

# Insights, Analytics & Modelling – Wie Reisende den ICE nutzen und nutzen werden

Dr. Jürgen Antes, Dr. Robert T. Firla, Dr. Martin Thust

Business Intelligence Marketing  
DB Fernverkehr AG  
Stephensonstraße 1  
60326 Frankfurt am Main  
juergen.antes@deutschebahn.com  
robert.firla@deutschebahn.com  
martin.thust@deutschebahn.com

**Abstract:** Geschäftsentscheidungen für die DB Fernverkehr AG basieren auf umfassendem Wissen über aktuell getätigte und prognostizierte zukünftige Reisen. Um entsprechende Entscheidungsunterstützung zu leisten integrieren unsere BI-Systeme dazu zunehmend vielfältigere Datenquellen. Wir beschreiben Hürden bei Aufbau und Nutzung der Systeme aus Bahn-systemischer-, Daten- und Nutzerperspektive sowie unsere Ansätze zu Integration, Strukturierung und Visualisierung zum Umgang damit.

## 1 Einleitung

Die DB Fernverkehr AG bietet Fernreisemöglichkeiten mit ICE- und IC/EC-Zügen im Linienverkehr für ihre Reisenden an, der mit rund 400 Zügen und etwa 1000 kommerziellen Zugfahrten pro Tag erbracht und von ca. 130 Mio. Reisenden im Jahr genutzt wird. Wir erreichten 2013 ca. 4 Mrd. EUR Umsatz und rund 8% EBIT-Marge [DB2014].

Unsere Geschäftsentscheidungen benötigen umfassendes Wissen darüber, wie unsere aktuellen Angebote von den Reisenden real genutzt werden und Prognosen, wie Reisende dies zukünftig bei veränderten Angeboten tun werden. Beispiele sind Fragen zur Flottenerneuerung (Wie viele Züge? Wie schnell? Wie viel 1./2. Klasse Plätze?), zur Fahrplangestaltung (z.B. neue Strecken) oder zur Kontingentsteuerung der Sparpreise.

Stetig wachsende Herausforderungen (Wirtschaftlichkeit, Wettbewerb) führen auch bei uns zu einem stetig steigenden Bedarf an verfügbaren Informationen bzgl. Genauigkeit, Detaillierung und Vielfalt, damit automatisch zu steigender Komplexität der BI-Systeme und entsprechenden **Nutzungshürden**. Konkret möchten wir drei hervorheben:

- **Systemische:** Im Gegensatz zu Airlines gibt es bei uns **keine vollständigen** flug-, ticket- und personengenauen **Datenerfassung**<sup>1</sup> für alle Einzelreisen, z.B.

---

<sup>1</sup> Ein Extrembeispiele: die Bahncard 100 erlaubt beliebiges Reisen im Netz der DB ohne weiteres Ticket. Es erfolgt keinerlei Erfassung der damit getätigten Reisen, auch nicht bei der Fahrscheinkontrolle.

beinhalten Normalpreistickets das Recht von einem Start- zu einem Zielbahnhof zu reisen, legen aber weder Zugverbindung noch das genaue Datum fest.

- **Datenseitige:** Wir benötigen daher eine **Vielzahl sehr heterogener Datenquellen**<sup>2</sup> um die notwendigen Entscheidungsgrundlagen zu schaffen.
- **Nutzerseitige:** Hohe **Nutzerakzeptanz** ist nur bei **Vollständigkeit** der Informationen, Anbieten **fachspezifischer Sichten** (z.B. von der Reise her über Quelle-Ziel-Relationen oder vom Fahrplan her über Züge), intuitive **Nachvollziehbarkeit** und **Einfachheit** der Nutzung zu gewinnen.

