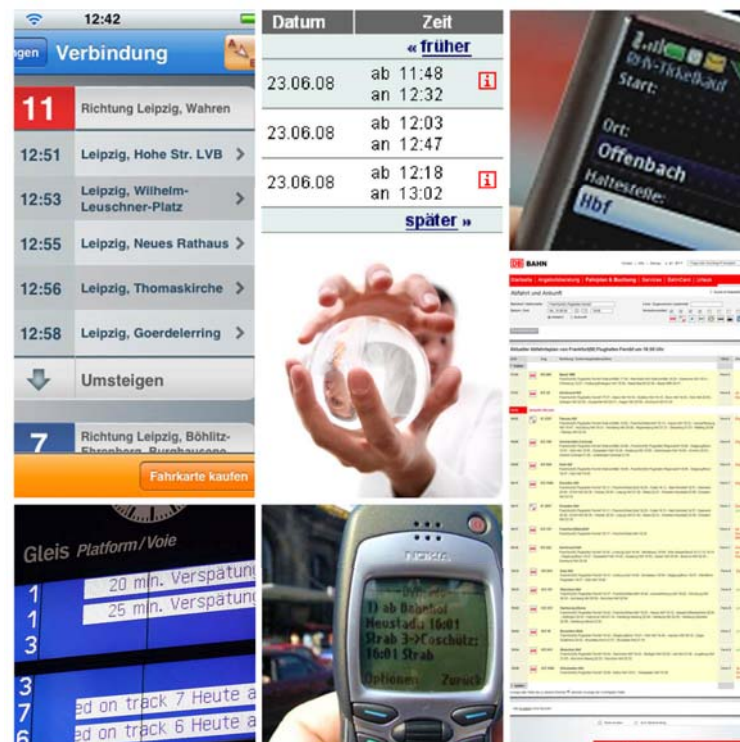
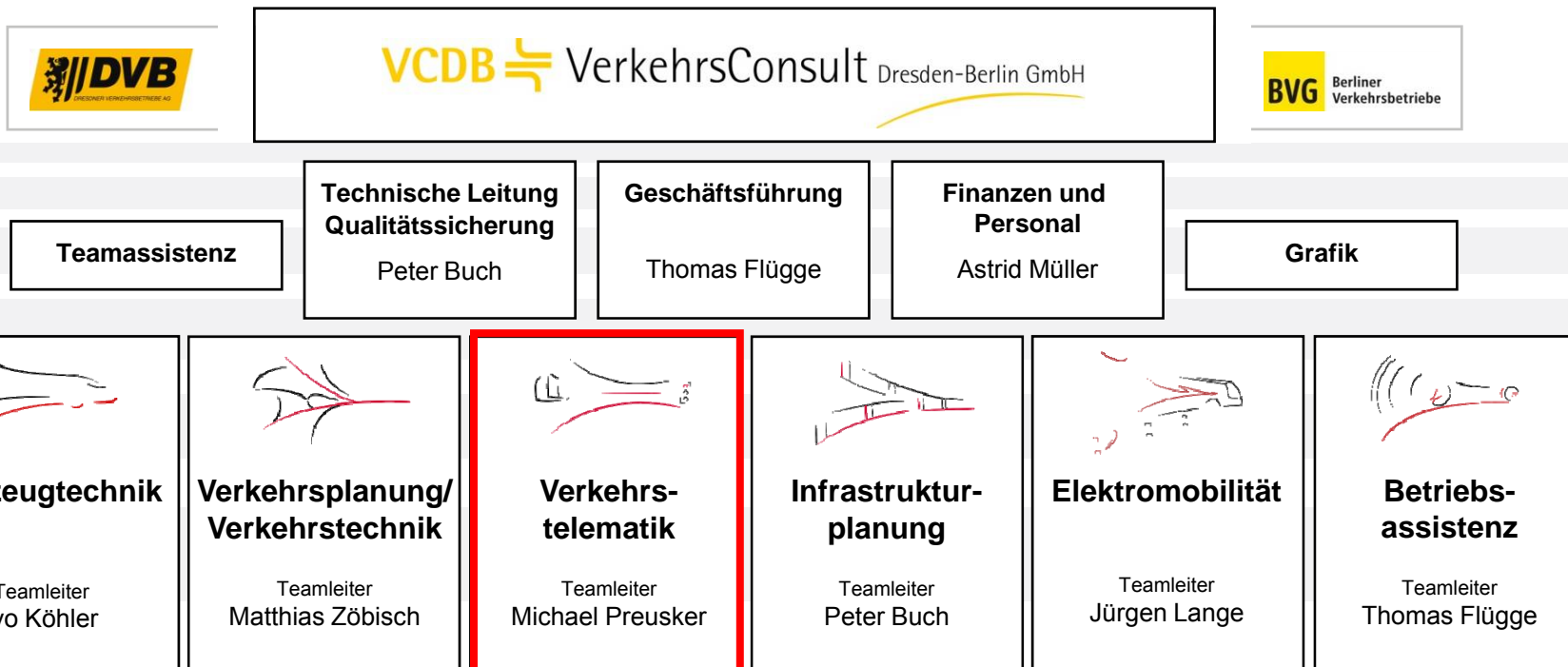


## Prognosen im ÖPNV — Blick in die Glaskugel ?



Hannover, 24. März 2011

## Die VCDB GmbH





























































## Begriffe

**Prognose** (griechisch, πρόγνωσις – wörtlich "das Vorwissen", die "Voraus-Kenntnis"), deutsch Vorhersage oder Voraussage, selten auch: Prädiktion (lat. praedicere – "voraussagen") ist eine Aussage über Ereignisse, Zustände oder Entwicklung in der Zukunft. (Quelle: Wikipedia)

- Fahrplanabweichung (Pünktlichkeit)
- Prognosewert
- Prognosehorizont
- Prognoseabweichung
- Prognosequalität (Verlässlichkeit)
- Quellsystem
- Zielsystem
- ...

