

Zurück in die Zukunft - Datenstrategien für Zeitreisende

Michael Brösdorf | InterSystems



DATENSTRATEGIEN FÜR ZEITREISENDE



A digital interface for time travel is displayed against a starry background. It features two main sections: 'DESTINATION TIME' and 'PRESENT TIME'. Each section includes a circular dial for months, a digital display for days, a digital display for years, and a digital display for hours and minutes. The 'DESTINATION TIME' section shows 'JAN 01 2018' and '09 22'. The 'PRESENT TIME' section shows 'NOV 22 2023' and '01 00'. There are also 'AM' and 'PM' indicators for each time section.

Section	Month	Day	Year	Hour	Minute
DESTINATION TIME	JAN	01	2018	09	22
PRESENT TIME	NOV	22	2023	01	00

Von Anekdoten zu Fakten



Fragen:

- nur Zufall?
- wie kommen wir an die Daten (hier: Herzfrequenz, codierte Diagnosen)?
- was heißt „erhöhte Herzfrequenz“ genau?
- in welchem zeitlichen Abstand suchen wir nach Korrelationen?

Antwort: eine Zeitreise in die (Daten-)Vergangenheit und zurück in die Zukunft.



Was sind Zeitreihendaten?

“alles mit einem Zeitstempel“:

- wiederholt auftretende Daten über einen Sachverhalt
- ...oder auch das Nicht-auftreten bestimmter Daten
- Zeitbezug (Zeitstempel oder Reihenfolge)
- potentiell endlos
- (alternativ: „Streaming-Daten“)



Was sind Zeitreihendaten?

Ein Datenpunkt repräsentiert zum Zeitpunkt x :

- eine Anzahl (z.B. Herzfrequenz) oder das Auftreten eines Ereignisses (Event)
- diskrete Messung eines kontinuierlichen Wertes (z.B. Körpertemperatur)
- einen kodierten Sachverhalt (z.B. eine Diagnose)
- Multimedia (z.B. ein Bild als Bestandteil eines Videos)
- kann Metadaten enthalten



Was sind Zeitreihendaten

Praktisch alle Daten können (auch) als Zeitreihe interpretiert werden:

- Sehr wenige, sehr langsam anfallende Daten (z.B. Krankenhausaufenthalte eines Patienten)
- Gelegentlich anfallende Daten (z.B. Labor- oder Diagnosedaten eines Patienten während eines Krankenhausaufenthaltes)
- Häufig anfallende Daten (z.B. regelmäßig von Wearables erfasste Vitaldaten)
- Hochfrequente Daten (Waveforms, Audio-/Videosignale etc.)



Was sind Zeitreihendaten

Charakterisierung von Zeitreihen, z.B.:

- Univariat / Multivariat (ein bzw. mehrere Variablen)
- Mehrdimensional (eine Variable an verschiedenen Orten)
- Äquidistanz (Messwerte im gleichen zeitlichen Abstand)
- Lücken
- Event time vs. Processing time (Eingangsreihenfolge!)
- Relevanz von Uhrzeit, Abstand, Reihenfolge
- ...

Herausforderungen bei der Handhabung von Zeitreihendaten



(insbesondere bei der Zusammenführung von Daten aus unterschiedlichen Quellen)

- Frequenz
- Latenz
- Reihenfolge
- Vollständigkeit
- Überlappungen
- Datenmenge
- Korrektheit



Auswertung von Zeitreihendaten

Typische Anwendungsfälle (inhaltlich):

- Erkennung von Mustern im zeitlichen Verlauf
- Korrelationen zwischen Zeitreihen und/oder diskreten Ereignissen

Typische Anwendungsfälle (zeitlich):

- Identifikation von Mustern/Korrelationen auf historischen Daten ← Data Science
- Reagieren auf Muster/Korrelationen in (Near) Real-Time-Daten ← Produktivbetrieb

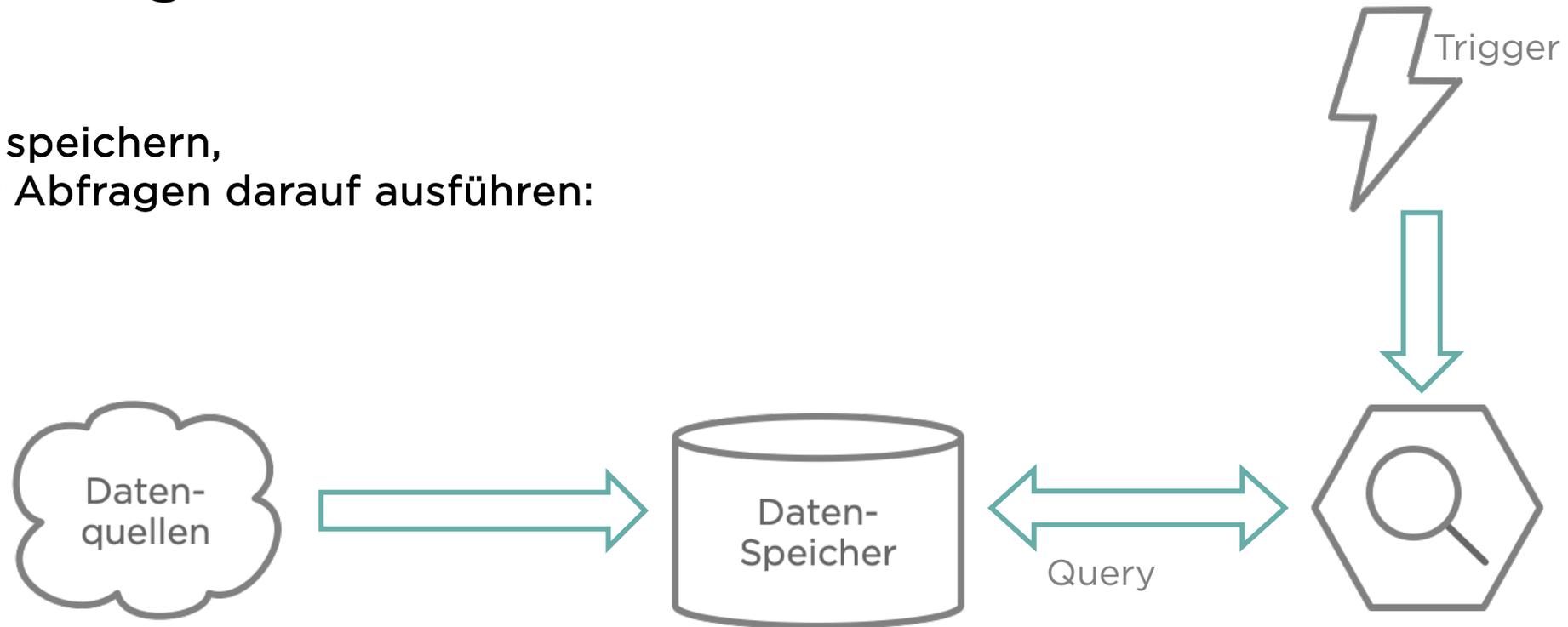
Grundsätzliche Vorgehensweisen:

- Daten speichern, dann Abfragen darauf ausführen ← "Data at Rest"; Pull
- Abfragen speichern, Daten permanent durchlaufen lassen ← „Data in Motion“; Push

Auswertung von Zeitreihendaten



Daten speichern,
später Abfragen darauf ausführen:

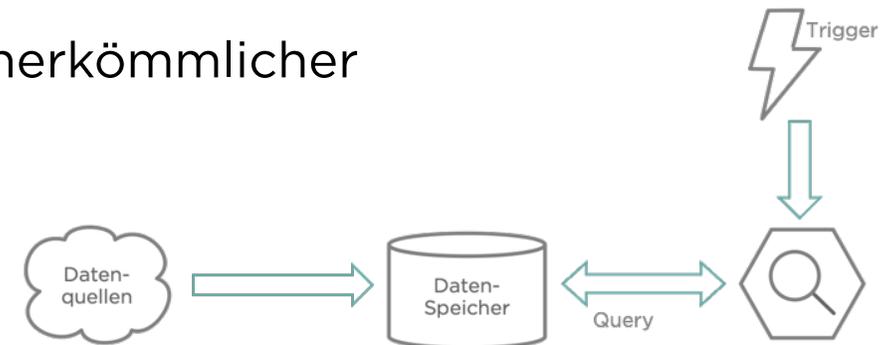


Auswertung von Zeitreihendaten



Daten speichern,
später Abfragen darauf ausführen:

