

# The Rocky Horror

# Horror

Data Show  
meets Ghostbusters:  
Keine Angst vor Datengeistern



# Einleitung

## Auf dem Weg ins Gruselkabinett...

- IT-PS Data & AI Team
  - Hat schon einige Projekte "erlebt"
  - Und auch darüber berichtet, aus Sicht des Use Cases
- Heutiger Workshop
  - Etwas anderer Blickwinkel
  - Unter dem **Motto "Datenstrategie"**
  - Denn viele Dinge passieren aufgrund des Umfeldes
- Workshop-Inhalte
  - Diskussion von Praxisproblemen in der Runde
  - Und deren Lösungen



# Vor dem Gruselkabinett

"Treten Sie näher, treten Sie ein!"

- Willkommen im Gruselkabinett
- Reise durch ein fiktives **Alptraum-Projekt**
- In Anlehnung an Erlebnisse aus unseren Projekten, allerdings überspitzt dargestellt und komprimiert
- Einzelne Projektphasen als **Anstoß für Diskussion:**
  - Use Case Idee (Projekt-Geburt)
  - Projektdefinition und Konzept
  - Projektstart
  - Prototyping
  - Produktionalisierung



# Projekt - Geburt

Oder die Ausgeburt eines Use Cases...

- Geschäftsführer (GF) und IT-Leiter haben in (nicht sehr seriöser) Sonntagsbeilage über AI gelesen
- Montag Früh beim Kaffee:
  - IT-Leiter: *"Schau, wir sammeln seit Jahren Daten in der Cloud... Mit den Daten kann man sicher was machen"*
  - GF: *"In der Zeitung hab ich gelesen, LLMs sind der Hit... Können wir den Teams-Chatverlauf unserer Mitarbeiter dazu verwenden, ihre Arbeitsmotivation zu messen?"*



# Vertreibung der Datengeister

Erfahrungsaustausch, Diskussion und Ideensammlung...



# Vertreibung der Datengeister

## Wichtige Punkte

- **Business-Problem** muss im Vordergrund stehen (nicht Werkzeuge / Daten)
- **Zweck** der Datensammlung wichtig
- **Owner** für Daten notwendig
- Kenntnis der **Regulatorien** wichtig (für C-Level und Mitarbeiter)
- **Prozess für geordnete Use Case Auswahl** hilfreich für Einhaltung der Regulatorien und für Erfolgchancen des Use Cases



# Projektdefinition und Konzept

"Willkommen zurück im Gruselkabinett!"

- GF setzt Meeting mit seinen besten Mitarbeitern an
- Derzeitige Produktionsmengensteuerung wird erklärt:
  - Eher Kunst als Wissenschaft
  - Erfahrene Mitarbeiter verwenden ihr "Bauchgefühl"
- Brainstorming über mögliche Lösungsvarianten
- GF bestimmt, dass ein Algorithmus implementiert wird, der Entscheidungen dieser Mitarbeiter nachbildet

*"Wenn Wetter schlecht & Kunde mies gelaunt, dann mehr Ware produzieren"*

# Vertreibung der Datengeister

Erfahrungsaustausch, Diskussion und Ideensammlung...





# Vertreibung der Datengeister

## Wichtige Punkte

- Ziel aller Daten-Initiativen: **Aus den Daten lernen**
  - Über das Unternehmen
  - Über die Zielerreichung
  - und die Abläufe
- **Wunsch nach Lernen** muss vorhanden sein  
(sonst Data Governance und Datenprojekte sinnlos)
- **Bereitschaft für Veränderung** muss vorhanden sein



# Projektstart

"Kommen sie weiter!"

- Daten wurden am Abend vor dem Kickoff übergeben
- enthalten Spalten wie X\_ORD\_DT\_PREV und X\_ORD\_DT2\_NX, ohne weiteren Kommentar
- Kickoff Meeting: Team kommt zusammen um das Projekt zu starten
- Gleich zu Beginn:
  - Experte aus Fachbereich: *"Ich hab keine Zeit für so ein Meeting",  
"und außerdem: ich stehe auch im Projekt nur eingeschränkt zur Verfügung"*

# Vertreibung der Datengeister

Erfahrungsaustausch, Diskussion und Ideensammlung...



# Vertreibung der Datengeister

## Wichtige Punkte

- **Ressourcen** wie Budget und Zeit (!) müssen **verfügbar** gemacht werden
- **Dokumentation** der Daten ist notwendig
  - Je mehr Datenprojekte, desto mehr rentiert sich das Investment
- **Zentrale Dokumentation** (Data Catalog) ist nachhaltig
  - Nachvollziehbarkeit
  - Rückführbarkeit der Infos und Learnings aus Datenprojekten



# Data Catalog

## Daten finden und verstehen



The screenshot displays the IBM Watson Knowledge Catalog interface for a dataset named 'Customer Information File'. The top navigation bar includes 'IBM Watson', 'Projects', 'Tools', 'Catalog', 'Community', and 'Services'. The current profile shows 163 classifiers, 14 columns, and 48 rows. Below this, there are five columns of classifier distribution charts for 'PARTY\_ID', 'FIRST\_NAME', 'LAST\_NAME', 'BIRTHDATE', and 'GENDER'. Each chart shows the percentage of data points classified by various categories. For example, the 'LAST\_NAME' classifier shows 100% for 'Inferred'. Below the classifier distributions are three frequency charts for 'PARTY\_ID', 'FIRST\_NAME', and 'BIRTHDATE', showing the count of data points for each unique value. The 'PARTY\_ID' chart shows values from 100004 to 100010. The 'FIRST\_NAME' chart shows names like NANCY, SEAN, JOE, DALE, ALEX, DAVID, and CONSUELO. The 'BIRTHDATE' chart shows dates from 1986-08-28 to 1983-09-15. The 'GENDER' chart shows a distribution between 'M' and 'F'. The interface also includes an 'Update Profile' button and a 'View Log' link.