Dem begegnen wir mit drei Ansätzen: **systematische Zusammenführung** aller relevanten **Daten**, **gleichartige Strukturierung** und damit einfache Vergleichbarkeit für **Ist- und Prognosedaten** sowie **sehr guten Visualisierungsmöglichkeiten**. Im Weiteren erläutern wir die dazu entwickelten „Nachbar“-Systeme TReK für Monitoring und Analyse der Ist-Reisen (Abs. 2), PRIMA zur Prognose und Bewertung von zukünftigen Angebotsszenarien (Abs. 3) und enden mit einer Skizze aktueller Fragestellungen (Abs. 4).

## 2 TReK – Wie Reisende den ICE/IC heute nutzen

Die Anforderung das laufende Geschäft sehr gut zu verstehen kumuliert in der Frage „*Wo, wann und wie nutzen Reisende<sup>3</sup> den Fernverkehr der DB?*“. *Wo* fragt nach Start- und Zielort der Reise sowie den dabei genutzten Zügen oder weiter gefasst, der Zugverbindung bzw. der Reisekette. *Wann* fragt nach Reisedatum sowie den einzelnen Zeitpunkten innerhalb der Reisekette (Abfahrt, Umstiege, Ankunft). *Wie* fragt u.a. nach Art des Tickets, Nutzung einer Bahncard, Preis, Klasse, Gruppengröße u.v.a. Die Bereitstellung von Daten zur Beantwortung dieser Fragen ist Aufgabe von TReK.

### 2.1 Datengrundlagen und Informationsschichten

Zentrale Datengrundlage von TReK sind die Buchungs- und Verbindungsdaten, die beim fahrplanbasierten Verkauf durch die IT-Systeme erfasst werden. **Stufe 1** umfasst die **Buchungsdaten für Sparpreise**, also zuggebundene Tickets inkl. der fest gebuchten Zugverbindung, **Stufe 2** die **Normalpreistickets** inkl. der angefragten Zugverbindungen. Aus Marktforschungen wissen wir, dass das Gros der Reisenden – trotz voller Flexibilität des Normalpreises bzgl. der gewählten Zugverbindung – die beim Ticketkauf angefragte Verbindung auch real nutzt.

Die Daten dieser beiden Stufen liefern also explizite Verbindungsinformationen und somit Transparenz über die Reisekette, umfassen jedoch nur 70% des Reiseaufkommens. Um Nutzerakzeptanz insb. im Punkt Vollständigkeit zu erreichen, sind die fehlenden 30% der Reisen z.B. solche mit Streckenzeitkarten oder Bahncard 100 zu ergänzen.

---

<sup>2</sup> z.B. Daten aus Verkaufssystemen, Reisendenzählungen, Marktforschungen, Daten zu Schienen und Bahnhöfen, Fahrplandaten, sozio-ökonomische Daten, Umfeldinformationen über Markt und Wettbewerber

<sup>3</sup> TReK kennt keine personenbezogenen Daten sondern nur anonyme Reisen

Dabei handelt es sich um Angebote deren Nutzung z.T. strukturelle Besonderheiten beinhalten (Reisezeitpunkt, Preis, Relation), womit eine einfache Hochrechnung un-zweckmäßig ist. Der für **Stufe 3** beschrittene Weg basiert daher auf der **modellbasierten Erzeugung von Reiseketten** für die **fehlenden Angebote** unter Beachtung jeweils verfügbaren Informationen und unter Einsatz des Verkehrsprognosemodells PRIMA (siehe Abs. 3). Da die verfügbaren Informationen zwischen den Angeboten stark variieren, wurden jeweils separate Modelle entwickelt, die die jeweiligen Reiseketten erzeugen, in identischer Struktur und Detaillierung wie die in Stufe 1 und 2 erfassten Reiseketten.

Der so geschaffene Datenbestand erlaubt es die 130 Mill. Reisen p.a. konsistent, durchgängig, einheitlich, in variierenden Sichten – zeitlich, räumlich, nach Relationen, nach Zügen, nach Tickets – und auch deren zeitliche Entwicklung auszuwerten.