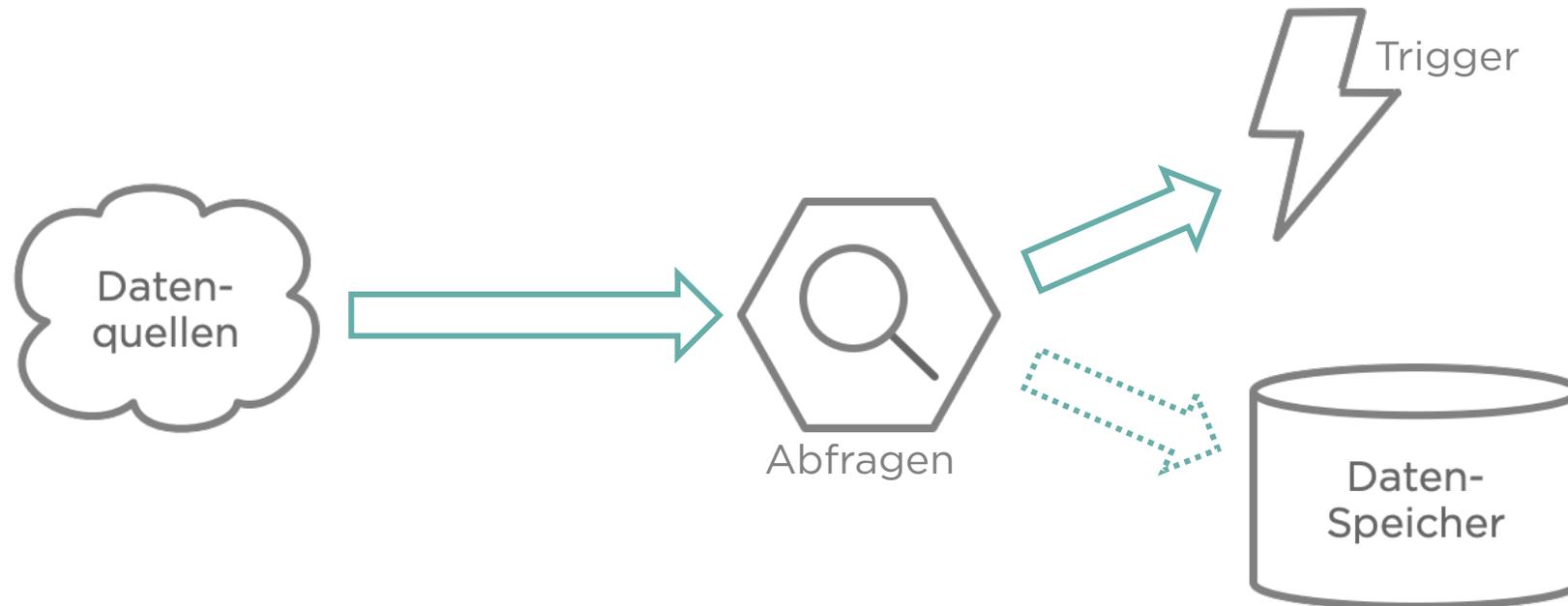
- Abfragen leicht auf dem vollen Datenbestand ausführbar
- Abfragen jederzeit wiederholbar
- Auswertung in Echtzeit nur bedingt möglich
- Berücksichtigung zeitlicher Aspekte bei Verwendung herkömmlicher Werkzeuge (SQL, BI etc.) vergleichsweise aufwendig



Auswertung von Zeitreihendaten



Abfragen speichern,
Daten kontinuierlich durchlaufen lassen:

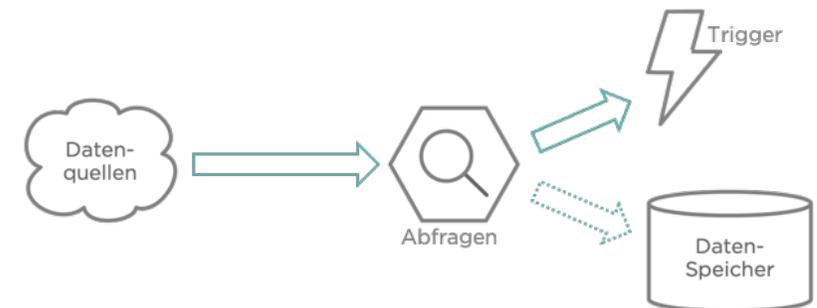


Auswertung von Zeitreihendaten

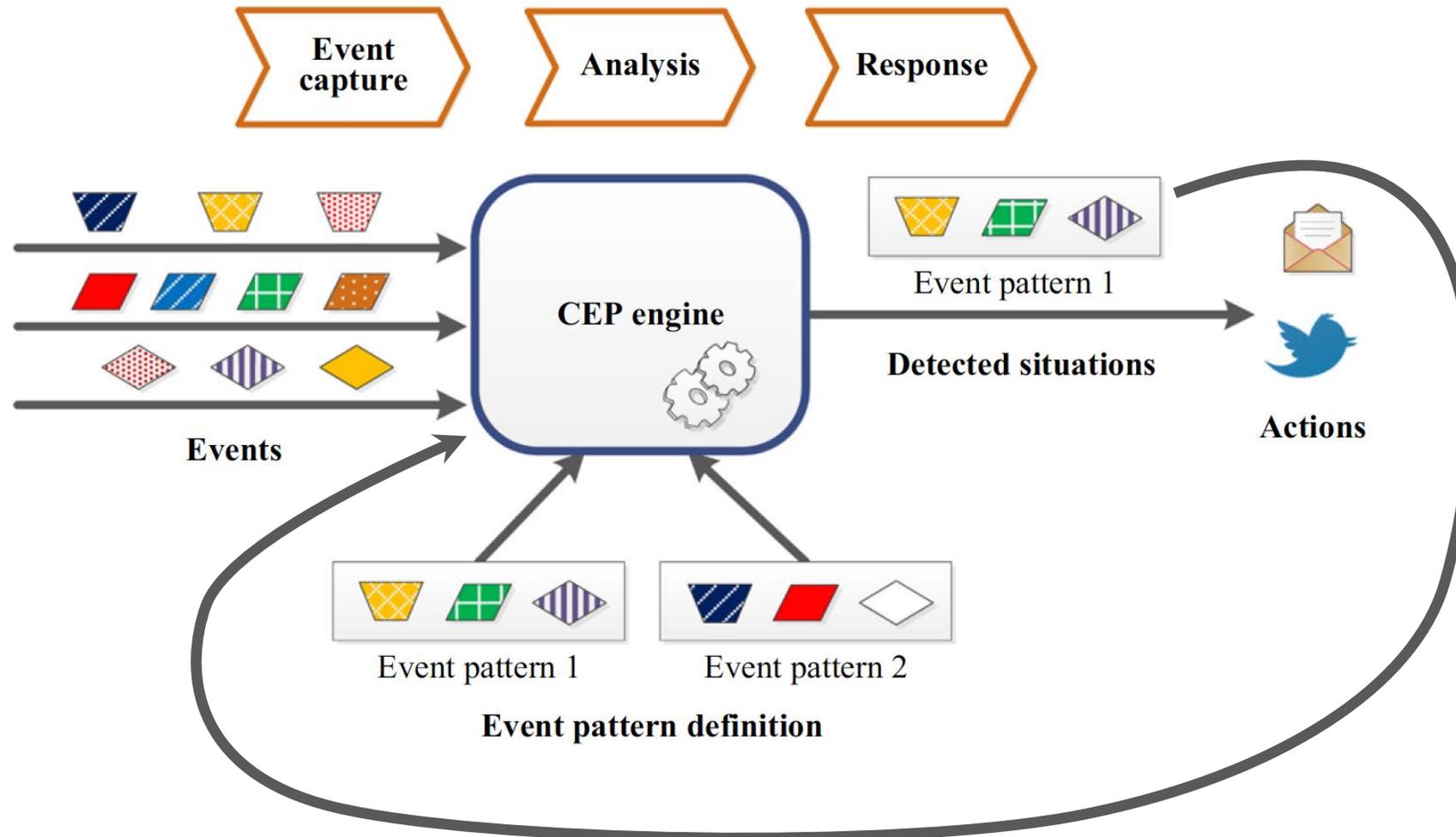


Abfragen speichern,
Daten kontinuierlich durchlaufen lassen:

- Gut geeignet für Echtzeit-Anwendungen
- Berücksichtigung zeitlicher Aspekte vergleichsweise einfach
- gut verteilbar
- Abfragen auf dem vollen Datenbestand aufwendig/teuer
- Abfragen nicht ohne Weiteres wiederholbar
- man sollte wissen, wonach man sucht



Auswertung von Zeitreihendaten - Complex Event Processing



Auswertung von Zeitreihendaten



Grundsätzliche Vorgehensweisen:

- Daten speichern, dann Abfragen darauf ausführen
- Abfragen speichern, Daten permanent durchlaufen lassen

→ Beides ist sinnvoll, ggf. auch parallel

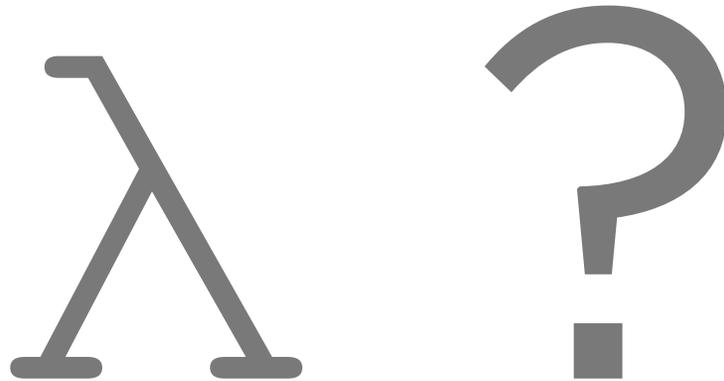
Kombination beider Welten / Architekturansätze