IBM  
Knowledge  
Catalog



# Prototyping

"Willkommen zurück im Gruselkabinett!"

Während Bearbeitung des UCs stellt sich heraus:

- Daten von vergangenen Produktionsmengen-Vorgaben vorhanden
- Aber:
  - Vorgaben wurden von Schichtleiter angepasst
  - Und Produktionsmenge ist durch Maschinen-Ausfälle vermindert
- Beides wurde nicht systematisch erfasst!



# Vertreibung der Datengeister

Erfahrungsaustausch, Diskussion und Ideensammlung...



# Vertreibung der Datengeister

## Wichtige Punkte

- Start der **Datendokumentation** im Business
  - Daten gehören zu Business-Domänen, nicht der IT
- Dokumentation der **Business-Prozesse** führt zur vollständig(er)en Daten
- Auswirkung auf Datenprojekte:
  - Dieses Projekt muss das Ziel anpassen an "Daten-Gegebenheiten"
  - Nächstes Datenprojekt hat zusätzliche Datenquellen zur Verfügung
- Findings aus Datenprojekten informieren die (Umsetzung der) **Datenstrategie**





# Produktionalisierung

*"Kommen Sie weiter zur letzten Attraktion im Gruselkabinett!"*

- Prototyp fertig, Prognose funktioniert gut: 95% der Prognosen machen einen akzeptablen Fehler von kleiner als 17 Stk. Produktionsmenge
- GF: *"Super Ergebnis! Ich hätte gerne täglich Vorhersagewerte für unsere Produktion! Ab wann steht das Tool für unsere Schichtleiter zur Verfügung?"*
- ...langes Schweigen...
- *"Ähem... Wir wissen nicht, wie wir das Datenprodukt im Unternehmen zur Verfügung stellen können..."*

# Vertreibung der Datengeister

Erfahrungsaustausch, Diskussion und Ideensammlung...



# Vertreibung der Datengeister

## Wichtige Punkte

- Verwendung der Lösung, sowie...
- **Einbettung der Lösung** müssen von Anfang an mitgedacht werden
- **Datenprodukt und Dokumentation** dazu gemeinsam zur Verfügung stellen



# Ausgang aus dem Gruselkabinett

Rocky Horror Data Show: "That's all, folks!"



# Zusammenfassung

"Hört nur! Ich glaube, ich rieche was." (Ghost Busters: Dr. Raymond Stanz (Dan Aykroyd))

- Zentraler Punkt:
  - Datenchaos verhindert sinnvolles Bild des Unternehmens und kann sogar schädlich sein
  - **Ordnung** in Daten bringen!
- **Maßnahmen** treffen, die **Ordnung begünstigen:**
  - Data Governance
  - Datenstrategie
  - Data Literacy



# Data Catalog

## Daten finden und verstehen



The screenshot displays the IBM Watson Knowledge Catalog interface for a 'Customer Information File'. The top navigation bar includes 'IBM Watson', 'Projects', 'Tools', 'Catalog', 'Community', and 'Services'. The breadcrumb trail shows 'Catalogs / Transform Catalog / Customer Information File'. The 'Profile' tab is active, showing 'Current profile: 163 classifiers' and 'Last profile: 8 Mar 2018 - 12:02 pm'. The file has 14 columns and 48 rows. Below this, a table lists columns: PARTY\_ID (Varchar), FIRST\_NAME (Varchar), LAST\_NAME (Varchar), BIRTHDATE (Varchar), and GENDER (Varchar). Each column has a corresponding 'CLASSIFIER DISTRIBUTION' chart showing the percentage of data inferred. For example, 'Last Name' is 100% inferred, while 'Date of Birth' is 94% inferred. Below the distributions are 'FREQUENCY' charts for each column, showing the count of unique values. For instance, the 'LAST\_NAME' frequency chart shows 'GOMEZ' as the most frequent name with 4 occurrences. The interface also includes 'Add to Catalog', 'Delete', and 'Update Profile' buttons.

IBM Knowledge Catalog





Alice: "Würdest du mir bitte sagen,  
wie ich von hier aus weitergehen soll?"

"Das hängt zum großen Teil davon ab,  
wohin du möchtest", sagte die Katze.

Lewis Carroll  
Alice im Wunderland  
1865





Der strategische Partner für AI.

[www.it-ps.at](http://www.it-ps.at)

**Ingo Nader**

Senior Data Scientist

[ingo.nader@it-ps.at](mailto:ingo.nader@it-ps.at)

**IT-Power Services GmbH**

Modcenterstraße 14, 1030 Wien