## 2.2 Von den Daten zur Entscheidungsunterstützung

TReK ist als klassisches fünfschichtiges Data Warehouse-System realisiert:

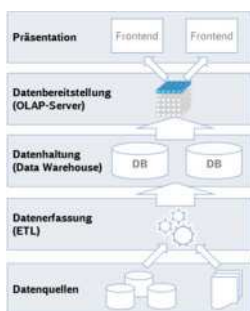


Abb. 1: Architektur und schematischer Aufbau des BI-Systems TReK

Wie bei (fast) allen BI-Systemen sind Beschaffung, Zusammenführung und Qualitätssicherung heterogener Datenquellen zentrale Kernaufgaben und Herausforderungen. Ergänzend zu den Reiseketten sind dies u.a. Infrastruktur-, Fahrplan- und Stammdaten, Prämissen und Geschäftsregeln. Die Datenerfassung und Aufbereitung umfasst auch Einzelmodule die Geschäftsregeln (u.a. Erlösaufteilung) implementieren. Das zentrale Data Warehouse ist mittels mehrschichtiger Datenbanken realisiert, die Datenbereitstellung mittels OLAP-Server. Da es uns letztlich um Erkenntnisgewinn und resultierende Wertschöpfung im Geschäft geht, kommt der Präsentation und dem Zugang zu den Systeminhalten eine besondere Bedeutung zu. Wir setzen auf eine Mischung aus klassischen BI-Komponenten (Standard- und *ad hoc*-Reports mit wachsendem *self-service*-Anteil) und spezifischen, kontextbezogenen Visualisierungen, die wir nun kurz vorstellen.

## 2.3 Visualisierung von Reisen und Reiseströmen

Nach unseren Erfahrungen ist es für die Nutzerakzeptanz essentiell, dass BI-Systeme den verschiedenen Anwendergruppen nicht „nur“ ein Meer an Daten anbieten, sondern dass dies jeweils fachspezifisch geeignet und zweckmäßig verdichtet geschieht. Visualisie-

nung kommt dabei besondere Bedeutung zu, da diese einen **Weg zur intuitiven Erfassbarkeit von Mustern in größeren Datenmengen** darstellt. Für die Visualisierung von Reisen und Reiseströmen wollen wir dies an zwei Beispielen erläutern.

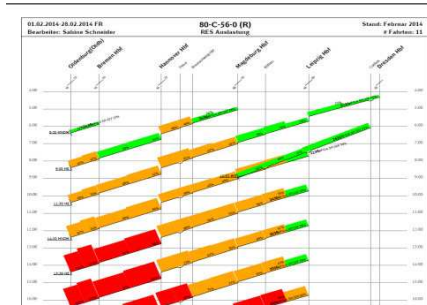


Abb. 2: Visualisierung Reiseströme im Weg-Zeit-Diagramm (Bildfahrplan)

**Visualisierung entlang der Zugfahrten:** Sogenannte Weg-Zeit-Diagramme (Bildfahrplan) sind eine klassische Darstellung von Fahrplänen. Sie zeigen entlang der X-Achse den Streckenverlauf mit Haltebahnhöfen und entlang der Y-Achse die Zeit. Einzelne Diagonalstriche zeigen Zugfahrten entlang von Ort und Zeit. Wir ergänzen dies durch die Visualisierung frei wählbarer Attribute entlang der einzelnen Zugfahrten: etwa die Balkenbreite als Besetzung und die Farbe als Auslastungsklasse. Vergleiche zwischen oder Identifikation von Strukturen oder Besonderheiten einzelner Fahrten sind auf diesem Wege substantiell einfacher als eine tabellarische Analyse der gleichen Daten.