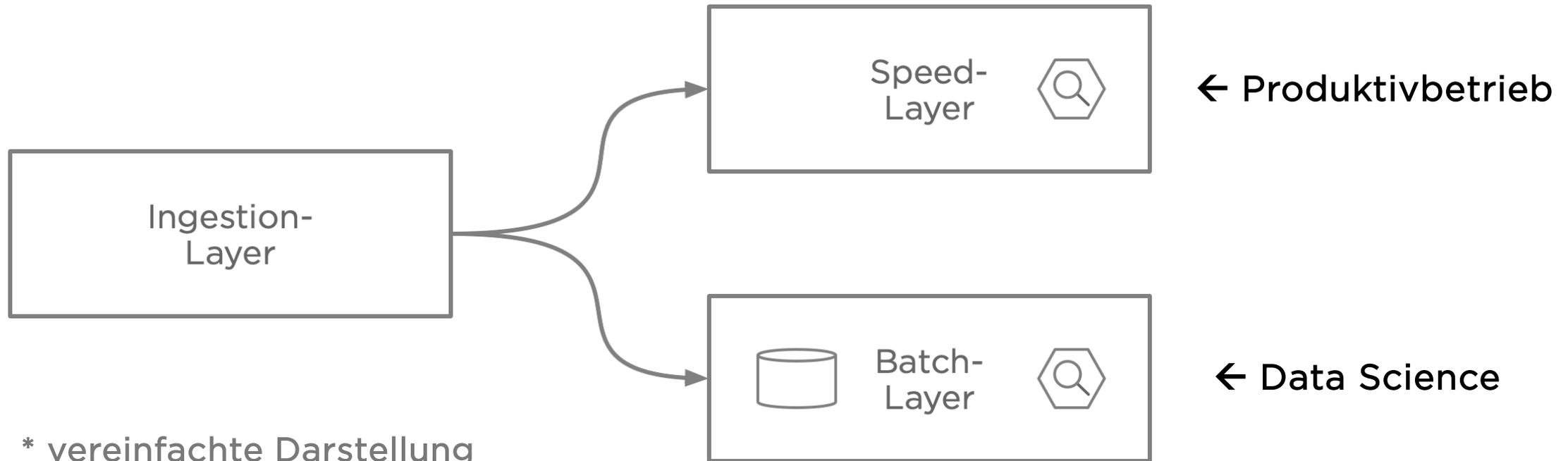
Einige Architekturansätze:

- Anlehnung an die Lambda-Architektur
- Kappa-Architektur
- Smart Data Fabric

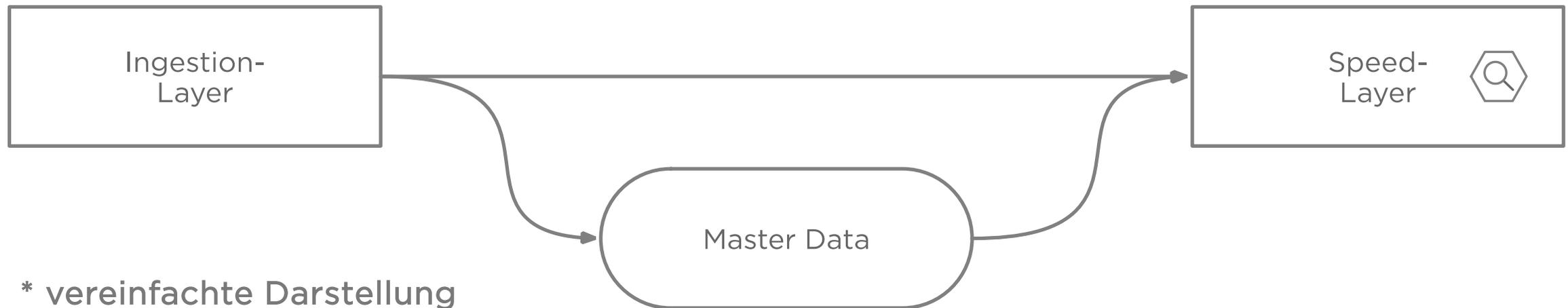
Kombination beider Welten - Lambda-Architektur



Kombination beider Welten - „Lambda-Architektur“*

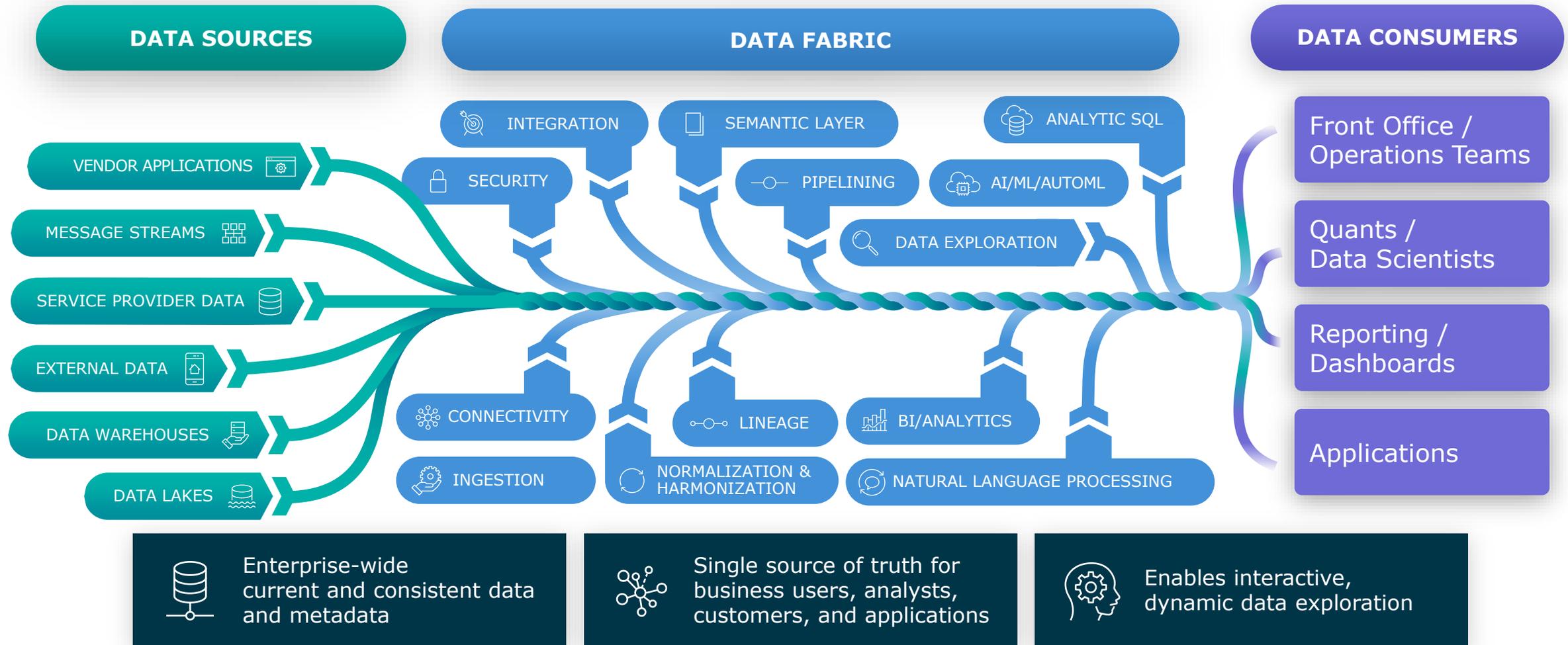


Kombination beider Welten - „Kappa-Architektur“*



* vereinfachte Darstellung

Kombination beider Welten - Smart Data Fabric



Bausteine für die Auswertung von Zeitreihendaten



Auswahl typischer Verfahren:

- Selektion/Pufferung jeweils zu betrachtender Datenpunkte
- Statistische Verfahren
- Chartanalyse
- Machine Learning
- Künstliche Intelligenz
- ...

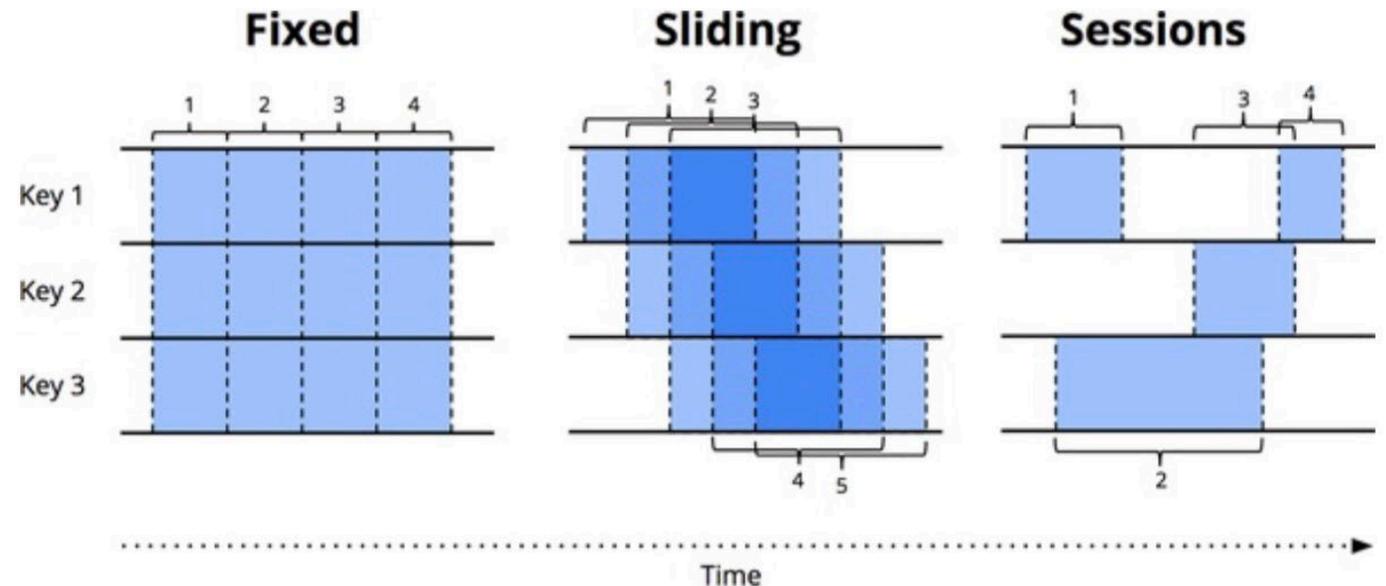
Windowing



Welche Datenpunkte sollen betrachtet werden?