# Beispiel Wetterprognose

Quelle: www.wetteronline.de

Datum	15.03.2011	16.03.2011	17.03.2011	18.03.2011	19.03.2011	20.03.2011	21.03.2011	22.03.2011
Do, 17.03.	 5° bis 10° ziemlich sicher							
Fr, 18.03.	 6° bis 11° weniger sicher	 6° bis 12° ziemlich sicher						
Sa, 19.03.	 5° bis 9° weniger sicher	 5° bis 13° weniger sicher	 4° bis 10° ziemlich sicher					
So, 20.03.	 4° bis 8° unsicher	 4° bis 8° weniger sicher	 4° bis 10° ziemlich sicher	 4° bis 10° sicher				
Mo, 21.03.	 4° bis 9° unsicher	 4° bis 9° unsicher	 4° bis 10° weniger sicher	 6° bis 11° ziemlich sicher	 6° bis 12° sicher			
Di, 22.03.	 6° bis 10° unsicher	 6° bis 10° unsicher	 6° bis 12° weniger sicher	 7° bis 13° ziemlich sicher	 8° bis 14° sicher	 8° bis 15° sicher		
Mi, 23.03.	 5° bis 11° unsicher	 5° bis 11° unsicher	 6° bis 13° unsicher	 9° bis 15° weniger sicher	 11° bis 15° ziemlich sicher	 10° bis 17° sicher	 9° bis 18° ziemlich sicher	
Do, 24.03.	 6° bis 13° unsicher	 6° bis 13° unsicher	 6° bis 12° unsicher	 9° bis 14° unsicher	 9° bis 17° ziemlich sicher	 9° bis 18° ziemlich sicher	 10° bis 18° ziemlich sicher	 10° bis 19° ziemlich sicher
		Fr, 25.03.	 7° bis 12° unsicher	 7° bis 14° unsicher	 5° bis 16° weniger sicher	 4° bis 16° weniger sicher	 3° bis 17° weniger sicher	 4° bis 20° ziemlich sicher
			Sa, 26.03.	 6° bis 12° unsicher	 6° bis 12° unsicher	 7° bis 16° unsicher	 5° bis 15° weniger sicher	 3° bis 14° weniger sicher
				So, 27.03.	 3° bis 10° unsicher	 7° bis 15° unsicher	 6° bis 12° unsicher	 3° bis 12° unsicher
					Mo, 28.03.	 7° bis 14° unsicher	 6° bis 13° unsicher	 8° bis 17° unsicher
						Di, 29.03.	 7° bis 14° unsicher	 4° bis 13° unsicher
							Mi, 30.03.	 6° bis 15° unsicher
								Do, 31.03.
								 9° bis 16° unsicher

## Erwartungen des ÖPV-Kunden (Fahrgast)

### Vor Auswahl der Fahrt bzw. Reisekette:

- voraussichtliche Reisedauer, Anzahl der Umsteigevorgänge, Verkehrsmittel etc.
- genaue Informationen über die Verlässlichkeit der Reisekette

### Vor Antritt und während der Fahrt bzw. Reisekette:

- genaue Informationen über die tatsächlichen Abfahrts- und Ankunftszeiten
- aktuelle, umfassende und verlässliche (Zusatz-) Informationen, besonders im Störfall, möglichst individuell:
  - Informationen über mögliche Anschlüsse, Zustand der Anschlusssicherung
  - Hinweis auf Alternativen im Störfall

→ Der ÖPV-Kunde will "Echtzeit-Informationen", dabei handelt es sich *a priori* immer um Prognosen !

## Stand der Technik: Kollektive FGI



### Dagegen Aushangfahrplan:


- noch aktuell?
- Sonderzeichen übersehen?
- Baustelleninfo übersehen?

- gehören zum Stadtbild, in Ballungsräumen Standard, zunehmend auch im ländlichen Raum an Übergangsstellen
- Unterscheidung Soll- und Ist-Daten nicht immer eindeutig, dennoch erzeugen DFI-Anzeiger ein "beruhigendes" Gefühl bei den Fahrgästen

## Stand der Technik: Individuelle FGI

Bahnhof/Haltestelle	Datum	Zeit	Gleis	Produkte
Dresden-Neustadt	Mi, 23.03.11	ab 06:14	5	RB 17100 Regionalbahn
Dresden Hbf	Mi, 23.03.11	an 06:20	2	Fahrradmitnahme begrenzt möglich
 Umsteigezeit 3 Min.				-> Umsteigezeit anpassen
Dresden Hbf	Mi, 23.03.11	ab 06:23	1	IC 2238 Intercity
Magdeburg Hbf	Mi, 23.03.11	an 08:55	3	Fahrradmitnahme reservierungspflichtig, Fahrradmitnahme begrenzt möglich, Bordbistro
 Umsteigezeit 6 Min.				-> Umsteigezeit anpassen
Magdeburg Hbf	Mi, 23.03.11	ab 09:01	7	IC 2432 Intercity
Hannover Hbf	Mi, 23.03.11	an 10:23 	11	Fahrradmitnahme reservierungspflichtig, Fahrradmitnahme begrenzt möglich, Bordbistro
 Fußweg 8 Min.				-> Umsteigezeit anpassen
Hauptbahnhof (U), Hannover	Mi, 23.03.11	ab 10:32		STB 8 Stadtbahn Richtung: Messe/Nord
Schlägerstraße (U), Hannover	Mi, 23.03.11	an 10:36		

### Aktuelle Informationen zu Ihrer Verbindung

 Aufgrund von Verspätungen kann es sein, dass der Anschlusszug nicht mehr rechtzeitig erreicht wird. Bitte beachten Sie die Ansagen im Zug und in den Bahnhöfen.

IC 2432: Magdeburg Hbf - Hannover Hbf: Bauarbeiten. Ankunftsverspätungen von 10 Min. möglich.

