Data Refinement in a Market Research Applications’ Data Production Process

Thomas Ruf, Thomas Kirsche
GfK Marketing Services, Nuremberg


**Agenda**

1. GfK Retail & Technology
2. The **S**t**A**r**T**r**A**c**K** system
3. Data Refinement in Master Data Management
4. Data Refinement in Data-In
5. Data Refinement in Analysis & Production
6. Data Refinement in Reporting
GfK Mission Statement

Companies need to make decisions,
Knowledge is the best for decision-making.

Our business information services provide the essential knowledge that industry, retail and service sectors and the media need in order to make their decisions.

As a knowledge provider, we aim to be at the top in all the global markets in which we operate – in the interest of our clients, our employees, our company, our shareholders and the general public.
## GfK Group: key features (1)

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sales</strong></td>
<td>in 2004: EUR 672 million</td>
</tr>
<tr>
<td></td>
<td>Previous year: EUR 595 million</td>
</tr>
<tr>
<td></td>
<td>Increase on the previous year: approx. +12.8%</td>
</tr>
<tr>
<td><strong>Employees</strong></td>
<td>5,539 full-time staff</td>
</tr>
<tr>
<td></td>
<td>nearly 75% of which outside Germany</td>
</tr>
<tr>
<td><strong>Services</strong></td>
<td>Comprehensive information services on consumer goods, healthcare and services markets</td>
</tr>
<tr>
<td></td>
<td>Research know how</td>
</tr>
<tr>
<td></td>
<td>Marketing consultancy</td>
</tr>
<tr>
<td><strong>Network</strong></td>
<td>About 120 subsidiaries, branches and participations in 61 countries on five continents</td>
</tr>
</tbody>
</table>
GfK Group: key features (II)

**Core business**
Pure market research player delivering information services to major global players in the consumer goods, services and healthcare industries.

**Market position**
Largest market research company in Germany, 4th largest in Europe and 5th in the world.

**USP**
Synergies across our business divisions: out of the top 50 clients 39 are consulting GfK in more than one business division.
## GfK Historic stages

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 1999</td>
<td>GfK shares start trading on the Frankfurt stock exchange</td>
</tr>
<tr>
<td>1990</td>
<td>Conversion into a joint stock company</td>
</tr>
<tr>
<td>1967</td>
<td>Start of internationalization</td>
</tr>
<tr>
<td>1949</td>
<td>Re-opening of GfK</td>
</tr>
<tr>
<td>1934</td>
<td>Founded as GfK-Nürnberg e.V.</td>
</tr>
<tr>
<td>1925</td>
<td>Founded: “Institut für Wirtschaftsbeobachtung der deutschen Fertigware”</td>
</tr>
</tbody>
</table>
GfK Group Sales by division¹)

1) in per cent

- Custom Research: 37.9%
- Retail and Technology: 28.2%
- Consumer Tracking: 14.2%
- Media: 9.4%
- HealthCare: 10.3%
GfK Retail and Technology: key services

Focus
Information services in 61 countries regarding marketing, logistics and sales in industry and retail for companies operating in the consumer technology markets

The advantage for clients
Direct access to databases and/or transmission of standard analyses used to support, track and manage short, medium and long-term planning of product and pricing policies, advertising, distribution, sales and logistics

Market position
Market leader in Europe, Asia and the Pacific as well as in the Arab countries; market leader in the US, in conjunction with our partner NPD Intelect
Example: Number of digital cameras sold by region

Source: GfK Marketing Services
Our Product
Our Product
The StarTrack system
What is S*T*A*R*T*R*A*C*K?
It is NOT:
- What is STARTACK? -

**STARTACK** = **S**ystem **T**o **A**nalize and **R**eport on **T**RACKing **data**

- New international production system of GfK Retail & Technology division
- Fully productive since March 2004
- Replaced inflexible mainframe system by a modern technology
- Relational database backend, VB & Active X frontend
- Why START Track? -

Global Markets, Global GfK ➔ Global System
Concepts of STARTRACK

- One global platform for the entire Marketing Services workflow
- Easy Integration of additional countries into one global system
- Global Consolidation in coding and reporting:
  - Global Codeplans (Feature Linking on MDM)
  - Common Coding (on MDM, -> global hitlists in DWH)
  - Global Reporting Structures (Library Objects in DWH)
  - Web access
- Flexible and fast => shorter reporting periodicities
Creating value through knowledge for our clients
**StarTrack Quality Benefits**

- **International Standardization**
  - consistent and consolidated product and shop master data base
  - global code plans
  - standardized segmentations in international reporting tools

- **Improved Data Checks**
  - improved shop quality check routines
  - easy to use and very flexible production analysis tool

- **Improved Report Production**
  - clear text definition reporting tool
  - centralized set up and maintenance of report pages
Kernaufgabe: „Kippen“ von Datenlieferungen zu Marktberichten