Mögliche Definitionskriterien:

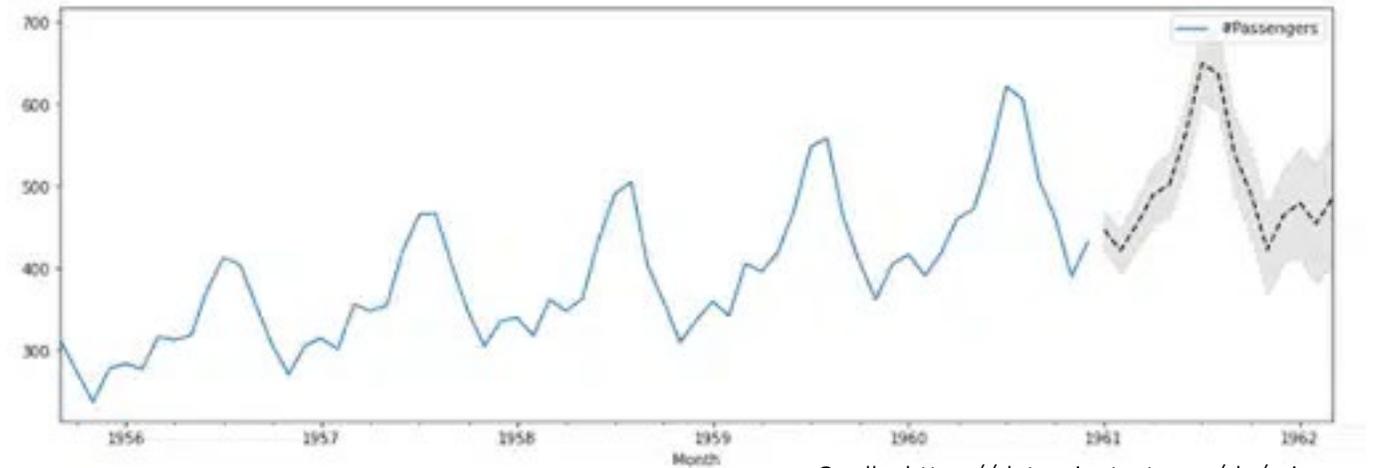
- Zeitintervalle
- Anzahl Datenpunkte
(fest oder geclustert/*Sessions*)
- Aufeinanderfolgend
(*Fixed*)
- Gleitend
(*Sliding*)
- ...



Statistische Verfahren



- Glättung ← Vorsicht!
- Trendanalyse
- Korrelationsanalysen (Lag, Autokorrelation, Kreuzkorrelation etc.)
- ARIMA (*Auto-Regressive Integrated Moving Average*)
- ...



Quelle: <https://datascientest.com/de/arma>

Statistische Verfahren – Glättung

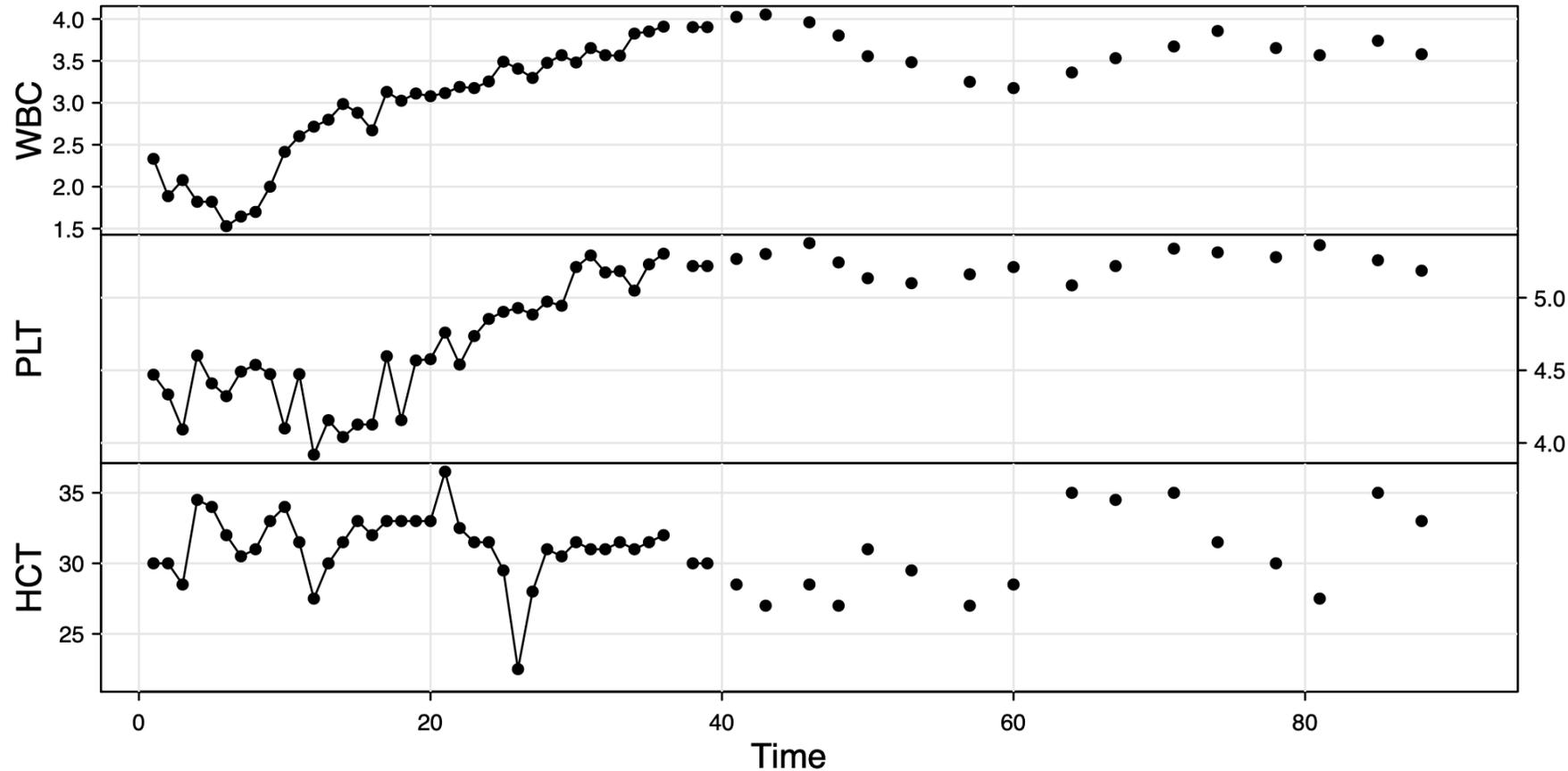
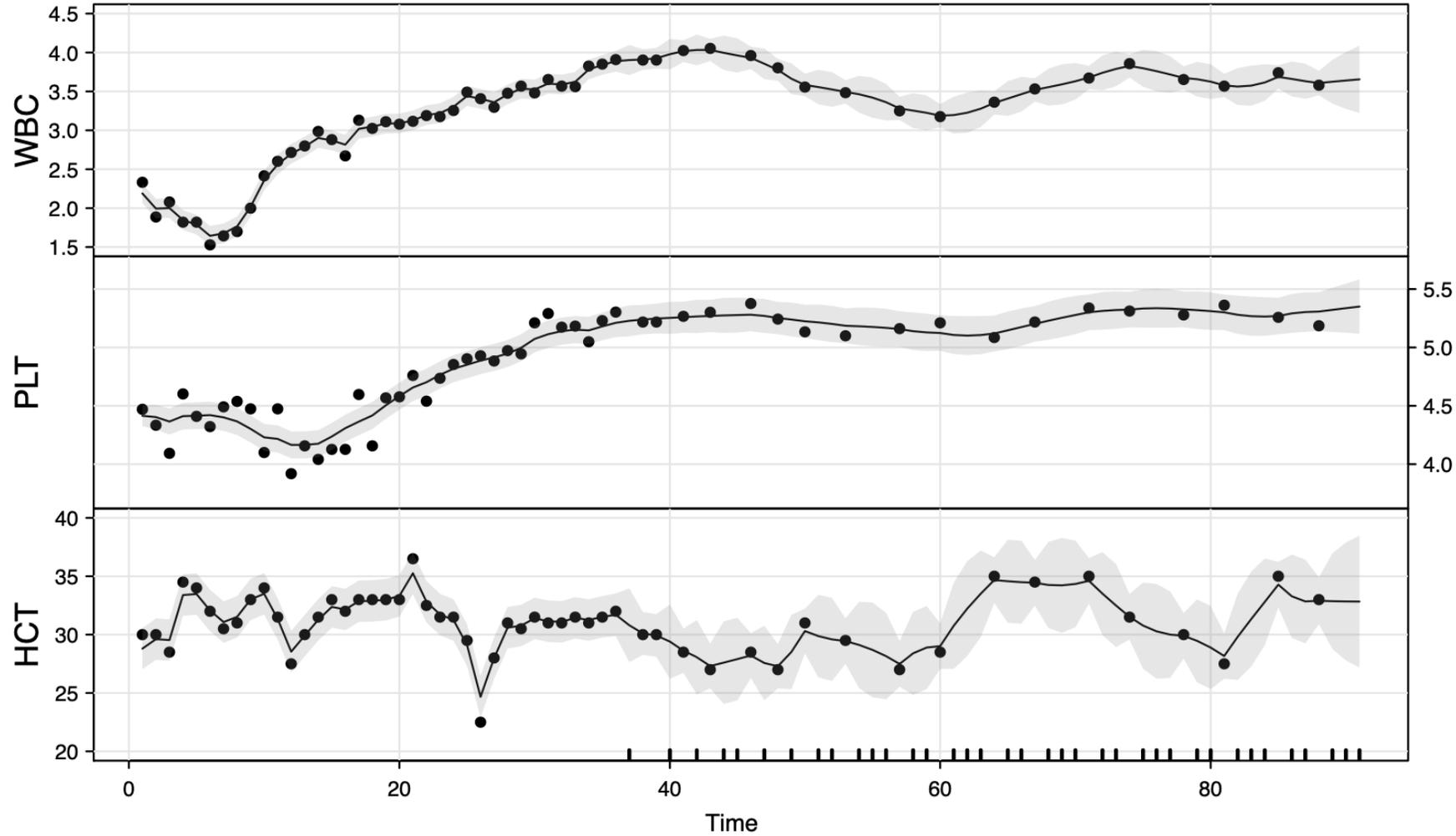