-> Weitere Informationen

Quelle: www.bahn.de

- Verbindungsauskunft fast nur mit impliziter Angabe der Verlässlichkeit ("Es ist mit Verspätungen zu rechnen")

## Problematik

- Linearprognose versagt regelmäßig bei nicht determinierten Betriebszuständen und spontanen Ereignissen
- Prognosegenauigkeit sinkt mit zunehmenden Prognosehorizont
- Qualität der Prognose unbekannt (bzw. wird nicht kommuniziert)
- Regelmäßig auftretende Abweichungen der Fahrzeiten (z.B. HVZ, Nachtverkehr) werden nicht mitgeteilt

→ Bei der Planung bzw. Auswahl von Reiseketten benötigt der Fahrgast genaue Informationen über:

- Prognosegenauigkeit der Ankunfts- und Abfahrtszeiten
- Anschlusssicherheit
- Verlässlichkeit der gesamten Reisekette
- Alternativen mit ggf. längerer Reisedauer, aber höherer Verlässlichkeit

## Forschungsprojekt FOPS 70.822

### "Steigerung der Qualität der Anschlusssicherung und Fahrgastinformation unter Einbeziehung personalisierter Dienste"

#### Hauptziele:

- Verbesserung der Prognosegenauigkeit von Ankunfts- und Abfahrtszeiten im ÖPV, insbesondere bei zunehmendem Prognosehorizont und Störungen
- Verbesserung der personalisierten Fahrgastinformation entlang von Reiseketten durch Bekanntgabe einer *a priori* Prognosegenauigkeit

#### Praxispartner:



## Forschungsgegenstände

### Theorie:

- Klassifizierung von Betriebszuständen
- Definition der Prognosequalität
- Weiterentwicklung der einschlägigen VDV/SIRI-Schnittstellen
- Entwicklung/Auswahl von angepassten Prognoseverfahren

### Praxis:

- Erkennen von Betriebszuständen
- Anwenden des "besten" Prognoseverfahrens je Betriebszustand
- Bestimmung der Prognosequalität (*a priori*)
- Verifizierung der Prognosequalität (*a posteriori*); ggf. rekursive Anpassung

→ Entwicklung eines Demonstratormoduls als "Sternpunkt" zwischen Quell- und Zielsystemen

## Betriebszustände (Anwendungsfälle, Usecases)

Kriterien der Klassifizierung:

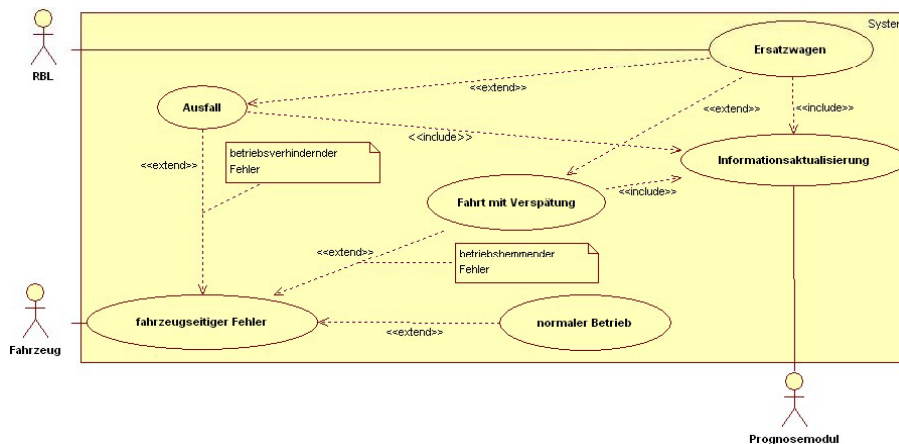
### Ursache:

- sagt meist etwas über die voraussichtliche Dauer des Betriebszustandes aus

### Wirkung:

- Einfluss auf den prognostizierten Wert und die Prognosequalität

Betriebszustände lösen einander ab



### Wirkungsspezifische Betriebszustände:

(Beispiele)

- Fahrt in Fahrplanlage (Normalbetrieb) bzw. gleichbleibender Verspätung
- Fahrt mit zunehmender Verspätung
- Fahrt mit abnehmender Verspätung
- Fahrt mit Verfrühung
- Unplanmäßiger Aufenthalt (Stau)
- Fahrt fällt aus
- ...

## Prognosequalität ...

... entspricht der Prognosegenauigkeit:

- Berechnung:

$$\Delta t(x) = |t(x)_{ist} - t(x)_{prog}|$$

- Optimalitätsbedingung:  $\sum \Delta t \rightarrow 0$

Fazit:

- Prognosegenauigkeit exakt nur *a-posteriori* bestimmbar, aber
- Kenntnis der Genauigkeit des Prognoseverfahrens erlaubt diese Angabe *a-priori*

Fragen:

- Genauigkeit bzw. Usecases im VDV-Standard 453/454 übermittelbar ?
- Wenn ja, wird dies bereits in den Quellsystemen umgesetzt ?

## Analyse VDV 453/454

- Prognose-  
Qualitätskennziffern (z. B. *Verlässlichkeit*) sind in ihren Abstufungen nicht deterministisch
- Usecase-Attribute (z. B. *Stauindikator*, *Ursache*) sind optional und stark interpretationsbedürftig