Warengruppenübergreifende Datenlieferung

Warengruppenbezogener Bericht

Geschäft\(_1\)  \quad \text{Bericht}_1

Geschäft\(_n\)  \quad \text{Bericht}_n

Geschäft\(_3\)  \quad \text{Bericht}_3

Geschäft\(_2\)  \quad \text{Bericht}_2
STARTRACK Integrated Workflow

**QUALITY CHECK**
- GfK WinCos Separation
- Raw Tracking

**QUALITY CHECK**
- WebTAS
- MDM items & shops

**QUALITY CHECK**
- DWH Suite
- DWH Extrapolation
- DWH Builder
- DWH QC project

**QUALITY CHECK**
- IDAS Info System
- IDAS Output Pool
- IDAS2DWH
- IDAS2BH
- "Good" data

**QUALITY CHECK**
- Preprocessing
- Fact Tool
- DWH Loader

**QUALITY CHECK**
- Reporting projects
- Base projects
- DWH Reporting

Data flow: 
- Identified data
- Data Release
- Data Order

Control flow: 
- To Do

Vision: 
- Data Orders

Dataflow:
- DWH2Tools
- DWH Explorer
- DWH Reporting

Controlflow:
Data Refinement in Master Data Management
# GfK MS product world structure

<table>
<thead>
<tr>
<th>Product Group Sectors</th>
<th>Product Groups (Selection)</th>
<th>Examples for Features and Their Feature Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large Appliances, Small Kitchen Appliances, Jewellery, Furniture</td>
<td>Washing Machines, Vacuum Cleaners, Shavers, Watches, Kitchen Furniture</td>
<td>Water Consumption (in litres), Cable Rewind (with, without), Dust Bags (with, without), etc.</td>
</tr>
<tr>
<td>Consumer Electronics</td>
<td>TV Sets, VCRs, Car Radios, DVD Players</td>
<td>Screen Size (in inches), Sound System (Stereo, Mono), TV Norm (NTSC, PAL, SECAM, PAL/SECAM multi), Remote (with/without), Number of Video Heads (2,4,5)</td>
</tr>
<tr>
<td>IT/Telecom</td>
<td>Cellular Phones, PCs, Printers, Software</td>
<td>Band (Single-, Dual- or Tri-band), Mounting Form (Mini Tower, Tower, 19”, Barebone), Pages per Minute, Application Field (learning game, entertainment, home office)</td>
</tr>
</tbody>
</table>
Item Quality Check via Feature Rules

Feature Rule

IF Screen size >= 55 cm
THEN Videotext = Yes

Browser
Mail via AutoNot

PGR II520
CRT-TV

<table>
<thead>
<tr>
<th>Item ID</th>
<th>Screen size</th>
<th>Videotext</th>
</tr>
</thead>
<tbody>
<tr>
<td>1714</td>
<td>55 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>1715</td>
<td>74 cm</td>
<td>No</td>
</tr>
<tr>
<td>1716</td>
<td>36 cm</td>
<td>No</td>
</tr>
<tr>
<td>1717</td>
<td>74 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>
## Feature Quality Control: Check Productgroup

<table>
<thead>
<tr>
<th>Rule Name</th>
<th>Mandatory</th>
<th>Last Edited</th>
<th>Edited By</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVD Audio → MP3</td>
<td>90%</td>
<td>Montag, 6. September 2004 Thomas Kirsche</td>
<td></td>
</tr>
</tbody>
</table>

**Description**
Most devices that support DVD audio playback also have MP3 playback

**Rule Body**
IF [DVD AUDIO (F50001)] = [DVD AUDIO (FV273994)] THEN [MP3 PLAYBACK (F58606)] = [MP3 PLAYBACK (FV258074)]

### Column Filter
- TABLE/PORTABLE
- DVD AUDIO
- SUPER AUDIO CD
- HEIGHT IN MM
- SHOW VIEW
- MP3 PLAYBACK
- REAL DOLBY
- PHOTO-CD
- PANAS/TECHNICS
- COMBI
- LOUDSPEAKERS
- SVCD
- EAN-CODES
- DVD RECORDING
- DIVX PLAYBACK
- BRAND
- EXCLUSIV/PROD ZA
- SERVER FUNCTION
- HDD CAPACITY
- PRODUKT URL
- SERVER FUNCTION

### Column Sort
- Layoutgroup, pos. PD
- Default Columns
- Global Features
- Standard Features

- check

Check only items with status: Existing

Check Product Group
Data refinement in Data-In
STARTRACK - IDAS
International Data Acquisition System

General Interface Manager: Conversion to GfK format

Check for unknown articles

Classification of unknown articles

Output Pool

Data is checked and available for production

WebTAS

Separation Run

Known

Unknown

MDM

Raw Tracking Data

Shop delivery quality control
**S.T.A.R.T.R.A.C.K - IDAS**

-> WINCOS

→ Checks the quality of the retailer deliveries

<table>
<thead>
<tr>
<th>Time</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 Mio</td>
<td>0</td>
<td>500T</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Delivery would have to be checked

Delivery OK from a volume point of view

Tunnel of acceptance for a retailer
Data delivery checks in WinCOS

Central Administration of Company Data and Error Borders

Error Borders - E-Pool Determination

- **BKZ**
  - Deviation previous week: Records ≤ 20
  - Purchase ≤ 30
  - Sales ≤ 30
  - New Stock ≤ 20
  - Turnover ≤ 20

- Deviation previous month: see Channels & Groups

- **RPG**
  - Deviation previous week / month: LB ≤ 50
  - Deviation previous year: LB ≤ 50

Error Check

To find out, if an error in the previous month should be shown, a check to the previous year is made.

- previous year: ≤ 20
Data Refinement in Data Analysis & Production
StarTrack Workflow

DWH

1. Take-over of
   - ProductGroups (incl. Features) and
   - Channels (incl. Features) from the MDM into DWH

2. Set-up of
   - Base Projects
   - Reporting Projects
   - Production Projects
   - Extrapolations

3. Definition of DataOrders and sending them to IDAS

4a. Get DataOrder related data on a Shop-ProductGroup-DeliveryPeriod level

4b. Extrapolation relevant compensations

5a. Saving compensations as LoadDefinition and Load

5b. LoadDefinition to IDAS

6a. Loading tracking data into LoadingTable

6b. Execution of compensations and loading into QC project

7a+b. Corrections and release to DWH

8. Reporting and Tools Export:
   - Inmarkt Express
   - Model Express
   - QuickView
   - Cobras
   - Excel
   - ...

IDAS

MDM

IOP Status Info Pool

IDAS Output Pool (IOP)
Automated price check & correction algorithm

DWH PriceCheck for the calc. price

Pricecheck Parameters - Productgroup - <DIG CAMERAS DE>

<table>
<thead>
<tr>
<th>General</th>
<th>Excluded Outlets I</th>
<th>Excluded Outlets II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method to calculate reference price:</td>
<td>Median</td>
<td></td>
</tr>
<tr>
<td>Number of valid prices needed to calculate reference price</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Order of reference price inclusion</td>
<td>Extend to Total market within current period first</td>
<td></td>
</tr>
<tr>
<td>Reject prices used to calculate reference price (no MEDIAN):</td>
<td>Lowest Prices [%]</td>
<td>Highest Prices [%]</td>
</tr>
<tr>
<td>Maximal deviation from reference price:</td>
<td>Lower Bound [%]</td>
<td>Upper Bound [%]</td>
</tr>
<tr>
<td>Maximal deviation from last period</td>
<td>Lower Bound [%]</td>
<td>Upper Bound [%]</td>
</tr>
<tr>
<td>Check only prices that are out of absolute range</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method to calculate correction price:</td>
<td>Reference Price</td>
<td></td>
</tr>
</tbody>
</table>

Are there enough valid mentionings in the current period per channel? 

YES  NO

Are there enough valid mentionings in the current period in the whole market? 

YES  NO

Are there enough valid mentionings all together in the current period and the pre period? 

YES  NO

User selects an average for calculation of reference price:
- Median
- Weighted average
- User average

User defines the Number of mentionings to calculate a reference price as mentionings per outlet

User defines absolute range of valid prices outside this range are wrong and don't influence reference price calculation

User defines % and absol. Range of price which influence the average. (Not needed if median)

User eventually excludes shops from reference price:

User eventually excludes shops from the price check:

User defines % deviation of prices from period price of same outlet (separately for lower and upper bound)

User defines % deviation of prices from the price found within the range if you exceed the % Range

User defines subproductgroups (not yet possible)

Price correction: the system is going to put in the reference price

Is the a reference price on the base of enough mentions? 

YES  NO

Price is going to be listed for manual input/release or working. The user is going to do the input needed (put the flag for subproductgroups that the price is not going to influence next period as a mentionning and not as a price of the pre period)
Fact manipulation settings

DWHExplorer - Fact Manipulation Form

- Sales Units
  - Recalculate Stock Old Units
  - Recalculate Purchase Units
  - Recalculate Stock New Units
- Purchase Units
  - Recalculate Stock Old Units
  - Recalculate Sales Units
  - Recalculate Stock New Units
- Stock New Units
  - Recalculate Stock Old Units
  - Recalculate Sales Units
  - Recalculate Purchase Units
- Stock Old Units
  - Recalculate Sales Units
  - Recalculate Purchase Units
  - Recalculate Stock New Units

- Negative Value Handling
  - Allow
  - Abort
  - Best Effort

- Update Values Action
  - Recalc current Sheet
  - Erase current Sheet
  - nothing

- Generate integer values
- Exclude Outlets

Default OK Cancel
6 Data Refinement in Reporting
DWH Explorer – Definition View
Thank You