Fig. 6.2. Longitudinal series of monitored blood parameters, log (white blood count) [WBC], log (platelet) [PLT], and hematocrit [HCT], after a bone marrow transplant (n = 91 days).

Statistische Verfahren - Glättung

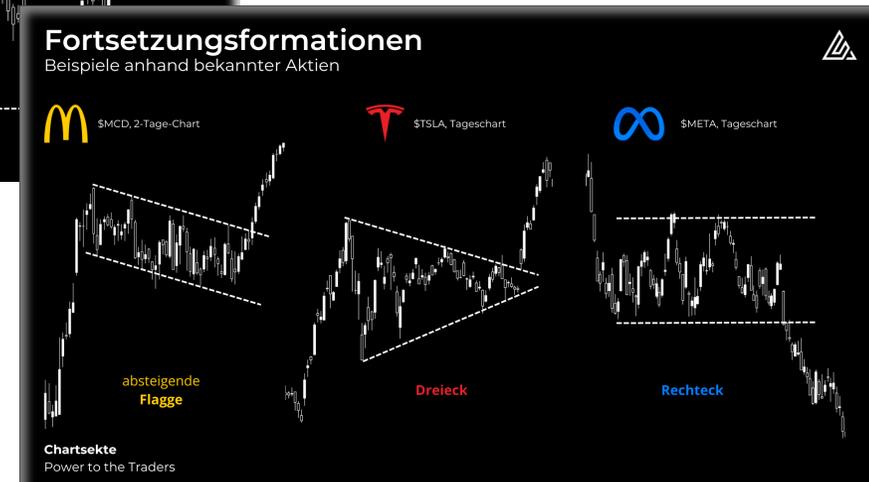
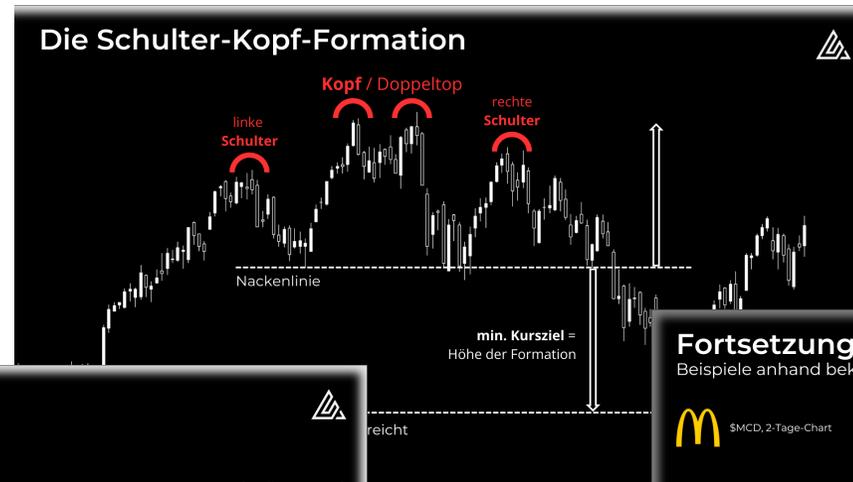
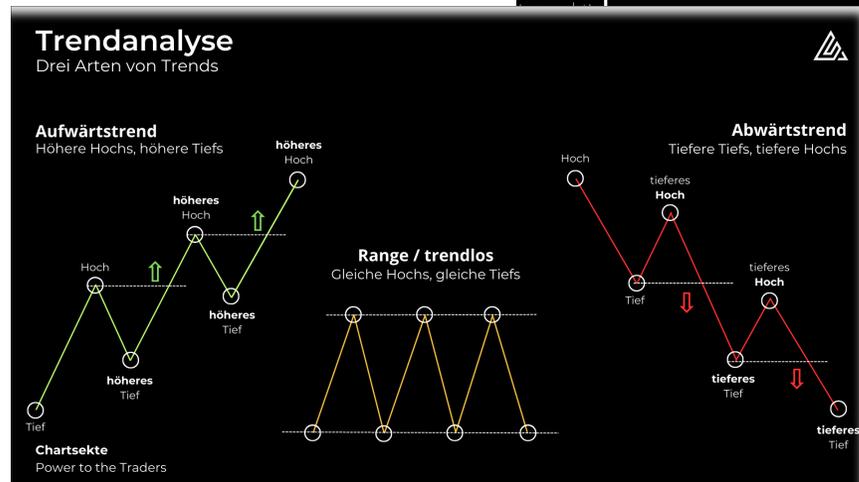


Chartanalyse



Grundannahme: alle für eine Vorhersage nötigen Informationen sind in den Werten der Zeitreihe enthalten (Aktienkurse)

- Trendanalyse
- Mustererkennung
- ...