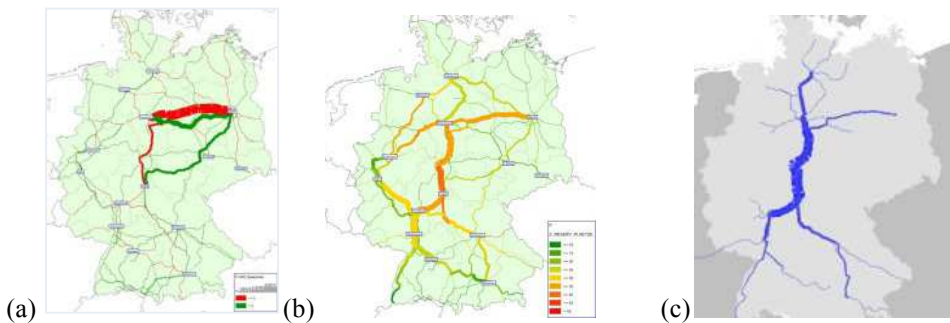


Abb. 3: Räumliche Visualisierung:  
 (a) Differenznetze, (b) Gesamtnetz inkl. spez. Kennzahlen (c) Spinnendarstellung

**Visualisierung räumliche/geographische Dimensionen:** Reiseströme in räumlicher Ausprägung darzustellen ist eine weitere Möglichkeit. Während dies für ortsbezogene Daten, z.B. Einsteiger je Bahnhof, ein etabliertes Feature vieler BI-Frontends ist, ist die Darstellung der Reiseströme (aus Streckensicht) deutlich komplexer, da die Einzelinformationen der Reisekette des Ticktes (von Bhf. A zu Bhf. B mit Zug C) den richtigen

Strecken zugeordnet werden müssen (*routing*). Dazu nutzen wir die Verkehrsplanungssoftware Visum<sup>4</sup>. Abb. 3 zeigt Beispiele für verschiedene mögliche Visualisierungen.

Obwohl Visualisierungen (fast) keine neuen Informationen beinhalten, sind wir sicher, dass die Erzeugung „bunter Bilder“ nicht nur den Erkenntnisgewinn erleichtert. Vielmehr wird damit ein intuitiverer Zugang zu den Daten, ein hohes Maß an Nachvollziehbarkeit und damit ein deutlicher Beitrag zur Nutzerakzeptanz erreicht.

### 3 PRIMA – Wie Reisende den ICE/IC zukünftig nutzen werden

Für Investitionsentscheidungen in Infrastruktur oder Fahrzeuge sowie die Gestaltung des Fahrplans genügt es nicht, die Nutzung der Fernverkehrszüge im Status quo zu kennen, sondern es ist auch die zukünftige Nutzung unter veränderten Bedingungen vorherzusagen. Diese Prognose der Wirtschaftlichkeit (Umsatz, Kosten) von Fahrplansenarien übernimmt das System PRIMA, dessen komplexester Teil die Nachfrageprognose ist.

#### 3.1 Das Nachfrageprognosemodell

Die Prima-Nachfrageprognose ist zeit- und verbindungsfein, erzeugt dabei identische Strukturen (Reiseketten) wie TReK<sup>5</sup> und ermöglicht dieselben detaillierten Auswertungen. Damit ist eine wesentliche Erfolgsvoraussetzung gegeben: direkte und strukturgleiche Vergleichbarkeit der Prognose- zu Ist-Daten. Der Ausgangspunkt des Modells ist ein klassisches 4-stufiges Verkehrsmodell, (z.B. [SL97], [SKV05] Kap. 5B):

1. Erzeugung: *Wie viele* Reisen werden je Verkehrszelle generiert?
2. Verkehrsverteilung: *Wohin* (zu welcher Zelle) gehen die Reisen?
3. Verkehrsmittelwahl: *Womit* (z.B. Pkw, Flug, etc.) erfolgt die Reise?
4. Verbindungswahl: *Wann* und *wo lang?* (hier: Mit welcher Zugverbindung?)

**Schwerpunkt** der Prognose in PRIMA ist die **Verkehrsmittel- und Verbindungswahl** in **Abhängigkeit** von den **fünf Angebotskenngrößen** einer Bahnreise: **Reisezeit**, **Fahrpreis**, **Anzahl und Dauer Umstiege** (zu knapp oder zu lang ist schlecht), **Zeitlage** (Wie oft/ wie gut passt das Angebot zur Nachfrage?), **Auslastung** (Wie viel Platz ist im Zug?).