Dienst	Nachricht	Meldung	Element	Hersteller 1	Hersteller 2	Hersteller 3	Hersteller 4
ANS	ZubringerNachricht	ASBFahrplanlage	FahrtStatus	ja	ja	ja	nein
		ASBFahrtLoeschen	Ursache	nein	nein	nein	nein
	AbbringerNachricht	WertBis	Verlässlichkeit	nein	nein	nein	nein
		AbbringerFahrtLoeschen	Ursache	nein	nein	nein	nein
DF1	AZBNachricht	AZBFahrplanlage	FahrtStatus	ja	ja	ja	ja
			Stauindikator	nein	ja, wenn Stauerkennungs-Modul vorh.	ja	ja
		AZBFahrtLoeschen	Ursache	nein	nein	nein	nein
VIS	VISNachricht	VISFahrplanlage	FahrtStatus	ja	ja	ja	ja
			Stauindikator	nein (geplant 06/2010)	ja	ja	nein
		VISFahrtLoeschen	Ursache	nein	nein	nein	nein
AUS	AUSNachricht	IstFahrt	HinweisText	nein	nein, in Entwicklung	nein, in Entwicklung	nein
			PrognoseMoeglich	ja	ja	ja	ja
			PrognoseUngenau	nein	nein, in Entwicklung	ja	ja
			Zusatzfahrt	ja	ja	ja	ja
			FaelltAus	ja	ja	ja	ja
		IstHalt	PrognoseUngenau	nein	nein, in Entwicklung	ja	ja
			HinweisText	nein	nein	nein, in Entwicklung	nein

→ Erweiterung der VDV 453/454 – Schnittstelle notwendig !

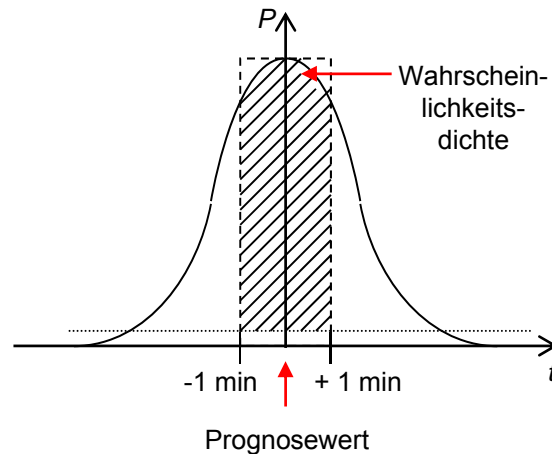
## Prämissen für die Definition des Qualitätsparameters

- Kein rein mathematischer Ansatz (System würde angreifbar)
- Fünf Stufen (überschaubar, bereits eingeführt mit anderer Semantik)
- Entkopplung von den Anwendungsfällen (keine 1-zu-1-Zuordnung möglich)
- Trennung von Ermittlung (im Quellsystem) und Verarbeitung (im Zielsystem)
- Vergleichbarkeit der gelieferten Qualitätsstufen (determinierte Werte)
- Beschränkung auf prognostizierte Ankunfts- und Abfahrtszeiten (AUS-Dienste)
  - Definition der Verlässlichkeit der Anschlusssicherung (ANS-Dienste) sollte in einem weiteren Arbeitsschritt erfolgen

## Definition der Prognosequalitätsstufen

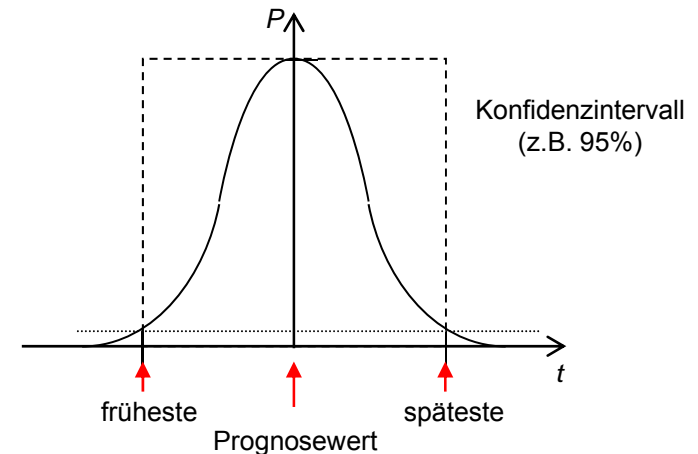
### Variante a:

- mittels Verlässlichkeit des Prognosewertes (Wahrscheinlichkeitsdichte der Verteilung im Bereich  $\pm 1$  min)



### Variante b:

- mittels Konfidenzintervall (eines sehr hohen Anteils der Prognosewerte)



### Variante c:

- mittels "griffiger" verbaler Beschreibungen mit dazugehörigen Anwendungsrichtlinien und Implementierungshinweisen



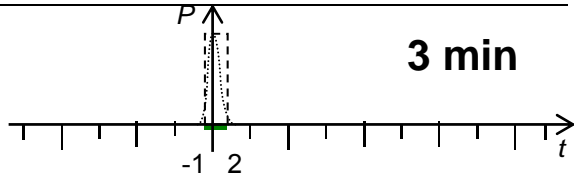
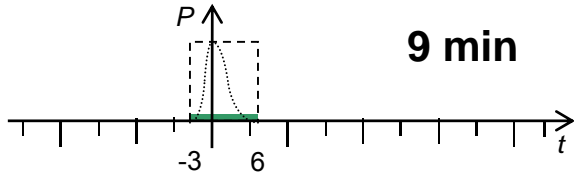
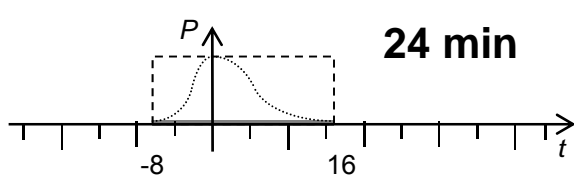
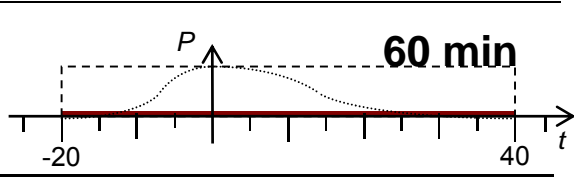
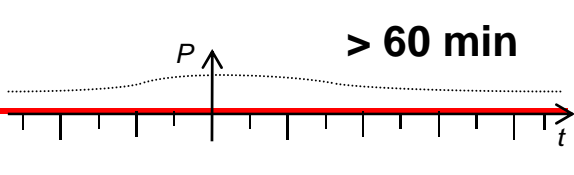
## Arbeitsgruppe "Ist-Datenschnittstellen"