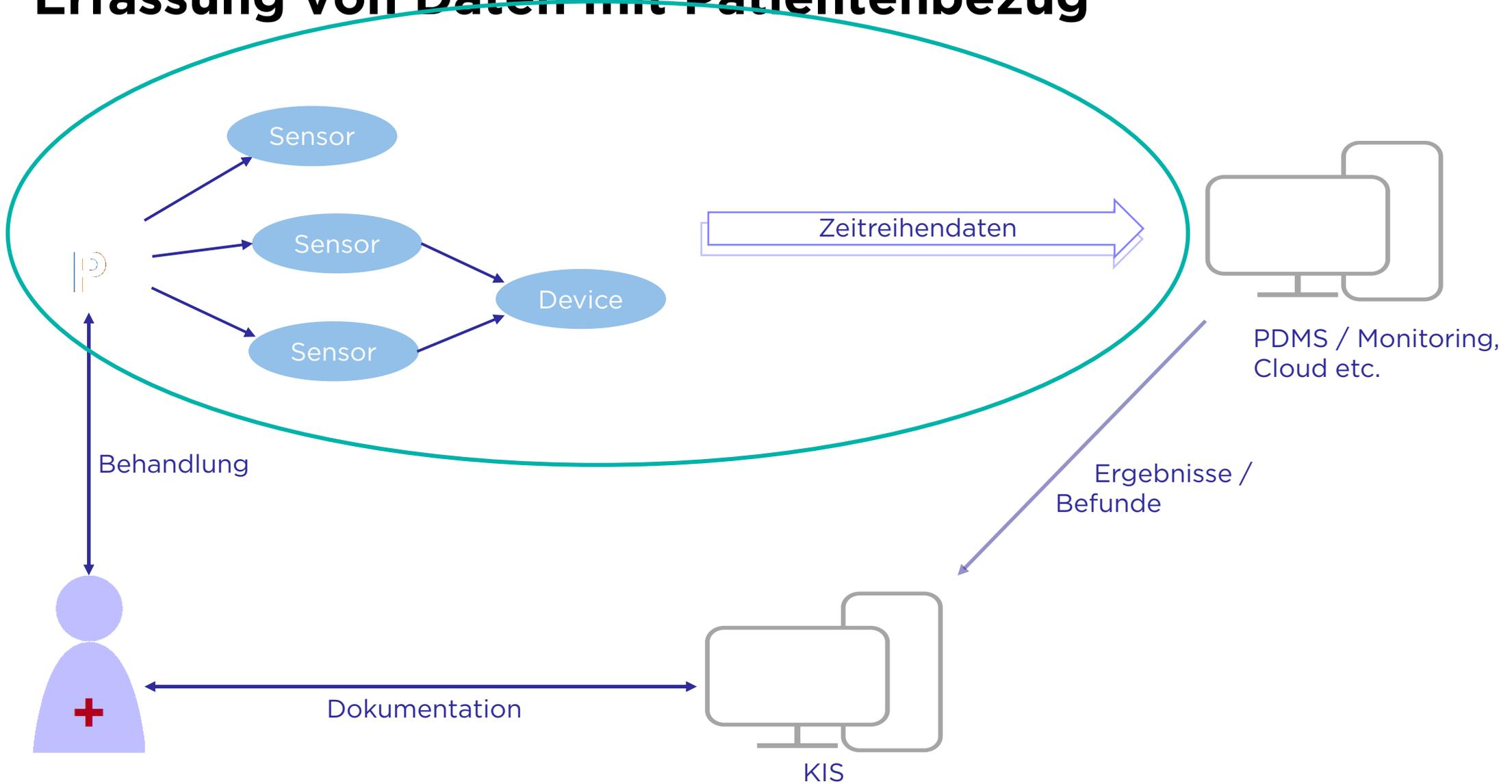
Machine Learning



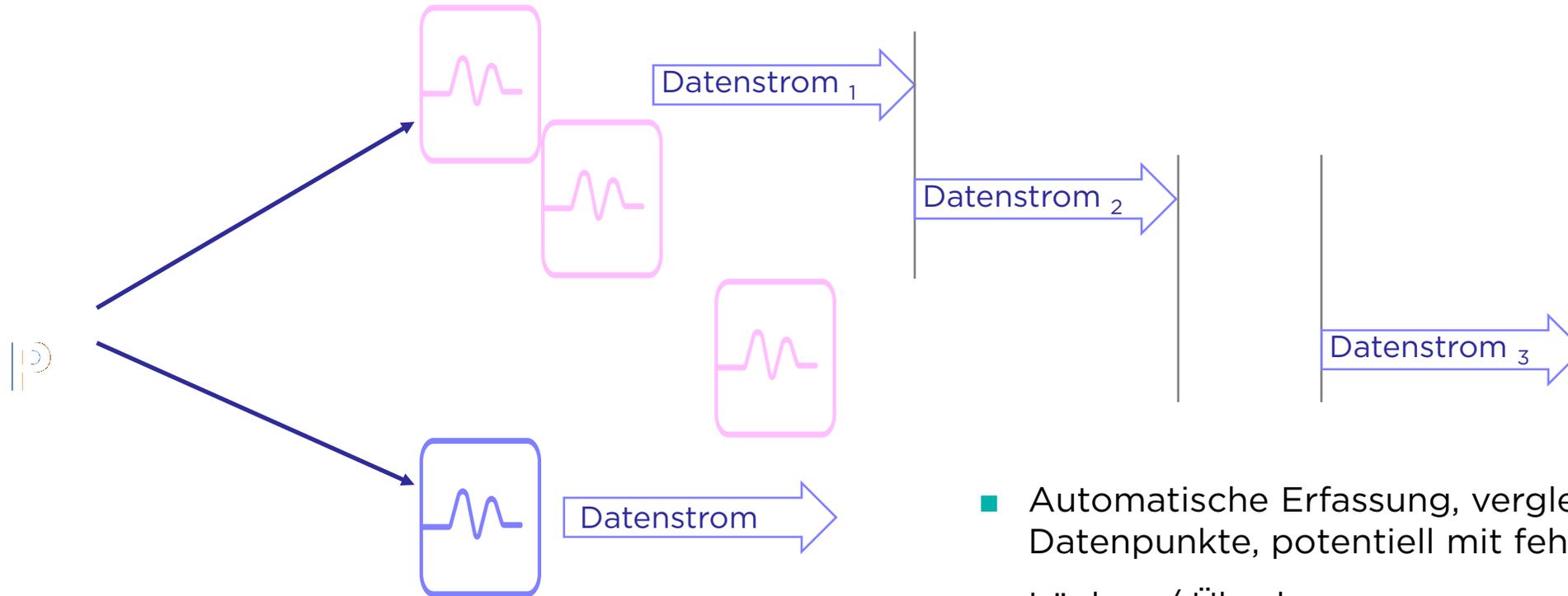
Traditional Programming



Erfassung von Daten mit Patientenbezug

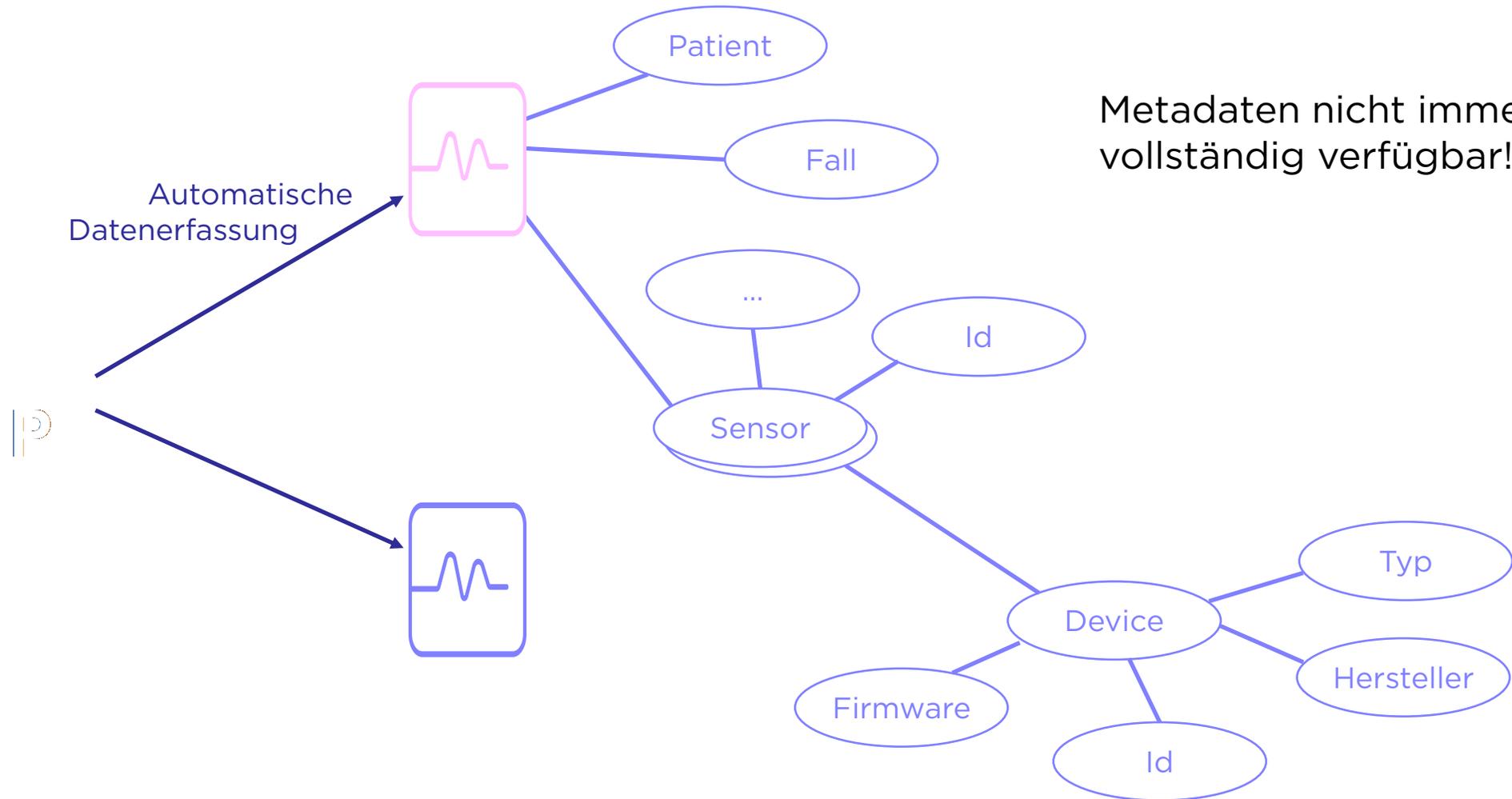


Erfassung automatisch generierter Daten

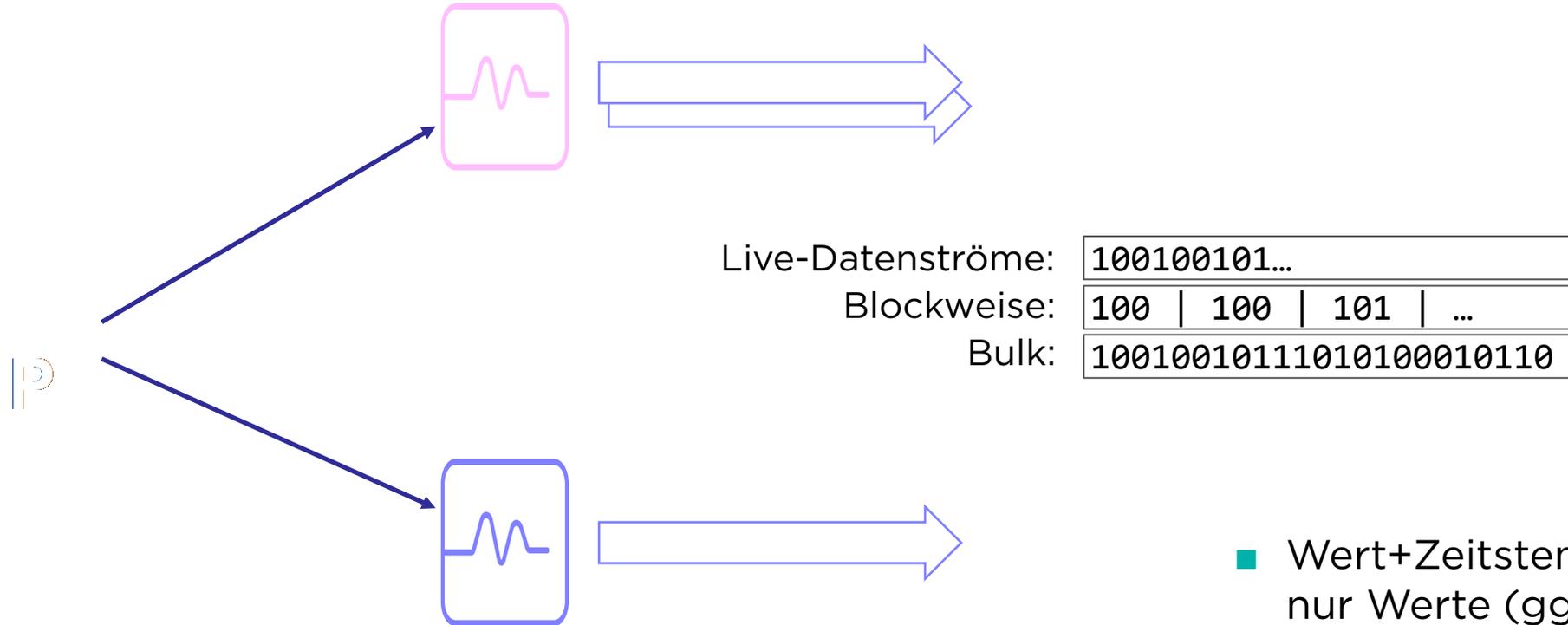


- Automatische Erfassung, vergleichsweise viele Datenpunkte, potentiell mit fehlerhaften Einträgen
- Lücken / Überlappungen
- [Datenpunkte evtl. ohne ID]

Erfassung automatisch generierter Daten

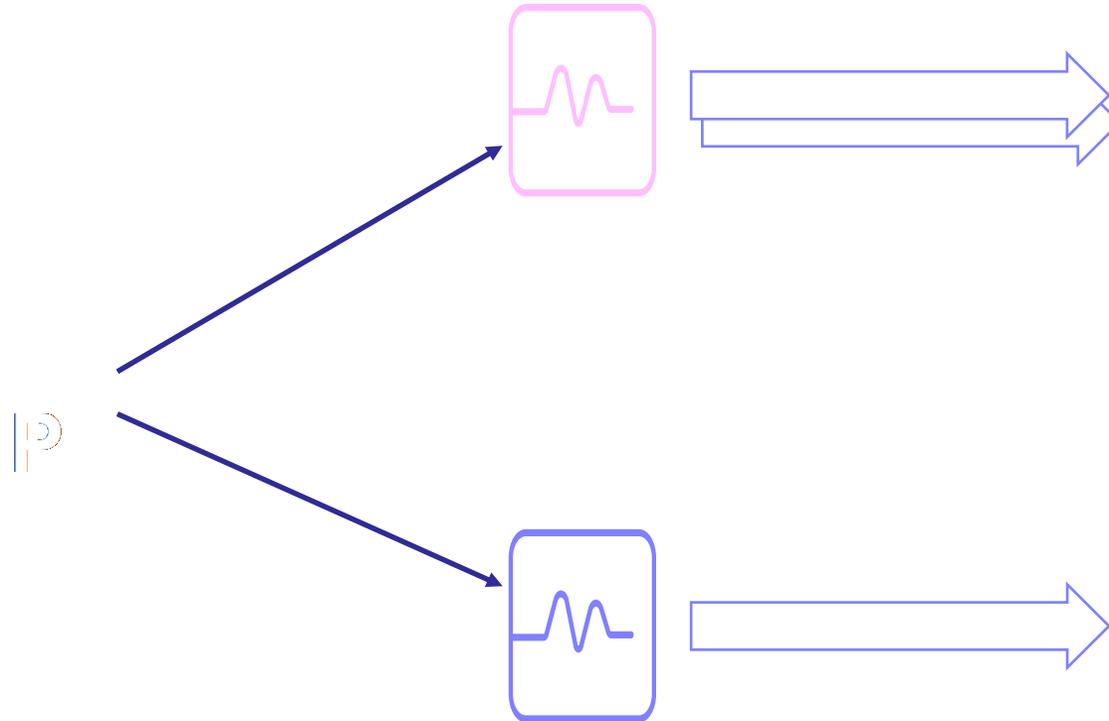


Erfassung automatisch generierter Daten



- Wert+Zeitstempel, Wert+Offset, nur Werte (ggf. äquidistant) ...
- Mit oder ohne Metadaten
- Reihenfolge

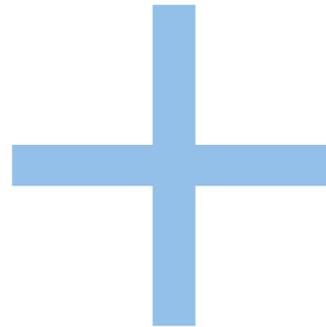
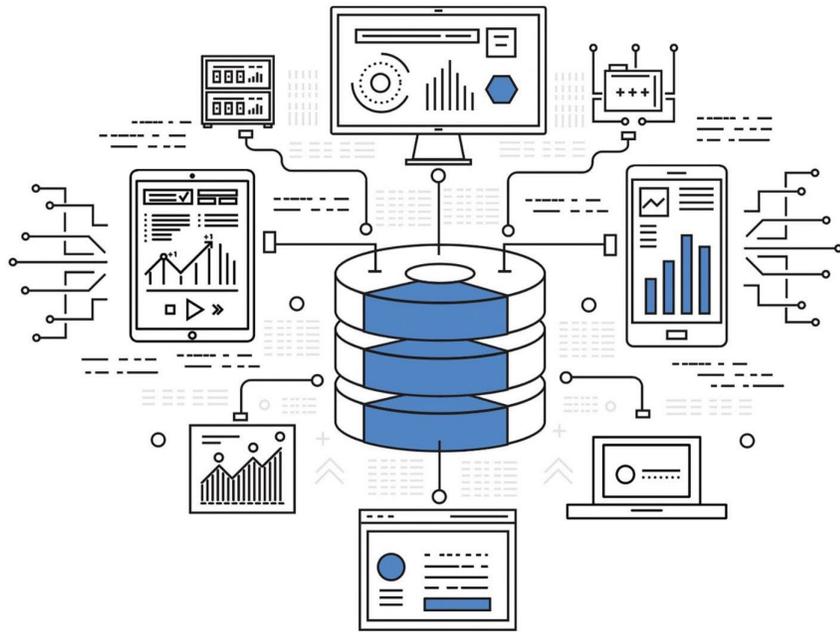
Erfassung automatisch generierter Daten



Datenvolumen:

- Ein Wert alle fünf Minuten:
>100k Datenpunkte pro Jahr und Patient
- Ein Wert alle fünf Sekunden:
>6m Datenpunkte pro Jahr und Patient

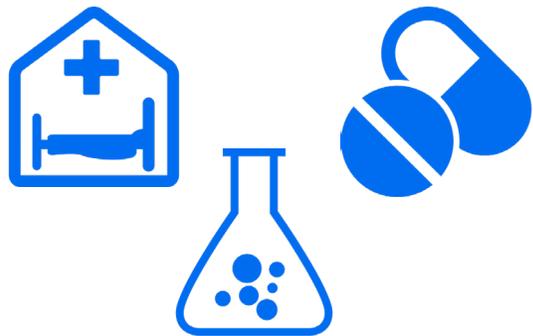
Die richtige Datenstrategie - mit InterSystems



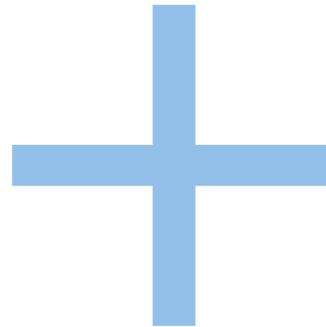
Standardisierung vs. Innovation



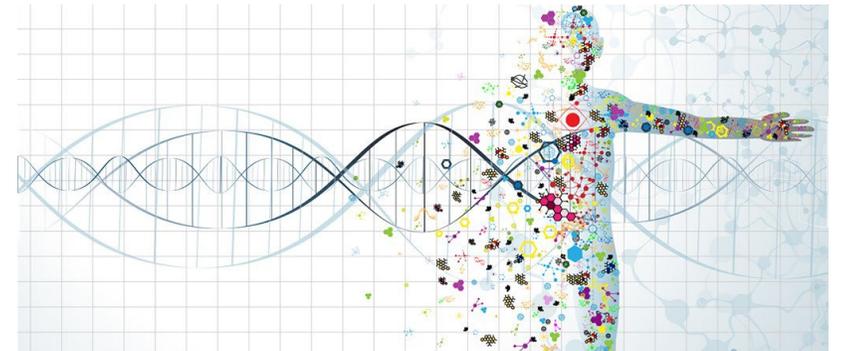
Operative Systeme



Solide, zuverlässig,
Interoperabilität auf der
Basis von Standards;
evolutionäre
Weiterentwicklung



Disruptive Ideen

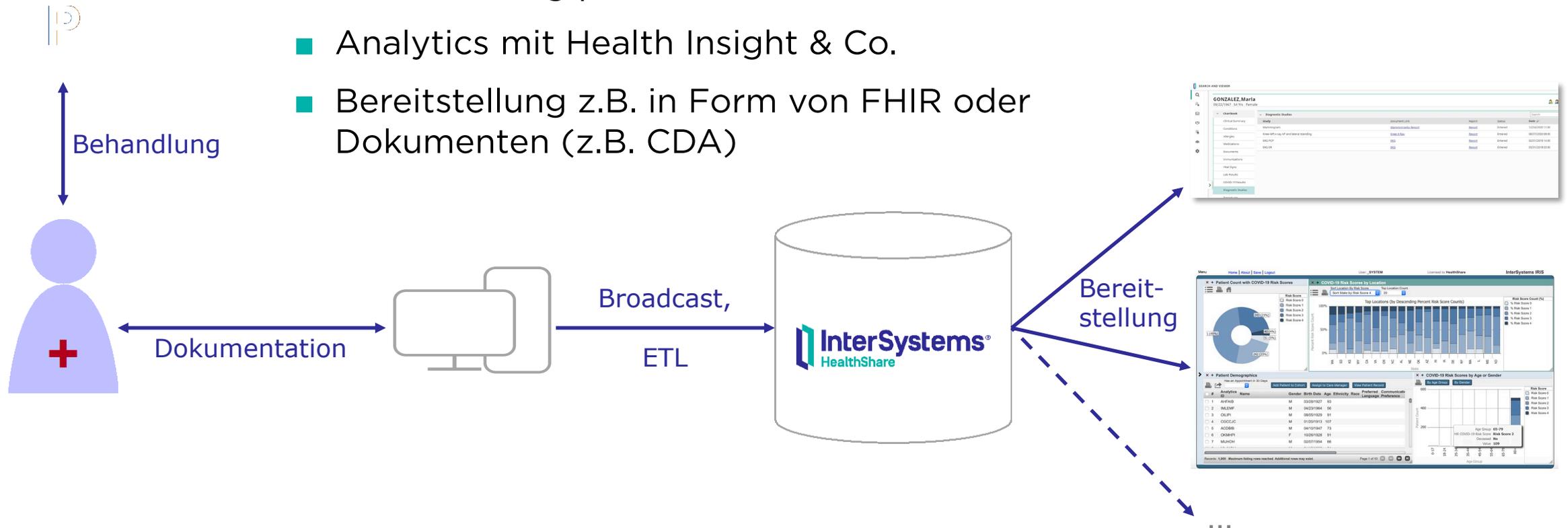


Hochgradig innovativ,
(noch) nur wenige übergreifende
Standards für Daten und
Interoperabilität;
revolutionäre Entwicklung

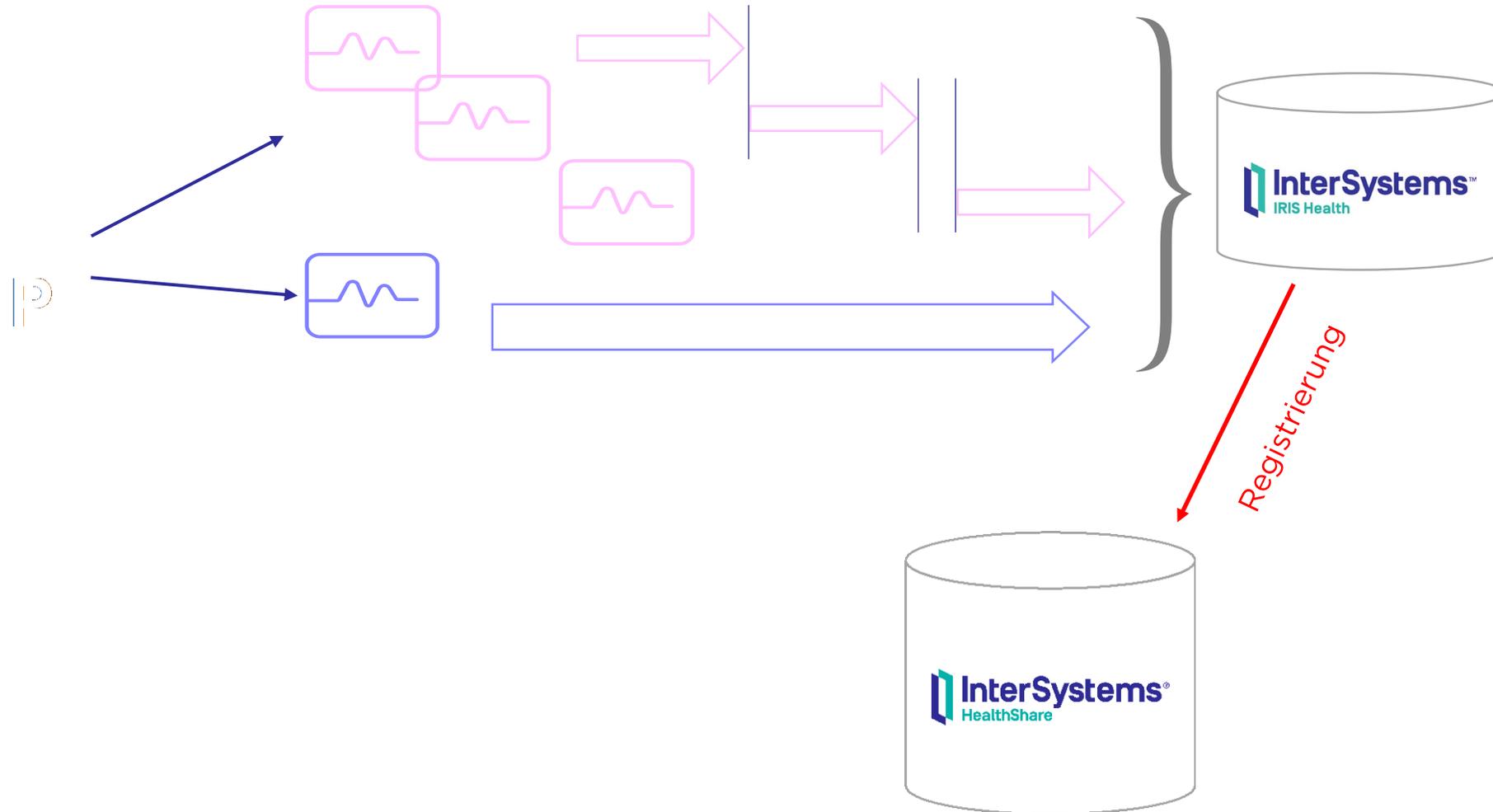
Datenbereitstellung via HealthShare



- (finale) Daten aus dem Behandlungsprozessen: vergleichsweise unregelmäßig, niedrige Frequenz pro Patient
- Visualisierung pro Patient im Clinical Viewer
- Analytics mit Health Insight & Co.
- Bereitstellung z.B. in Form von FHIR oder Dokumenten (z.B. CDA)



Datenbereitstellung via InterSystems IRIS for Health

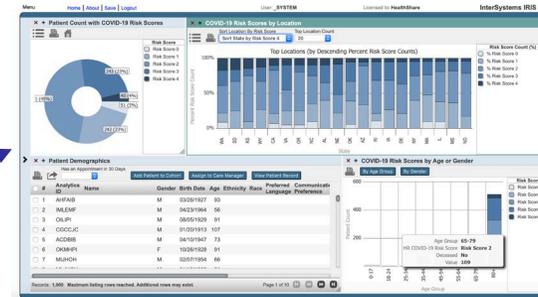


Datenbereitstellung via HealthShare & InterSystems IRIS



Zeitreihen-Analyse /
Complex Event Processing

Zeitreihen registrieren



Analytics



SEARCH AND VIEWER

GONZALEZ, Marla
09/22/1967 - 54 Yrs - Female

Study	Document Link	Report	Status	Date
Mammogram	Mammography_Report	Report	Entered	12/24/2020 11:00
Knee-Left x-ray AP and lateral standing	Knee_X-Ray	Report	Entered	08/07/2020 08:00
ENG FCP	ENG	Report	Entered	02/01/2019 14:00
ENG ER	ENG	Report	Entered	02/01/2019 14:00

Viewer
Plugin

Vielen Dank