PRIMA differenziert dazu aktuell 1.340 Verkehrszellen europaweit (460 in Deutschland), 4 Verkehrsmittel, 5 Reiseanlässe und 14 Tagestypen. Datengrundlagen klassischer Modelle sind räumlich aggregierte sozioökonomische Daten und Marktforschungen in Form stichprobenhafter Befragungen. Dieses Vorgehen ist state-of-the-art für neue Verkehrsmittel wie z.B. den Fernlinienbus in Deutschland (Abs. 3.2), für die es noch keine Nutzungsdaten gibt. Andererseits ist dieser Ansatz nicht fein genug auf dem Detailniveau von 13 Mio. Fernzugverbindungen je Tagestyp, die PRIMA generiert. Aktuelle

---

<sup>4</sup> Firma ptv AG, Karlsruhe

<sup>5</sup> Ausnahme: PRIMA kennt keine Ticketarten und Klassen

Ansätze, wie die genaue, detaillierte Kenntnis des Status quo durch TReK zur Verbesserung der Prognosen verwendet werden kann, werden in Abschnitt 3.3 vorgestellt.

PRIMA wird zur Bewertung von ca. 1.000 Fahrplanvarianten jährlich eingesetzt. Ein Szenario wird innerhalb weniger Stunden, i.d.R. über Nacht gerechnet. Fahrplanszenarien und Bewertungsergebnisse werden in einem eigenen Data Warehouse verwaltet und das aufgesetzte Berichtssystem ermöglicht – analog dem von TReK – standardisierte Auswertungen in zahlreichen Dimensionen insbesondere auch den paarweisen Vergleich von Fahrplanszenarien bzw. von Szenarien mit Ist-Daten.

Der hohe Automatisierungsgrad von PRIMA legt die Nutzung als *black box* nahe. Es ist uns jedoch wichtig, die Bewertungslogik transparent zu halten, damit Fahrplaner ihre eigene Einschätzung dagegen reflektieren und erfolgsversprechende Ideen generieren können. Auch hier helfen die Visualisierungsverfahren wie in 2.3 dargestellt erheblich.

### 3.2 Gemeisterte Herausforderungen: Fernlinienbus und Auslastung

Seit 2013 ist der Fernbusmarkt in Deutschland liberalisiert und Fernbuslinien dürfen nun in Konkurrenz zur Schiene angeboten werden. Der Markt entwickelt sich extrem dynamisch: Mittlerweile verkehren täglich mehr Fernbusse als Fernzüge im Linienbetrieb. Da es vor 2013 i.W. kein Fernlinienbusangebot gab, also Reisende im realen Verhalten nicht beobachtet werden konnten, war es eine große Herausforderung, die **Auswirkung des Fernbusses** als neues Verkehrsangebot **auf die Bahnachfrage** vorherzusagen.

Der empirische Input für das PRIMA-Verkehrsmittelwahlmodell ist eine Marktforschung in Form eines individualisierten *choice based conjoint* [GSG09] mit 5.700 Befragten im Fernverkehr. Jedem Befragten wurde zu seiner berichteten Reise neben den drei realen Alternativen Pkw, Bahn und Flugzeug auch der Fernbus als hypothetische Alternative angeboten. Die Alternativen wurden dabei durch Angebotskenngrößen beschrieben und die Wahlentscheidungen bei Variation selbiger festgehalten. Das auf Basis der Marktforschung mit dem *Maximum Likelihood* Ansatz ([BAL85], Kap. 2) geschätzte Verkehrsmittelwahlmodell prognostiziert, dass sich ein Fernbusangebot parallel zum Kernnetz der DB Fernverkehr AG rentieren kann, dass jeder 4. Busnutzer von der Bahn wechselt und sich die Busnachfrage auch bei steigenden Buspreisen in den kommenden Jahren etwa verdoppeln könnte.

