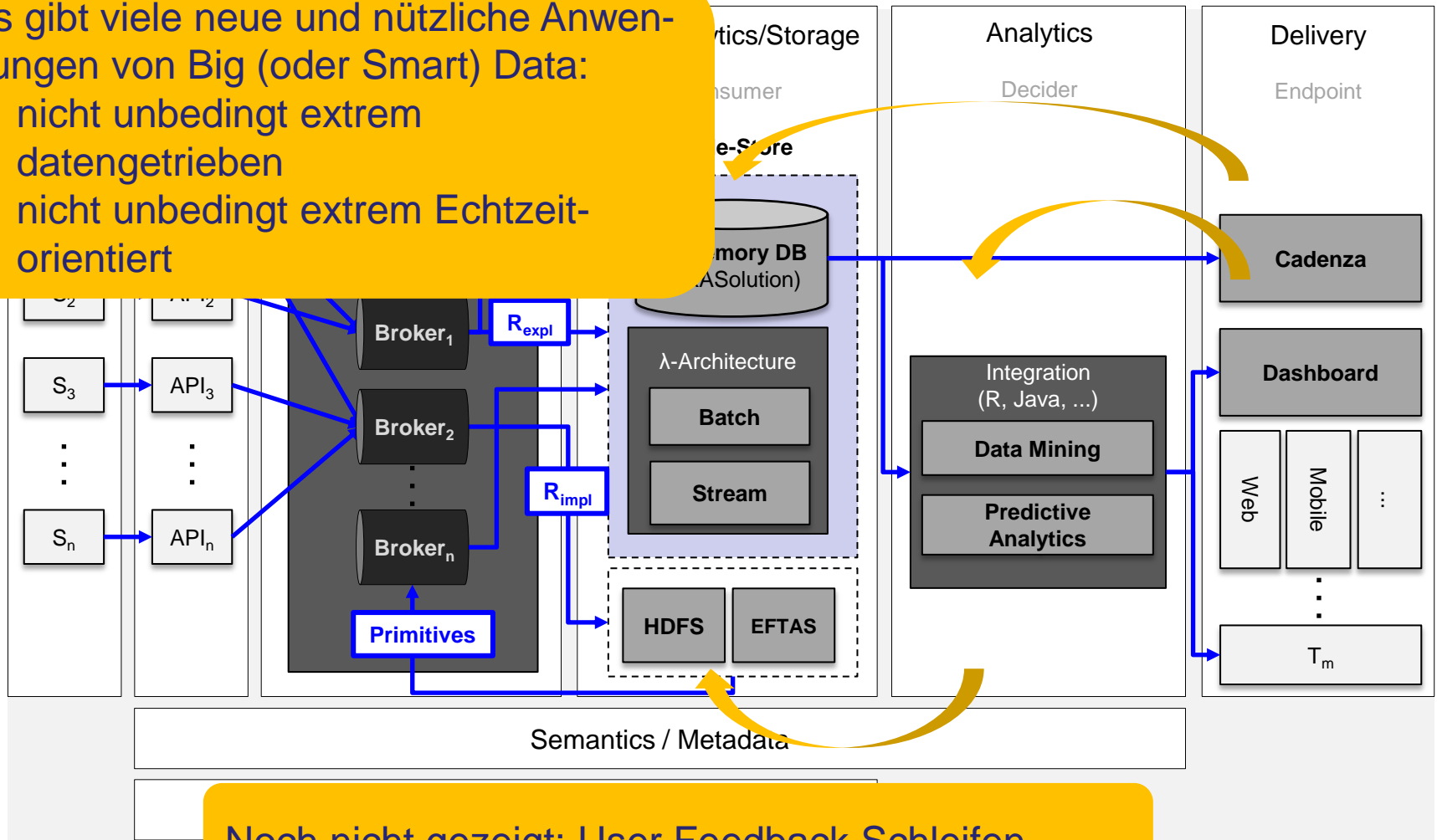
Überlegungen zur Stufen-Architektur für Geodaten(ströme)



Überlegungen zur Stufen-Architektur für Geodaten(ströme)