- UAG "Prognosequalität" empfiehlt mehrheitlich den Ansatz b) – Definition über Konfidenzintervalle
- Bedeutung: Prognoseabweichung wird mit "sehr hoher Wahrscheinlichkeit" innerhalb des durch die Qualitätsstufe definierten Intervalls liegen
- Unsymmetrische Intervalle, da in der Praxis Tendenz zur Verstärkung von Verspätungen besteht
- Zusätzlich können explizite Grenzwerte übergeben werden (optionale Parameter *ZeitMin*, *ZeitMax*)
- Zur Validierung der Grenzwerte werden die Ergebnisse dieses Forschungsprojektes abgewartet

→ Einigung auf einen fünfstufigen Qualitätsparameter für AUS-Dienste mit definierten Grenzwerten auf Basis von Konfidenzintervallen



## Prognosequalitätstufen der AUS-Dienste

Stufe / Verbal	Unterer Grenzwert (früheste Ankunfts-/Abfahrtszeit)	Oberer Grenzwert (späteste Ankunfts-/Abfahrtszeit)	Intervallbreite
<b>1</b> Sehr sicher	- 1 min	+ 2 min	 3 min
<b>2</b> Ziemlich sicher	- 3 min	+ 6 min	 9 min
<b>3</b> Unsicher	- 8 min	+ 16 min	 24 min
<b>4</b> Sehr unsicher	- 20 min	+ 40 min	 60 min
<b>5</b> Prognose unbestimmt	unbestimmt	unbestimmt	 > 60 min

## Konsequenz dieser Definition

### Prognosequalität (Stufe lt. Def.):

- kann sich im Verlauf der Fahrt verändern
- ist *a priori* keine "Garantie" für die Einhaltung der Grenzwerte

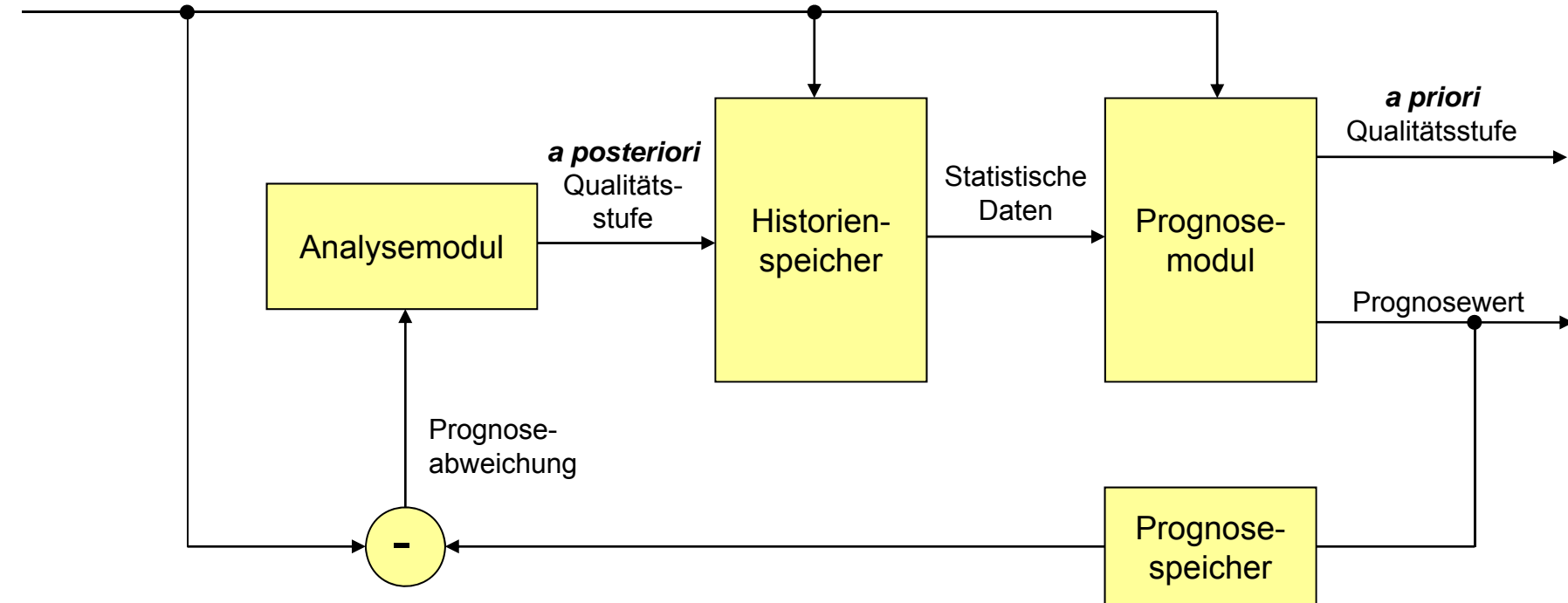
→ Konfidenzniveau sollte dennoch "sehr hoch" sein; ggf. ist ein bestimmtes anzustrebendes Konfidenzniveau zu vereinbaren (z.B. 95%)

### Ermittlung des Qualitätsparameters durch die Quellsysteme:

- algorithmisch aufgrund der Kenntnis interner Zustände des Systems (z.B. Mindestfahrzeiten, Dauer von Störungen)
- rekursiv durch *a posteriori* Analyse pro Anwendungsfall

## Rekursive Qualitätsbestimmung im Demonstratormodul

aktuelle  
Fahrtinformationen



- Prognosequalität ist abhängig von Betriebszustand (Usecase) und Prognosehorizont
- Übermittelte Qualitätsstufe wird rekursiv angepasst

## Prognosehorizont

- Ziel: Klassifizierung des Prognosehorizonts anhand der vorliegenden Daten und der zeitlichen Differenz zwischen Zeitpunkt der Prognoseerstellung und prognostizierten Ankunfts-/Abfahrtszeit
  
- Einteilung anhand der vorliegenden Daten:
  - Fahrt hat begonnen (= Echtzeitdaten vorhanden)
  - Fahrt hat noch nicht begonnen, aber Umlauf hat begonnen (= Echtzeitdaten der Vorgängerfahrten vorhanden)
  - Umlauf hat noch nicht begonnen (= keine Echtzeitdaten vorhanden)
  
- Zusätzlich weitere Unterteilung der Klassen in zeitliche Intervalle

## Demonstrator

- Hardware: Handelsüblicher PC (Intel i7 2,8 Ghz, 4 GB RAM) mit Windows 7 (32 Bit)
- Datenbank: Oracle 10g XE
- Programmiersprache: Java, IDE: Netbeans 6.9

### Vorteile:

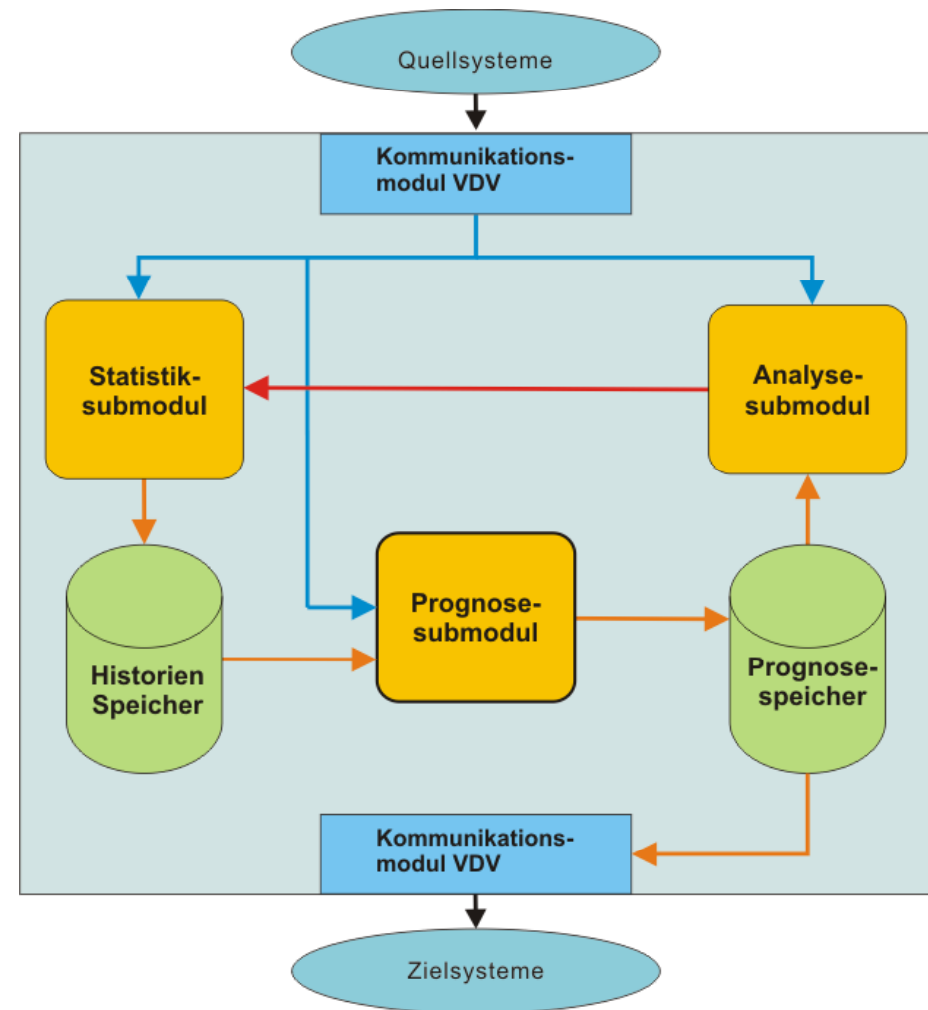
- Verwendung von Standard-Hardware und -Software
- Keine Lizenzkosten für genutzte Software
- Alle Komponenten auch auf anderen Betriebssystemen lauffähig (Linux, Mac OS, ...)
- Einfache Portierbarkeit

→ nach Abschluss des Forschungsvorhabens wird der Quelltext veröffentlicht

## Systemaufbau

### Modularer Aufbau:

- Schrittweise Implementierung & Validierung
- Erweiterung um zusätzliche Funktionalität bzw. Schnittstellen einfach möglich
- Leichter Vergleich der Prognosealgorithmen durch Austausch des Prognosesubmoduls ("Plug-In")
- Erlaubt parallele Verarbeitung der Daten ("Multi-Threading")
- Aufbau der Datenbank an VDV 452 orientiert

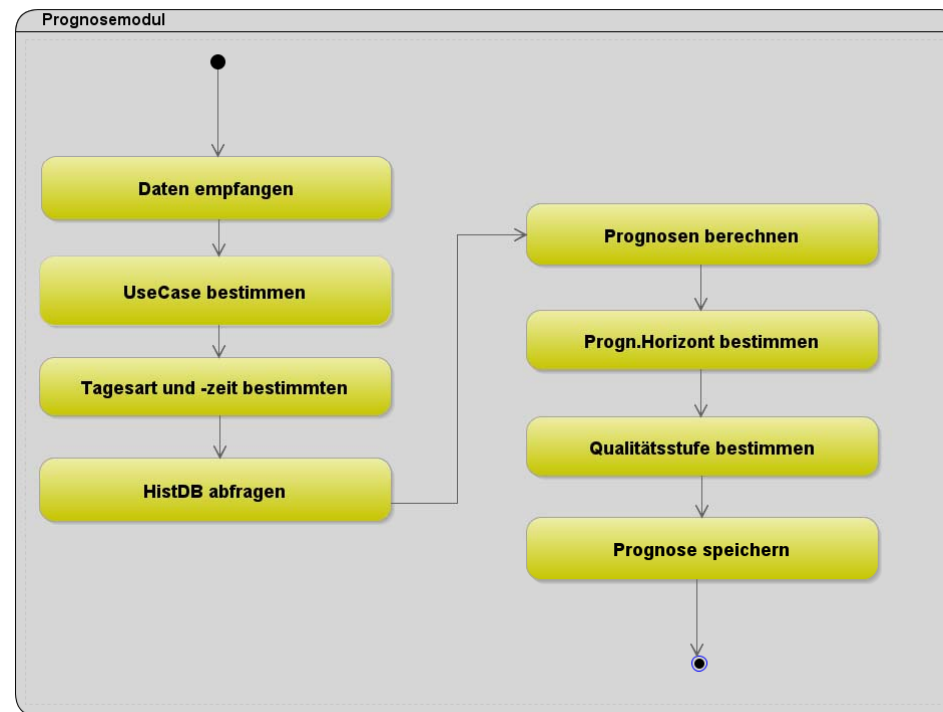


## Statistiksubmodul

- Folgende Daten werden erfasst:
  - Fahrzeiten zwischen den einzelnen Haltestellen
  - Fahrzeiten zwischen den Knotenpunkten (Teil-Routen)
  - Haltezeiten
- Alle Zeiten werden getrennt nach Tagesart, Tageszeit sowie Usecase ausgewertet und gespeichert
- Zusätzlich Auswertung der letzten x Fahrten auf einem Fahrtabschnitt
- Genutzte Auswerteverfahren:
  - Arithmetisches Mittel
  - Mean Square Error

## Prognosesubmodul

- Berechnet die Ankunft/Abfahrtprognose anhand von statistischen Daten und speichert diese in der Prognose-Datenbank ab
- Echtzeitdaten werden laufend "vorverarbeitet", um im Abfragefall schnell verfügbar zu sein



## Analyse der Prognoseverfahren

- Implementierte Verfahren:
  - Statistische Fahrzeiten (Annahme: feste Aufenthaltszeiten an Haltestellen)
  - zusätzlich statistische Aufenthaltszeiten
  - zusätzlich Fahrdaten Vorgängerfahrt(en) auf den Kanten
  
- Referenzverfahren: Lineare Progression
  
- Vergleich der erzielten Prognosestufen bzw. der Prognoseabweichungen

→ Ziel: Auswahl des besten Verfahrens in Abhängigkeit von Usecase /  
Prognosehorizont

## Stand und weitere Entwicklung des Demonstrators

- Framework und Datenbanken etc. sind erstellt
- Visualisierungsmodul implementiert (auf Basis GoogleEarth)
- Aktuell:
  - vergleichende Durchläufe der Algorithmen anhand der aufgezeichneten Echtzeitdaten
  - Auswertung der Daten nach "interessanten" Usecases
  - Erstellen von determinierten Störungsszenarien
  - Erprobung der "Usecase-Prediction"

## Demonstrator "Live"

**Bahnhof Mitte**

- Tram** 10 nach Tolkewitz ab 6:06 Prognose: 6:06:44 (1)
- Tram** 10 nach Friedrichstadt ab 6:04 Prognose: 6:06:46 (1)
- Tram** 1 nach Betriebshof Waltherstr. ab 6:09 Prognose: 6:09:40 (1)
- BUS** 94 nach Cossebaude Bf. ab 6:12 Prognose: 6:12:51 (1)
- Tram** 2 nach Kleinschachwitz ab 6:13 Prognose: 6:15:31 (1)
- BUS** 75 nach Goppeln ab 6:19 Prognose: 6:19:00 (1)
- Tram** 2 nach Gorbitz (Schleife) ab 6:19 Prognose: 6:20:21 (1)
- Tram** 6 nach Niedersedlitz ab 6:20 Prognose: 6:21:01 (1)
- Tram** 6 nach Niedersedlitz ab 6:21 Prognose: 6:22:04 (1)
- Tram** 1 nach Prohls Gleisschleife ab 6:23

Route: [Hier hin](#) - [Von hier](#)

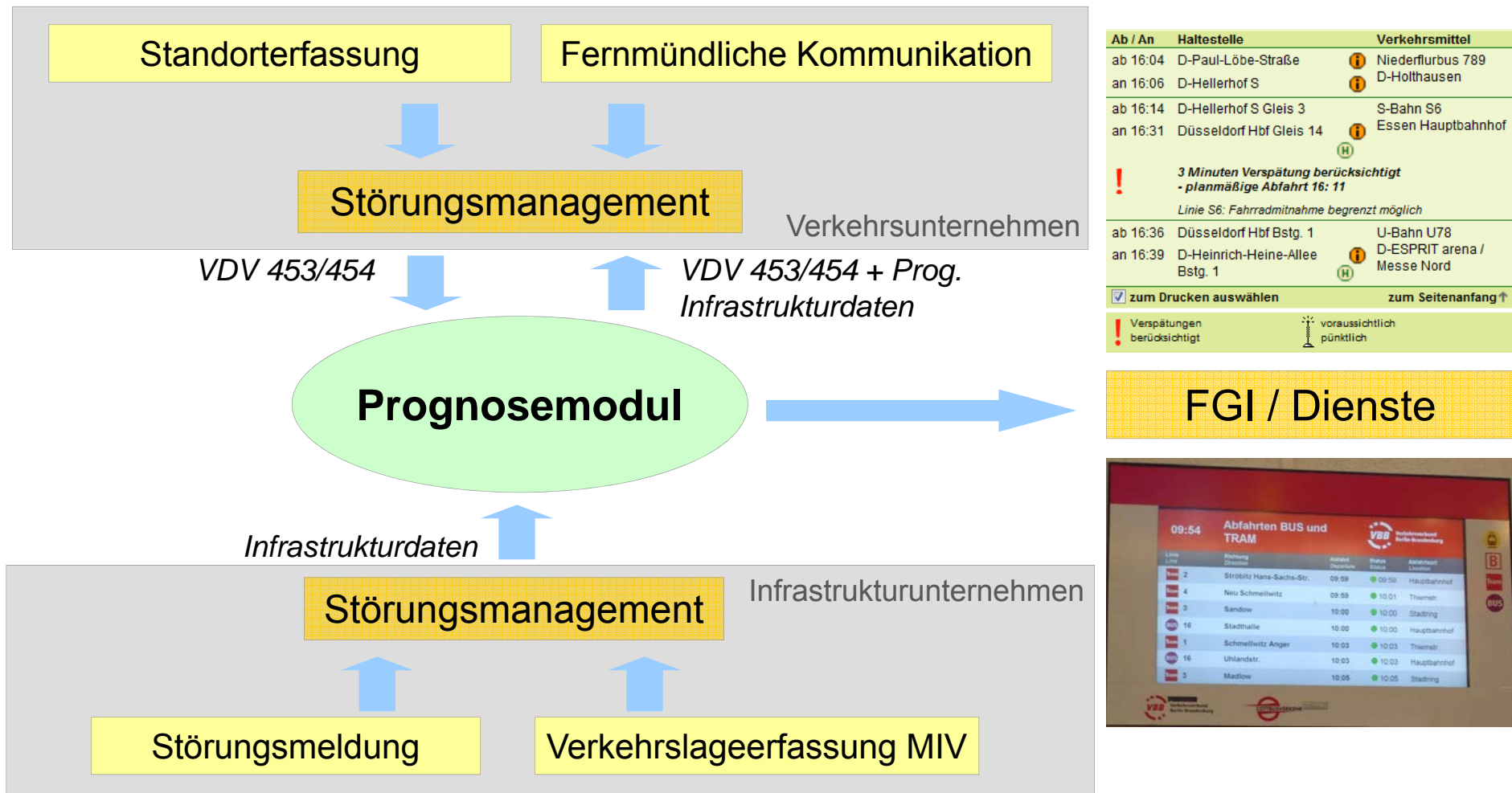
Map showing transport stops and arrival times:

- Lene-Glatzer-Straße: an 6:21 Prognose: 6:26:06 1
- Lene-Glatzer-Straße: ab 6:21 Prognose: 6:26:21 1
- Prellerstraße: an 6:22 Prognose: 6:27:14 1
- Prellerstraße: ab 6:22 Prognose: 6:27:29 1
- Schillerplatz: an 6:24 Prognose: 6:28:45 1
- Schillerplatz: ab 6:24 Prognose: 6:29:00 1
- Jüngststraße: an 6:26 Prognose: 6:30:47 1
- Jüngststraße: ab 6:26 Prognose: 6:31:02 1
- Heinrich-Schütz-Straße: an 6:27 Prognose: 6:31:41 1
- Heinrich-Schütz-Straße: ab 6:27 Prognose: 6:31:56 1
- Gustav-Freytag-Straße: an 6:28 Prognose: 6:33:16 2
- Gustav-Freytag-Straße: ab 6:28 Prognose: 6:33:31 2
- Ludwig-Hartmann-Straße: an 6:30 Prognose: 6:35:08 2
- Ludwig-Hartmann-Straße: ab 6:30 Prognose: 6:35:23 2
- Johannfriedhof: an 6:32 Prognose: 6:36:33 2
- Johannfriedhof: ab 6:32 Prognose: 6:36:48 2
- Tolkewitz: an 6:33 Prognose: 6:38:20 2
- Tolkewitz: ab 6:33 Prognose: 6:38:35 2
- Wasserwerk Tolkewitz: an 6:34 Prognose: 6:39:20 2
- Wasserwerk Tolkewitz: ab 6:34 Prognose: 6:39:35 2
- Alttolkewitz: an 6:36 Prognose: 6:41:23 2
- Alttolkewitz: ab 6:36 Prognose: 6:41:38 2
- Hermann-Seidel-Straße: an 6:37 Prognose: 6:42:14 2
- Hermann-Seidel-Straße: ab 6:37 Prognose: 6:42:29 2

## Ausblick / Nächste Aktivitäten

- Seit Anfang 2011 wird der Demonstrator mit historisierten Echtzeiten aus dem RBL der Dresdner Verkehrsbetriebe AG getestet
- In Kürze werden Online-Daten aus dem Regio-RBL der Verkehrsverbände VVO und ZVON erwartet
- Demonstrator wird dann im Dauerbetrieb mit Daten aus dem AUS- und VIS-Diensten getestet und validiert
- evtl. Weiterführung der Aktivitäten im Rahmen der VDV UAG  
Prognosequalität: Definition der Qualitätsparameter für ANS-Dienste
- Optional Feldtest mit Online-Daten und "echten" Zielsystemen in Produktionsumgebung ab Mitte 2011

## Anzustrebende Einbindung in itcs-Systemlandschaft



## Fazit

- Entscheidend ist die Erkennung von Betriebszuständen und deren voraussichtlicher Dauer
- Prognosegenauigkeit kann bei bestimmten Betriebszuständen nicht beliebig gesteigert werden
- Schnittstellenerweiterungen bei Quellsystemen hinsichtlich Art und Dauer des vorliegenden Störfalles wünschenswert
- Evtl. Definition eines Ist-Datendienstes Verkehrsinfrastruktur, insbesondere für Fahrtabschnitte, die von mehreren VU genutzt werden
- Für optimale Reisekettenbildung in Zielsystemen ist zusätzlich die Definition der Verlässlichkeit der Anschlusssicherung anzustreben

→ Die "Glaskugel" wird nicht benötigt, wohl aber strukturierte Informationen aus den itcs-Quellsystemen (insbesondere aus dem Störungsmanagement)

## Prognosen im ÖPNV — Blick in die Glaskugel ?



**Vielen Dank für Ihre Aufmerksamkeit !**