Eine weitere große Herausforderung war die **Integration der Angebotskenngröße „Auslastung“** in das Nachfragemodell PRIMA. Zuvor konnten Züge im Modell ohne Rückwirkung auf die Nachfrage beliebig und unrealistisch überlastet werden, Verdrängung wurde nicht berücksichtigt. Aus Marktforschungen ist bekannt, dass ein einfaches Abschneiden der Besetzung bei 100% Auslastung kein geeignetes Modell ist: Die ersten Kunden springen bereits bei 50% Auslastung ab, die letzten sind sogar bereit zu stehen.

Die Integration der Kenngröße „Auslastung“ in das Modell war vor allem algorithmisch schwierig. Anders als die übrigen vier Kenngrößen hängt die Auslastung direkt von der prognostizierten Nachfrage ab, wodurch sich ein negativ rückgekoppelter Wirkungszusammenhang ergibt: Hohe Auslastung senkt die Bahnachfrage, diese wiederum senkt

die Auslastung. Um den Gleichgewichtszustand ([SKV05], Abschnitt 5.10) zwischen Angebot und Nachfrage zu finden, wurde ein intelligent gedämpftes, iteratives und sehr performantes Verfahren entwickelt. Die notwendige zweistellige Anzahl von Iterationen erfordert nur die doppelte Rechenzeit ggü. dem Modell mit nur vier Kenngrößen.

### 3.3 Aktuelle Herausforderungen und Ansätze zur weiteren Modellverbesserung

Planungen werden immer detaillierter (Einzelzüge, Verkehrstage) und stellen höhere Ansprüche an PRIMA: Hohe Prognosegüte auf hohem Detailniveau. TReK ist dabei für PRIMA Herausforderung und Chance zugleich. Es ermöglicht einerseits den Vergleich von Prognose vs. Ist auf maximalem Detailniveau und deckt damit die Schwächen eines Prognosemodells auf: Je höher das Detail, desto ungenauer sind statistische Modelle. Andererseits kann TReK eine Datengrundlage für ein besseres Modell sein. Wir sehen zwei Ansätze, Informationen aus TReK für PRIMA zu nutzen:

1. **Schätzung der Verhaltensparameter der (Zug)Verbindungswahl**  
 Als Alternative zu hypothetischen, stichprobenhaften Aussagen einer Marktforschung liefert TReK für jede Relation explizit, unter welchen Angebotsbedingungen welche Bahnverbindungen in der Realität gewählt werden.
2. **Inkrementelle Nachfrageprognose auf Basis IST-Nachfrage**  
 Aus PRIMA wird nur die Änderung der Nachfrage übernommen und auf die tatsächliche Nachfrage aus TReK angewendet. Gegenüber herkömmlichen Ansätzen [DFP12] möchten wir auf der Ist-Nachfrage nicht nur in räumlicher, sondern auch in zeitlicher Schneidung aufsetzen.

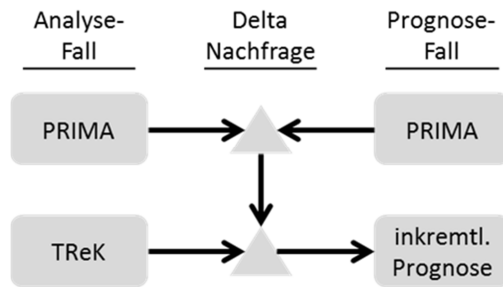


Abb. 4: Inkrementelle Nachfrageprognose auf Basis TReK

Um zu prognostizieren, wie viel Nachfrage die Bahn wann und wo von anderen Verkehrsmitteln gewinnen kann, ist es wichtig, deren Quelle-Ziel-Nachfrage genau zu kennen. Während das für den Flugverkehr aus Angebot und Buchungen der Fall ist, ist der MIV (motorisierte Individualverkehr) mit 80% Verkehrsleistungsanteil im Personenfernverkehr die große Unbekannte. Hier versprechen wir uns deutliche Fortschritte durch Nutzung anonymisierter Bewegungsdaten von Navigations- oder Mobilfunkgeräten.

Beide Datenquellen haben aber noch ihre Unzulänglichkeiten: Die Abdeckung von verfügbaren anonymisierten Pkw-Navigationsdaten beträgt derzeit nur etwa 1% des MIV-

Marktes und ist im Detail so disproportional, dass eine brauchbare Hochrechnung auf den MIV-Markt nicht ohne weiteres möglich erscheint. Die Verwendbarkeit von Mobilfunkdaten wird durch Anonymisierungstechniken sehr stark eingeschränkt. Der regelmäßige Wechsel der Telefon-ID verhindert die Nachvollziehbarkeit von längeren Fernreisen und die Nicht-Weitergabe von Einzelereignissen detaillierte Auswertungen.

## 4 Aktuelle Fragestellungen und Ausblick

Die durchgängige Digitalisierung der Gesellschaft und der Dinge (*internet of things*) prägen die Erwartungshaltungen unserer internen Kunden und die der Reisenden. Erstere erwarten einfachsten Zugang zu allen, stets top-aktuellen Daten auf allen Endgeräten. Letztere lernen, dass individuell relevante Ansprachen und Informationen von Unternehmen anderer Branchen beherrscht werden und erwarten dies zunehmend auch von uns. Daher halten wir umfassende, gut zugängliche Datengrundlagen und Fähigkeit diese als Marketingorganisation zweckorientiert nutzen zu können, für wesentliche Schlüsselkompetenzen. Unsere aktuellen Herausforderungen liegen nun insbesondere darin das Nützliche und Zweckmäßige vom technisch Möglichen zu trennen. Einige herausgehobene Themen und *trade-off* dazu sind:

- Wie kann viel *self-service* erreicht werden und wie viel/welche *Governance* soll dabei sein (u.a. Agilität Fachbereiche vs. Risiko Fehlinterpretationen)?
- Ist BI auf mobilen Endgeräten nützlich oder bloß schick?
- Wie aktuell müssen Daten sein ggü. wie vollständig/strukturiert?
- Welche Möglichkeiten zur persönlich relevanten Ansprache der Reisenden können wir u.a. mit neueren *Big-Data*, *predictive analytics* Ansätzen realisieren (z.B. *churn rate reduction*, *next best offer*, *liquid feedback*, ...)?

Die DB Fernverkehr AG strebt danach in den Augen der Reisenden und der Mitarbeiter erste Wahl zu sein. Wir freuen uns mit unseren Themenstellungen in Zeiten gesellschaftlicher Veränderungen und einem hochdynamischen technischen Umfeld dazu einen Beitrag zu leisten.

## Literaturverzeichnis

- [BAL85] Ben Akiva, M.; Lerman, S.: Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press 1985
- [DB14] Geschäftsbericht DB AG zum Jahr 2013
- [DFP12] Daly, A.; Fox, J.; Patrui, B.; Milthorpe, F.: Pivoting in Travel Demand Models. BTS Conference Paper, Australasian Transport Research Forum, 26-28 September 2012
- [GSG09] Günthel, D.; Sturm, L.; Gärtner, C.: Anwendung der choice-based-conjoint-Analyse zur Prognose von Kaufentscheidungen im ÖPNV. TU Dresden. Diskussionsbeiträge aus dem Institut für Wirtschaft und Verkehr Nr.1/2009
- [SKV05] Steierwald, S.; Künne, H-D.; Vogt, W.: Stadtverkehrsplanung: Grundlagen, Methoden, Ziele, Springer Berlin Heidelberg New York 2005
- [SL97] Schnabel, Lohse: Grundlagen der Strassenverkehrstechnik und der Verkehrsplanung. Verlag für Bauwesen, Berlin, 1997.