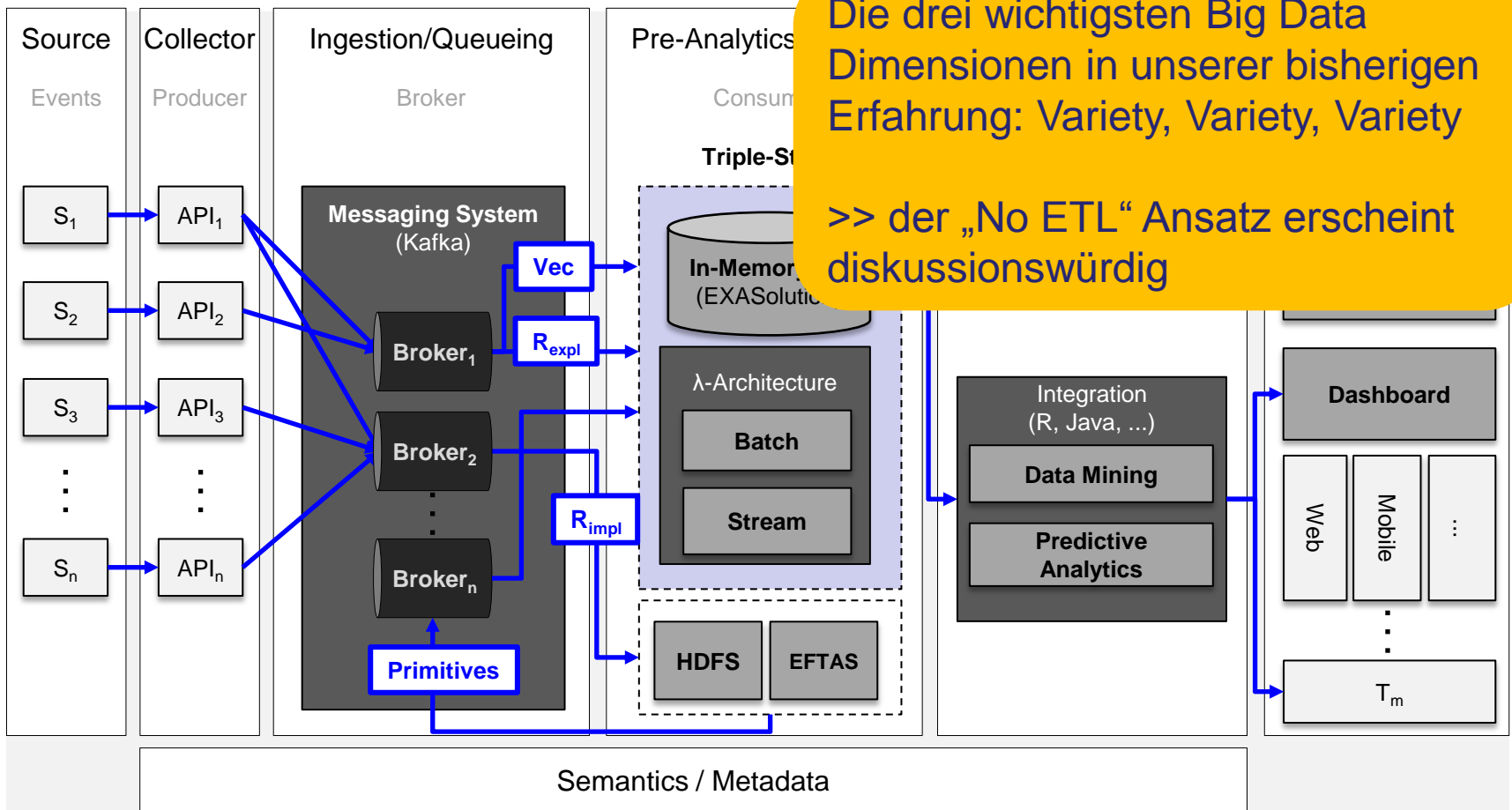
Es gibt viele neue und nützliche Anwendungen von Big (oder Smart) Data:

- nicht unbedingt extrem datengetrieben
- nicht unbedingt extrem Echtzeit-orientiert



Noch nicht gezeigt: User Feedback Schleifen

Überlegungen zur Stufen-Architektur für Geodaten(ströme)



Die drei wichtigsten Big Data Dimensionen in unserer bisherigen Erfahrung: Variety, Variety, Variety

>> der „No ETL“ Ansatz erscheint diskussionswürdig

Hier noch nicht gezeigt:

- Semantische Harmonisierung, Geocoding, ...
- Idealerweise, automatisch konfiguriert, mithilfe semantischer Metadaten über Quellen und Algorithmen

- Nach unserer Erfahrung ist im Geodatenbereich heute **Variety** der Schlüssel
 - Volume entsteht mit zunehmenden Rohdaten aus der Fernerkundung
 - Velocity wird mit mehr und mehr Sensorik kommen
 - Aber auch heutige Anwendungen sind bereits herausfordernd genug !
- Big Data Technologien können **wertvolle** Bausteine liefern (In-Memory DB, verteilte Speicherung und Verarbeitung, Virtualisierung)
- Die optimale **Einbettung** räumlicher Big Data Anwendungen in bestehende Hardware/Software Landschaften erfordert noch Ideen und Erfahrung
- Es gibt eine große „**Usability Gap**“ zwischen Rohdaten / Domänenwissen und dem technischen Wissen zu Machine Learning und Decision Support
- Dankbare Themen für die Algorithmiker:
 - ML mit dynamisch veränderlichen räumlichen Aggregationen
 - Spatial Complex-Event Processing

Vielen Dank für Ihre Aufmerksamkeit !

Dr. Andreas Abecker
Dipl.-Inform.
Leiter Innovationsmanagement



Disy Informationssysteme GmbH

Ludwig-Erhard-Allee 6
76131 Karlsruhe
www.disy.net

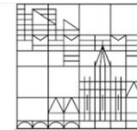
Tel. +49 721 16006-256
Fax +49 721 16006-05
andreas.abecker@disy.net



www.disy.net



Universität
Konstanz



<http://biggis-project.eu/>

- Das Projekt BigGIS wird ko-finanziert vom Bundesministerium für Bildung und Forschung (BMBF)
- Dankeschön für Ideen, Graphiken und Folien an:
 - Dr. Bodo Bernsdorf, EFTAS GmbH
 - Julian Bruns, FZI Forschungszentrum Informatik am Karlsruher Institut für Technologie

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung