

## Data Analytics @ Schwabmünchen

Ingo Hild, Dr. Bernd Eberhard, Dr. Jürgen Almanstötter, Martina Flörchinger

### **Inhalt**



1.	<ul> <li>Ingo Hild, Dr. Bernd Eberhard: Einführung</li> <li>Data Analytics @ Schwabmünchen</li> <li>Wertstrom / ETL &amp; Cloud</li> </ul>	~ 5 Min.
2.	Ingo Hild, Dr. Bernd Eberhard: Daten & Produktionsverfahren  • Entwicklung & Implementierung eines Backtrackingsystems entlang des Wolfram-Wertstroms	~ 10 Min.
3.	<ul> <li>Dr. Jürgen Almanstötter: Algorithmen &amp; Datenanalyse</li> <li>Technische Prozessanalyse mit Deep-Learning</li> <li>Dynamic Time warping (DTW): Ein Algorithmus zur Prozessoptimierung</li> </ul>	~ 15 Min.
4.	<ul> <li>Martina Flörchinger: Scripting &amp; Ad-hoc Analyse</li> <li>Ad-hoc Analyse</li> <li>Tutorial: Data wrangling toolbox – von Big Data zu neuen Einsichten</li> </ul>	~ 15 Min.
<b>5</b> .	Ingo Hild, Dr. Bernd Eberhard: Wrap Up	~ 5 Min.

### **Inhalt**



1.	Ingo Hild, Dr. Bernd Eberhard: Einführung  • Data Analytics @ Schwabmünchen  • Wertstrom / ETL & Cloud	~ 5 Min.
2.	Ingo Hild, Dr. Bernd Eberhard: Daten & Produktionsverfahren  • Entwicklung & Implementierung eines Backtrackingsystems entlang des Wolfram-Wertstroms	~ 10 Min.
3.	<ul> <li>Dr. Jürgen Almanstötter: Algorithmen &amp; Datenanalyse</li> <li>Technische Prozessanalyse mit Deep-Learning</li> <li>Dynamic Time warping (DTW): Ein Algorithmus zur Prozessoptimierung</li> </ul>	~ 15 Min.
4.	<ul> <li>Martina Flörchinger: Scripting &amp; Ad-hoc Analyse</li> <li>Ad-hoc Analyse</li> <li>Tutorial: Data wrangling toolbox – von Big Data zu neuen Einsichten</li> </ul>	~ 15 Min.
5.	Ingo Hild, Dr. Bernd Eberhard: Wrap Up	~ 5 Min.

### Data Analytics @ Schwabmünchen

#### Wertstrom – vom APW zum Wolframdraht









Endprodukt





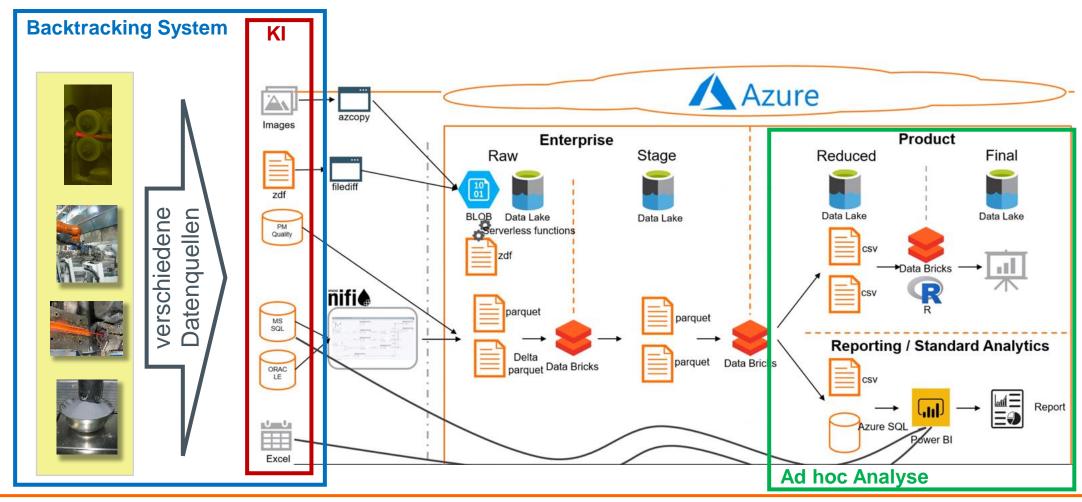




Wolframdraht

### Data Analytics @ Schwabmünchen

#### Wertstrom und Infrastruktur der Daten



vom Wertstrom zu einer effizienten Datenlandschaft!



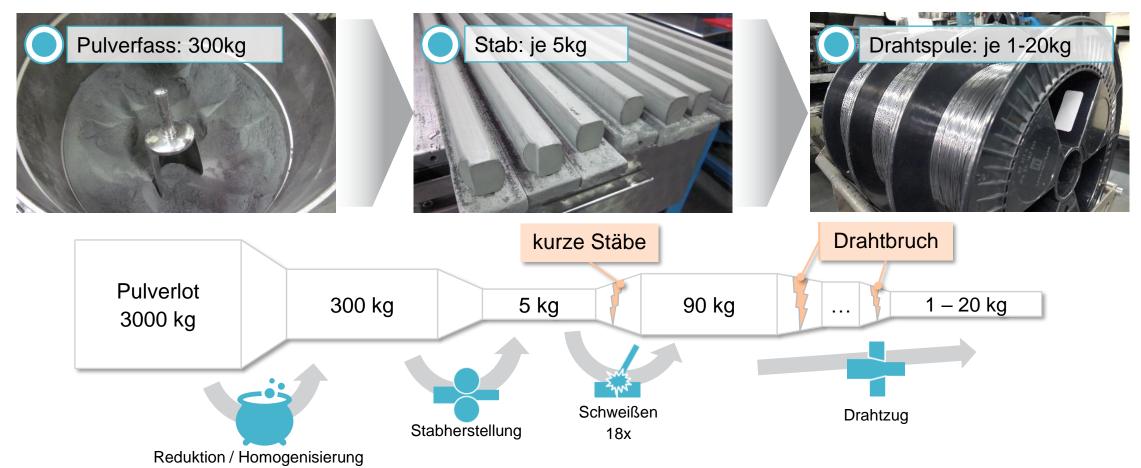
### **Inhalt**



1.	Ingo Hild, Dr. Bernd Eberhard: Einführung  • Data Analytics @ Schwabmünchen  • Wertstrom / ETL & Cloud	~ 5 Min.
2.	Ingo Hild, Dr. Bernd Eberhard: Daten & Produktionsverfahren  • Entwicklung & Implementierung eines Backtrackingsystems entlang des Wolfram-Wertstroms	~ 10 Min.
3.	<ul> <li>Dr. Jürgen Almanstötter: Algorithmen &amp; Datenanalyse</li> <li>Technische Prozessanalyse mit Deep-Learning</li> <li>Dynamic Time warping (DTW): Ein Algorithmus zur Prozessoptimierung</li> </ul>	~ 15 Min.
4.	<ul> <li>Martina Flörchinger: Scripting &amp; Ad-hoc Analyse</li> <li>Ad-hoc Analyse</li> <li>Tutorial: Data wrangling toolbox – von Big Data zu neuen Einsichten</li> </ul>	~ 15 Min.
5.	Ingo Hild, Dr. Bernd Eberhard: Wrap Up	~ 5 Min.

#### Über den Materialfluss

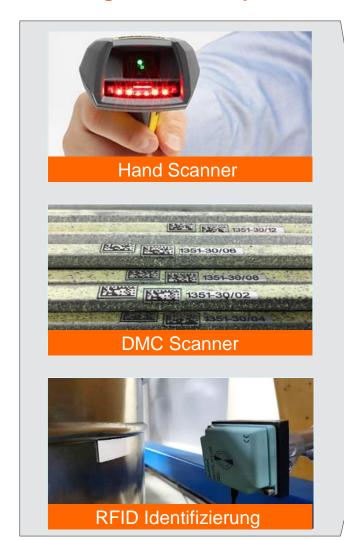


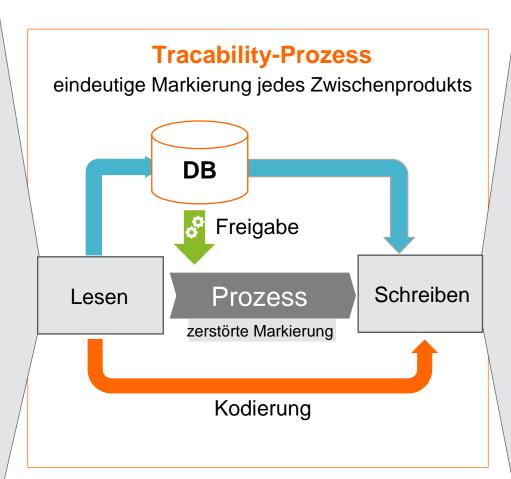


Rückverfolgbarkeit ist notwendig für die Multiprozessanalyse – Aber leider nicht einfach in der W-Drahtherstellung!

#### technologisches Konzept









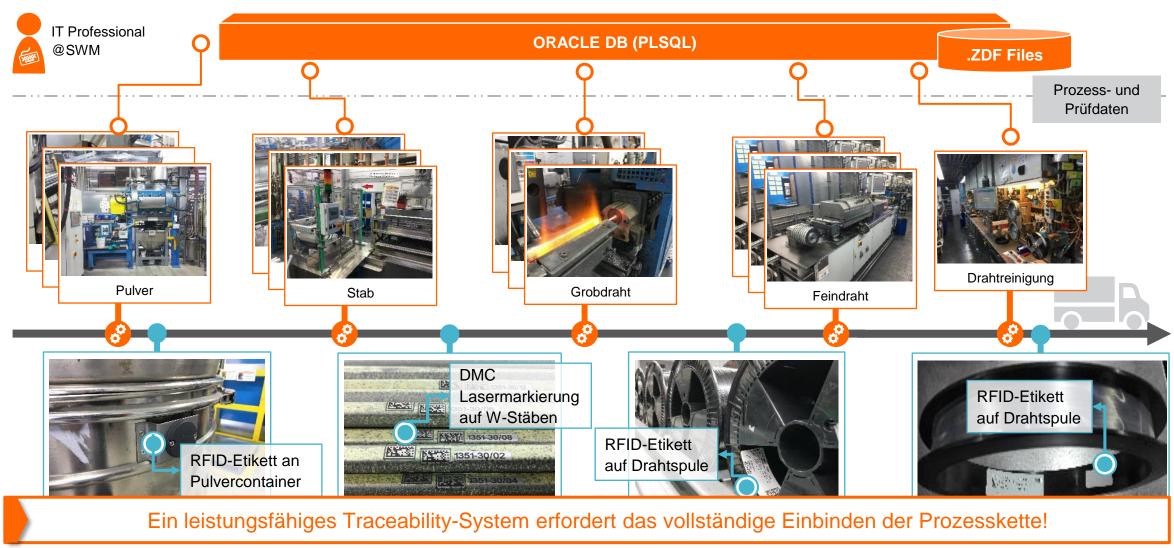
#### Der Draht bekommt einen Namen und zugehörigen Stammbaum

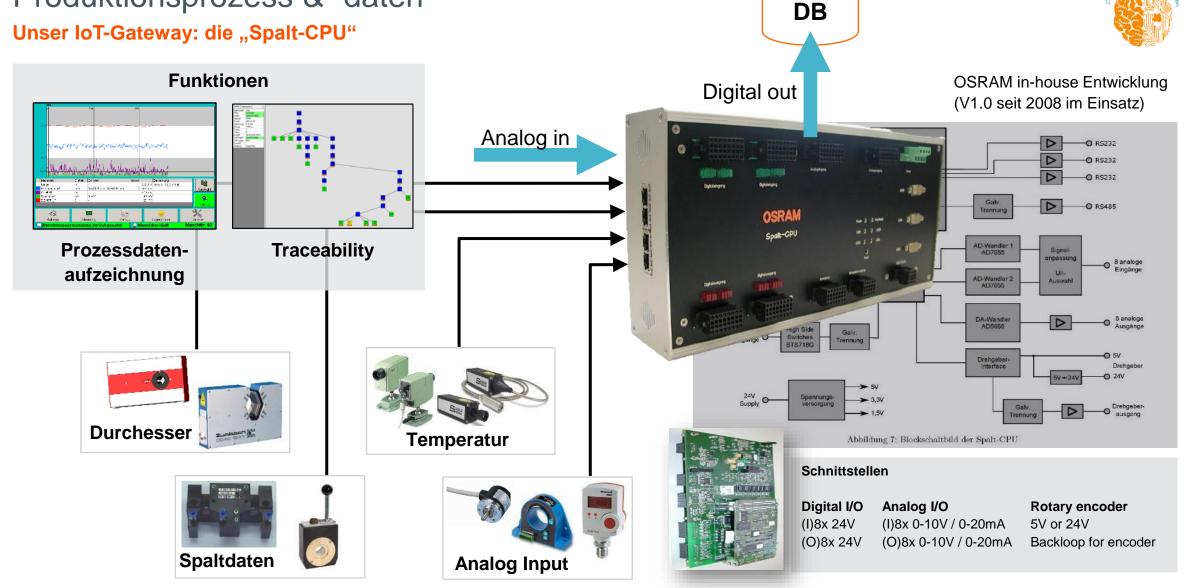






Implementierung entlang der kompletten Prozesskette (~60 Prozessschritte, 34x Lesen und Schreiben)

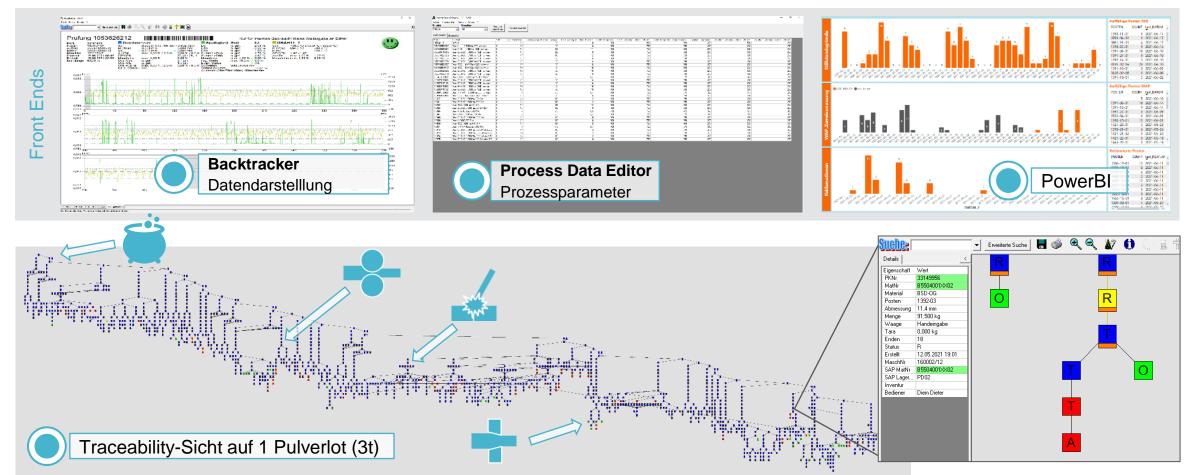






#### **Datenvisualisierung & -interaktion**





Shopfloor Management: tägliche Datenvisualisierung und -zeitreihenbeurteilung der gesamten Prozesskette!

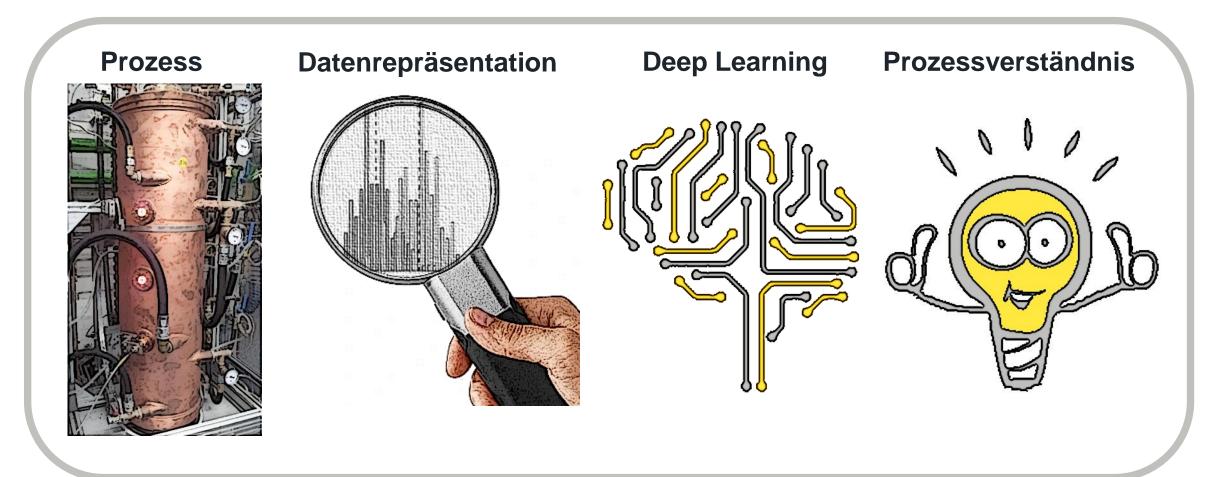
### **Inhalt**



1.	Ingo Hild, Dr. Bernd Eberhard: Einführung  • Data Analytics @ Schwabmünchen  • Wertstrom / ETL & Cloud	~ 5 Min.
2.	Ingo Hild, Dr. Bernd Eberhard: Daten & Produktionsverfahren  • Entwicklung & Implementierung eines Backtrackingsystems entlang des Wolfram-Wertstroms	~ 10 Min.
3.	<ul> <li>Dr. Jürgen Almanstötter: Algorithmen &amp; Datenanalyse</li> <li>Technische Prozessanalyse mit Deep-Learning</li> <li>Dynamic Time warping (DTW): Ein Algorithmus zur Prozessoptimierung</li> </ul>	~ 15 Min.
4.	<ul> <li>Martina Flörchinger: Scripting &amp; Ad-hoc Analyse</li> <li>Ad-hoc Analyse</li> <li>Tutorial: Data wrangling toolbox – von Big Data zu neuen Einsichten</li> </ul>	~ 15 Min.
5.	Ingo Hild, Dr. Bernd Eberhard: Wrap Up	~ 5 Min.

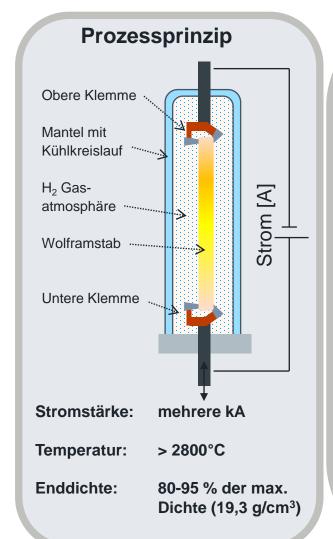
Zielsetzung: Generierung von interpretierbaren technischen Zusammenhängen

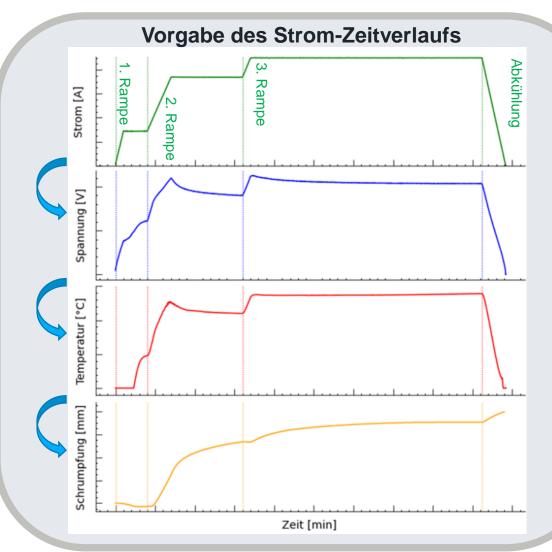


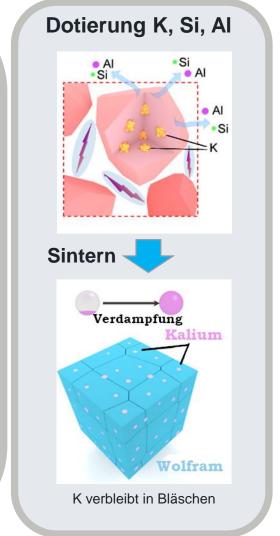


#### Prozess: Sintern von Wolframstäben



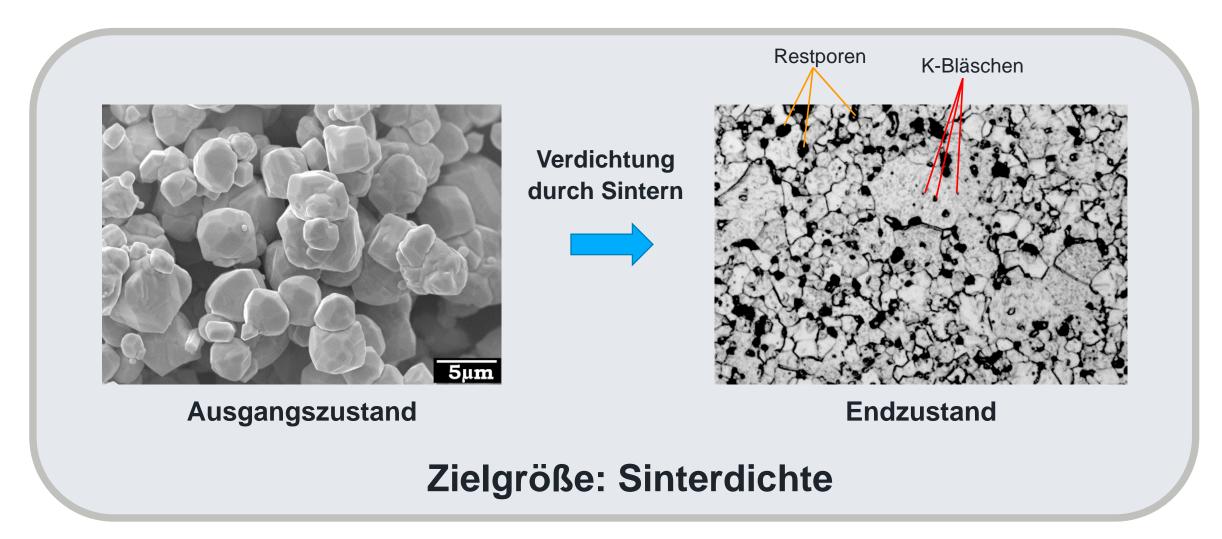






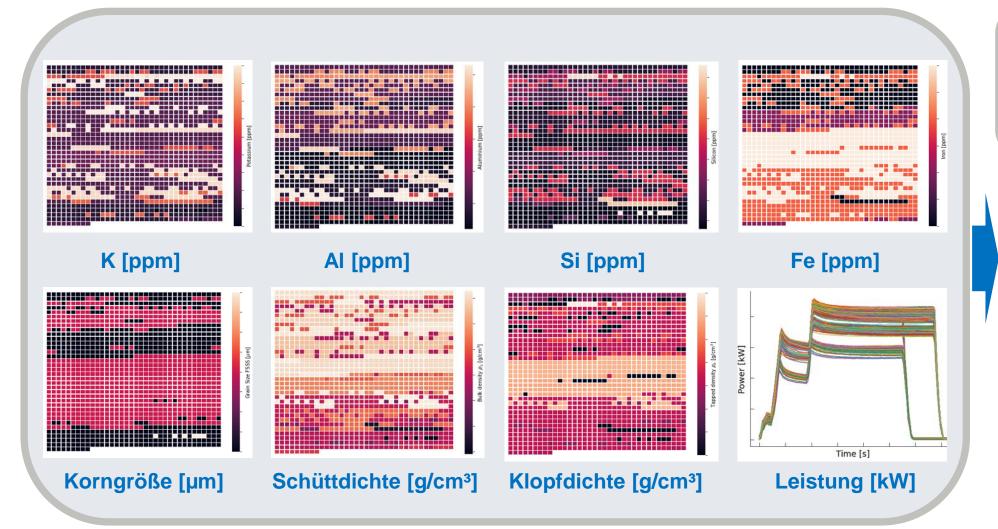
**Prozesswirkung auf das Material** 

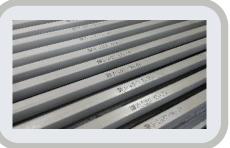


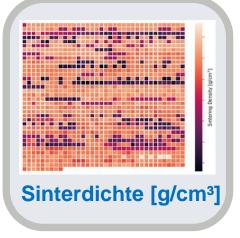


Datenrepräsentation: Datensätze von 1270 Wolframstäben









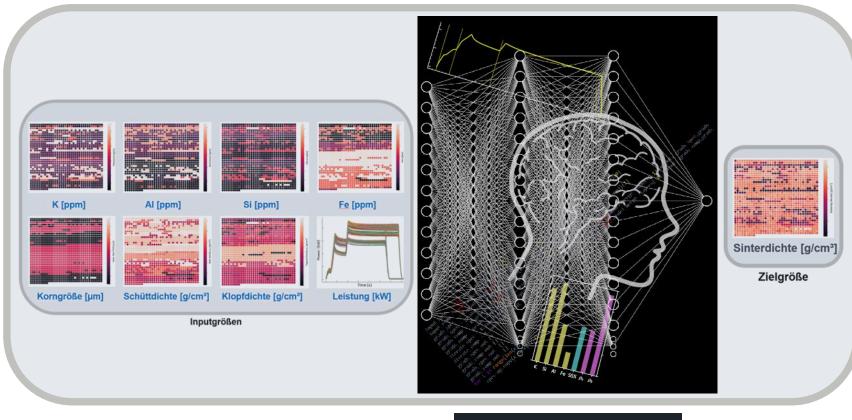
Zielgröße

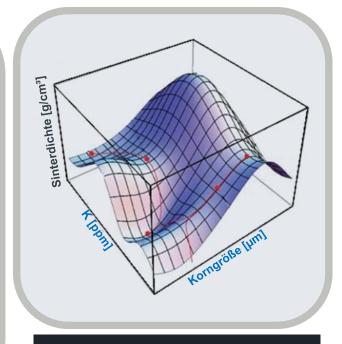




#### **Programmierung und Training des neuronales Netzwerks**







Unbekannte hochdimensionale Funktion  $f(x_1,x_2,\dots)$ 

Neuronales Netzwerk

Repräsentation











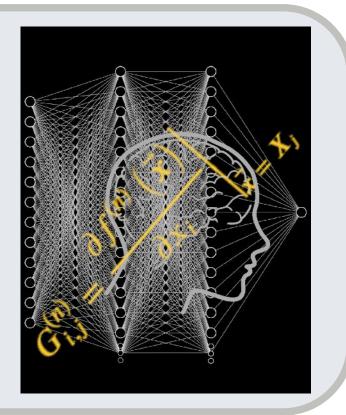
#### **Multidimensionaler Gradient**

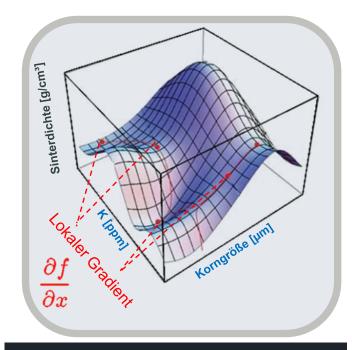


Zweck des trainierten Netzwerks ist <u>nicht</u> eine <u>Vorhersage</u> der Sinterdichte

**Ziel:** Ermittlung des **Gradienten** zur Darstellung der Variation der Zielgröße unter Änderung der Inputgrößen

**Ergebnis:** Quantitatives Maß für den Einfluss der einzelnen Prozeßparameter auf die Sinterdichte



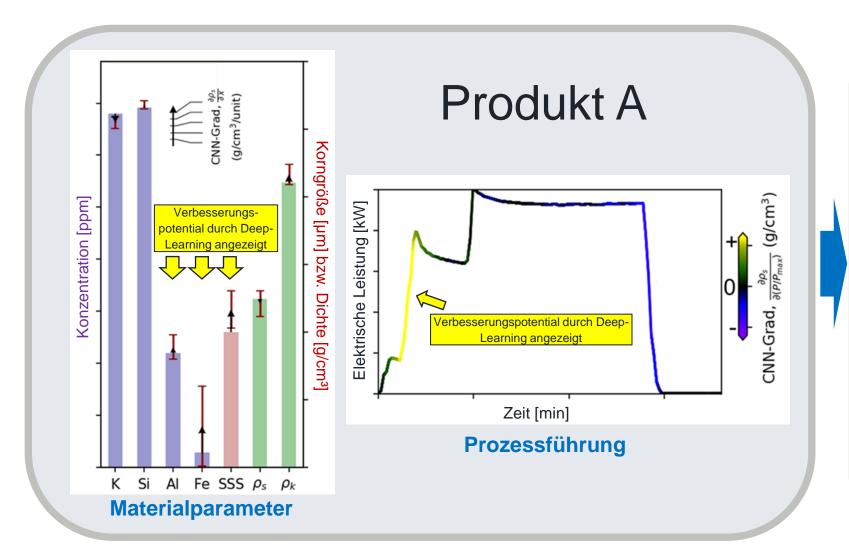


Berechnung des lokalen Gradienten von f(x<sub>1</sub>,x<sub>2</sub>,...) mittels trainiertem neuronalen Netzwerk



**Anwendungsfall: Erhöhung Sinterdichte** 





## Maßnahmen zur Erhöhung der Sinterdichte

#### 1. Materialparameter

- Erhöhung Fe-Konzentration
- Erhöhung Korngröße (SSS)
- Erhöhung Al-Konzentration

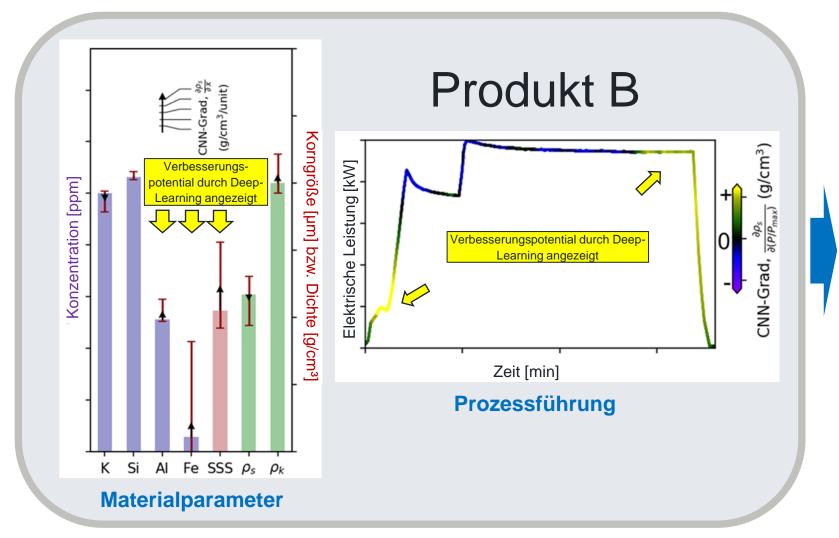
#### 2. Prozessführung

Verbesserungspotential an2. Rampe



#### Einfluss der Prozessparameter auf die Sinterdichte





## Maßnahmen zur Erhöhung der Sinterdichte

#### 1. Materialeigenschaften

- Erhöhung Fe-Konzentration
- Erhöhung Korngröße (SSS)
- Erhöhung Al-Konzentration

#### 2. Prozessführung

 Verbesserungspotential bei 1. Haltezeit und in Endstufe der 3. Haltezeit

**Zusammenfassung & Fazit** 



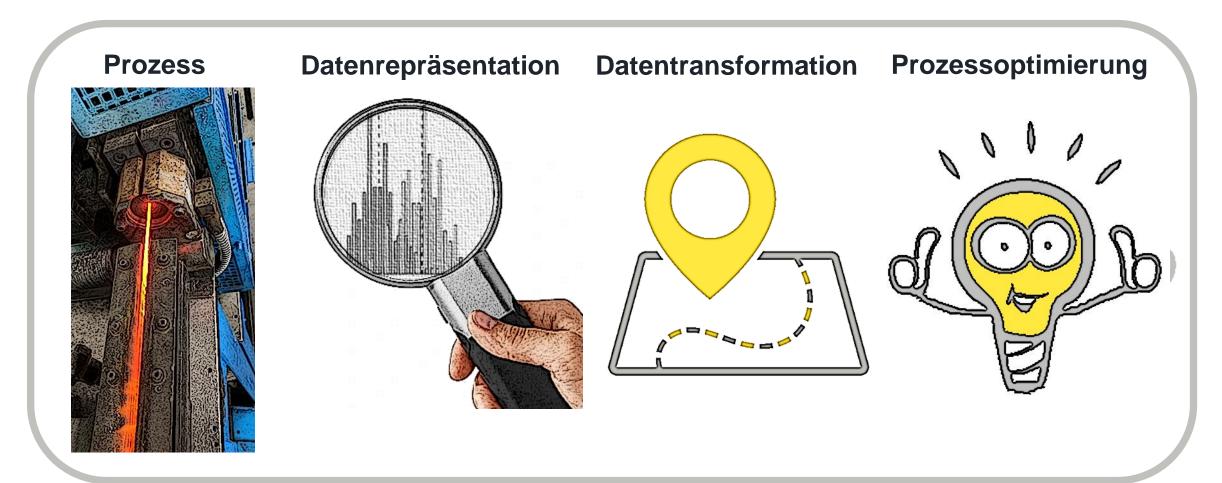
Die Analyse von Gradienten in einem trainierten neuronalen Netzwerk zeigt Schlüsselzusammenhänge, die durch traditionelle statistische Analysen nicht ohne weiteres identifiziert werden können

Die Gradientenanalyse automatisiert die Vorhersage der Richtungen im Parameterraum (Konzentration spezifischer Dop-Elemente oder Prozessparameter) zur Optimierung der Zielgröße (Sinterdichte)

Die vorgestellte Methode ist universell und kann auf eine Vielzahl von Problemen angewendet werden, von Produktionsprozessen bis zu F&E-Projekten

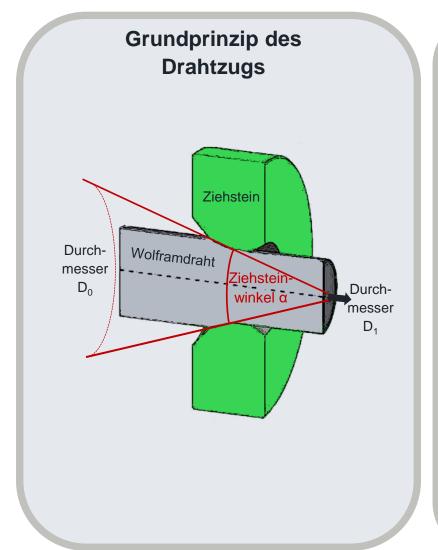
**Zielsetzung: Prozessoptimierung** 

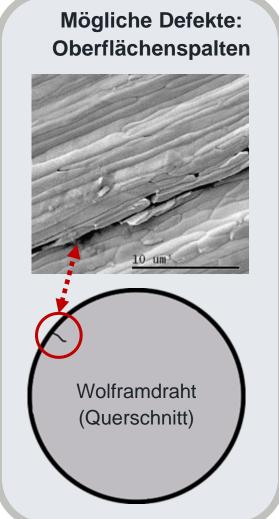


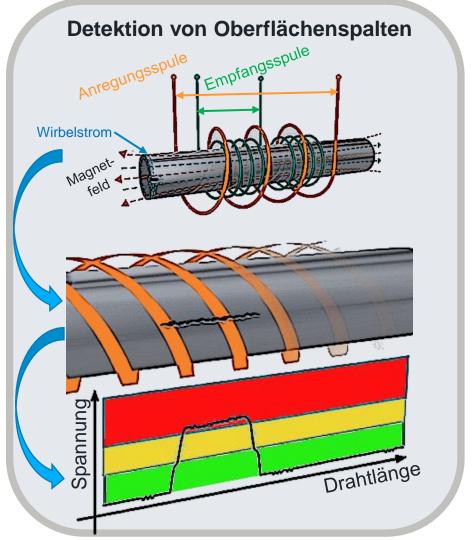


#### Prozess: Ziehen von Wolframdrähten



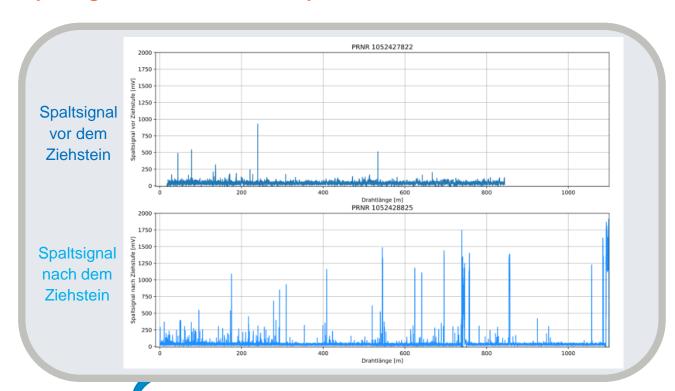






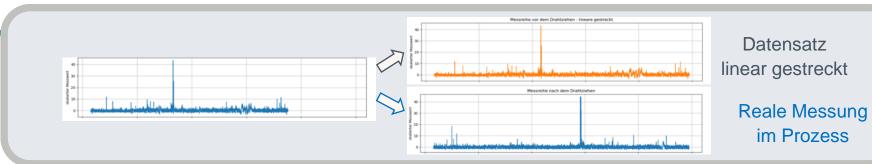
#### **Spaltsignal vor und nach Ziehpozess**





## Herausforderungen in der Datenanalyse

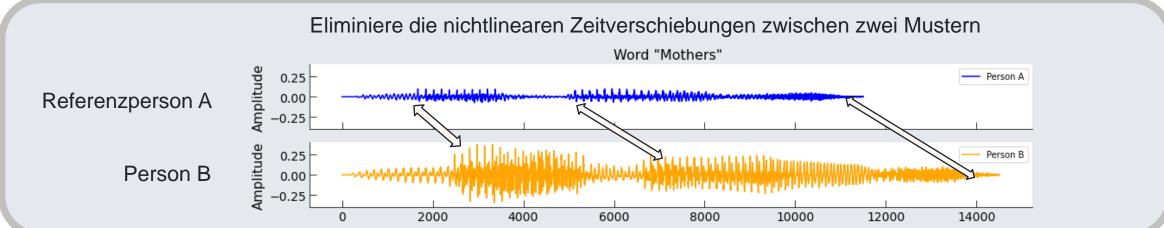
- 1. Signale haben unterschiedliche Gesamtlänge
- 2. Nichtlineare Streckung über die Drahtlänge
- Entstehung neuer Risse und Änderung existierender Spalten
- 4. Sehr große Datensätze: Auflösung 1 cm bei bis zum mehreren 1000 m Gesamtdrahtlänge





# **Verwandte Problemstellung: Spracherkennung**





IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-26, NO. 1, FEBRUARY 1978

#### Dynamic Programming Algorithm Optimization for Spoken Word Recognition

HIROAKI SAKOE AND SEIBI CHIBA

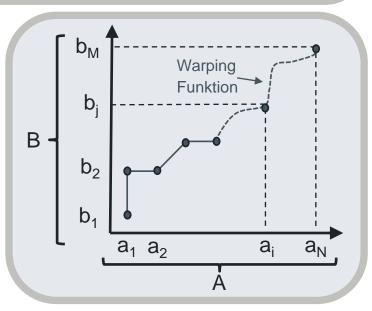
Abstract-This paper reports on an optimum dynamic programs (DP) based time-normalization algorithm for spoken word recognition. First, a general principle of time-normalization is given using timewarping function. Then, two time-normalized distance definitions, called symmetric and asymmetric forms, are derived from the principle These two forms are compared with each other through theoretical discussions and experimental studies. The symmetric form algorithm superiority is established. A new technique, called slope constraint, is successfully introduced, in which the warping function slope is restricted so as to improve discrimination between words in different categories. The effective slone constraint characteristic is qualitatively analyzed. and the optimum slope constraint condition is determined through experiments. The optimized algorithm is then extensively subjected to experimental comparison with various DP-algorithms, previously applied to spoken word recognition by different research groups. The experiment shows that the present algorithm gives no more than about twothirds errors, even compared to the best conventional algorithm

vestigations were made, based on the assumption that speech patterns are time-sampled with a common and uniform sampling period, as in most general cases. One of the problems discussed in this paper involves the relative superiority of either a symmetric form of DP-matching or an asymmetric one. In the asymmetric form, time-normalization is achieved by transforming the time axis of a speech pattern onto that of the other. In the symmetric form, on the other hand, both time axes are transformed onto a temporarily defined common axis. Theoretical and experimental comparisons show that the symmetric form gives better recognition than the asymmetric one. Another problem discussed concerns slope constraint technique. Since too much of the warping function flexibility sometimes results in poor discrimination between words in different

### **Gegenseitige Abbildung**

Suche den Pfad mit minimaler Distanz (i,j) um maximale Koinzidenz der Signale zu erreichen

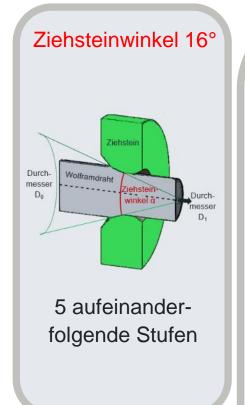
$$d(C_n) = d(i,j) = ||a_i - b_j||$$

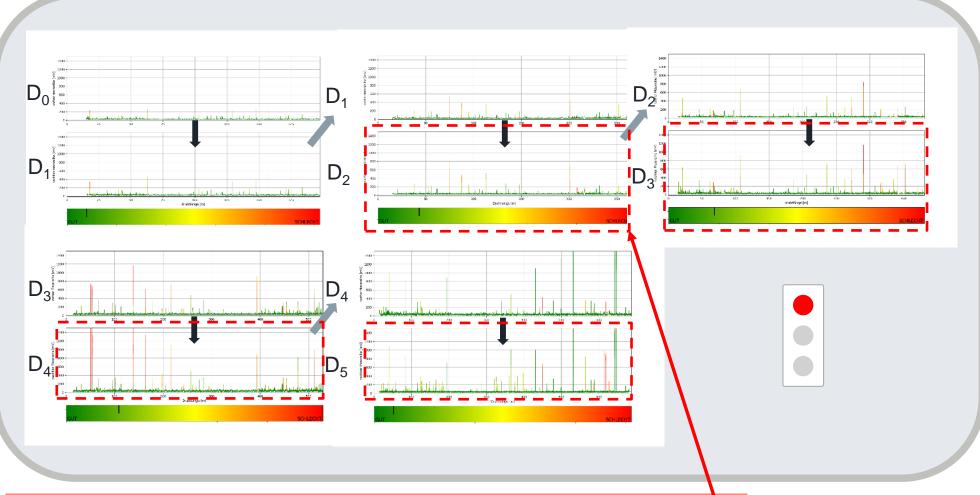




**Anwendung: Optimierung des Ziehsteinwinkels hinsichtlich Spaltigkeit** 





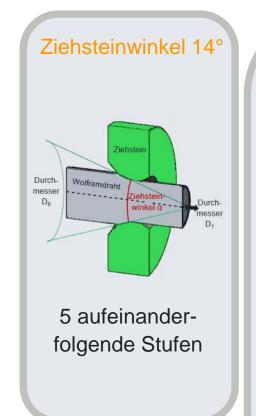


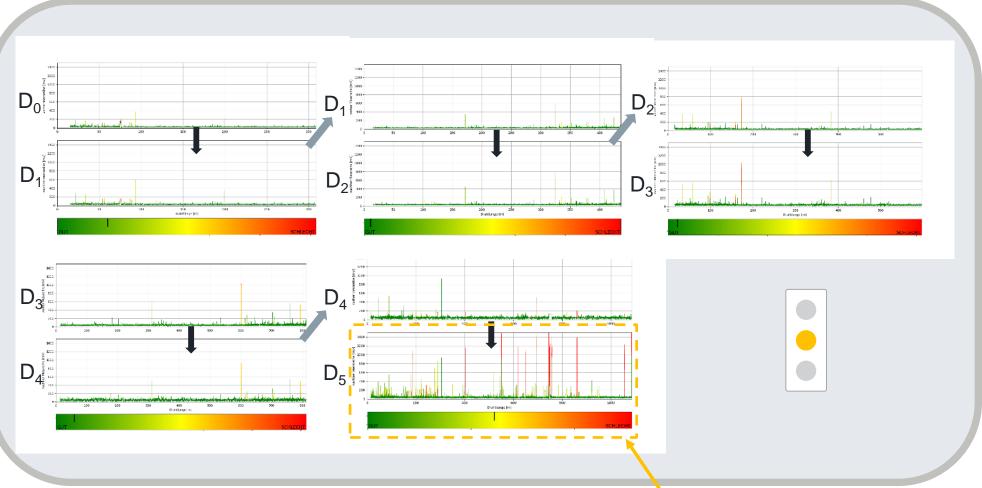
→ Starke Zunahme der Anzahl von Spalten bereits in der zweite Ziehstufen

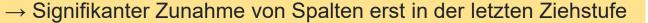








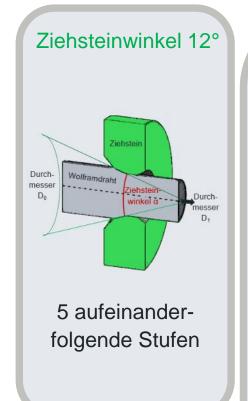


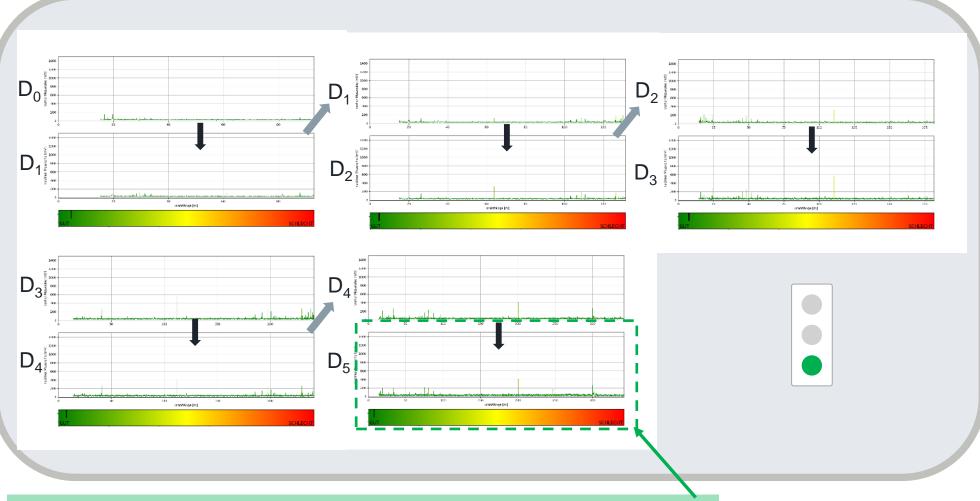












→ Keine signifikante Zunahme von Spalten → Optimaler Ziehsteinwinkel



**Zusammenfassung & Fazit** 



Der DTW-Algorithmus dient allgemein der Erkennung von Ähnlichkeiten zwischen Signalen oder Zeitreihen

Die Anwendung des DTW-Algorithmus ermöglicht die Verfolgung der Änderung von Spaltsignalen mit hoher Auflösung entlang der Drahtlänge

Nutzen: Identifikation kritischer Prozessparameter (hier Ziehsteinwinkel) über die Wertschöpfungskette

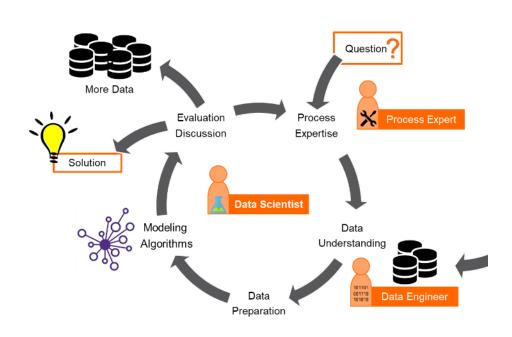
### **Inhalt**

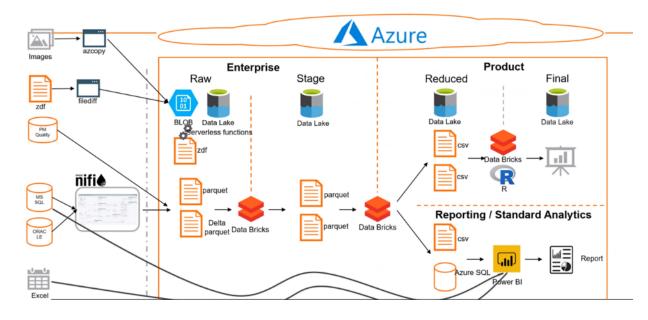


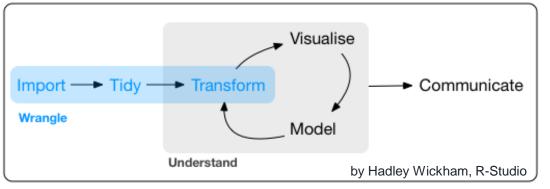
1.	Ingo Hild, Dr. Bernd Eberhard: Einführung  • Data Analytics @ Schwabmünchen  • Wertstrom / ETL & Cloud	~ 5 Min.
2.	Ingo Hild, Dr. Bernd Eberhard: Daten & Produktionsverfahren  • Entwicklung & Implementierung eines Backtrackingsystems entlang des Wolfram-Wertstroms	~ 10 Min.
3.	<ul> <li>Dr. Jürgen Almanstötter: Algorithmen &amp; Datenanalyse</li> <li>Technische Prozessanalyse mit Deep-Learning</li> <li>Dynamic Time warping (DTW): Ein Algorithmus zur Prozessoptimierung</li> </ul>	~ 15 Min.
4.	<ul> <li>Martina Flörchinger: Scripting &amp; Ad-hoc Analyse</li> <li>Ad-hoc Analyse</li> <li>Tutorial: Data wrangling toolbox – von Big Data zu neuen Einsichten</li> </ul>	~ 15 Min.
5.	Ingo Hild, Dr. Bernd Eberhard: Wrap Up	~ 5 Min.

#### Eine praxisnaher Herangehensweise um Einsichten aus Daten zu generieren













Wir starten mit "Big Data" – Laut Definition sind das viele, unsortierte und schnell neu generierende Datensätze

Big Data

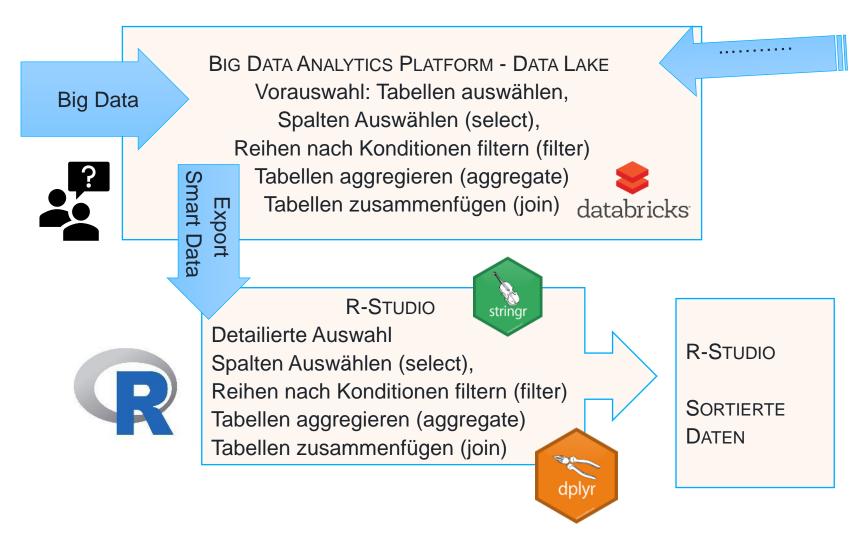


BIG DATA ANALYTICS PLATFORM - DATA LAKE
Vorauswahl: Tabellen auswählen,
Spalten Auswählen (select),
Reihen nach Konditionen filtern (filter)
Tabellen aggregieren (aggregate)
Tabellen zusammenfügen (join)



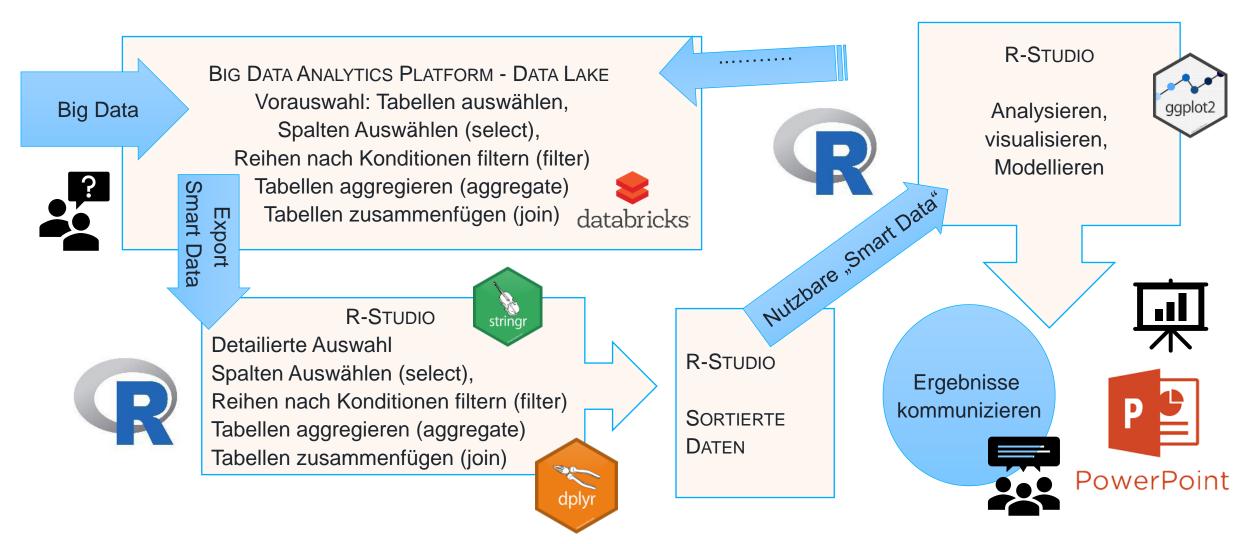


Wir starten mit "Big Data" – Laut Definition sind das viele, unsortierte und schnell neu generierende Datensätze





Wir starten mit "Big Data" – Laut Definition sind das viele, unsortierte und schnell neu generierende Datensätze



#### Beispiel





#### FRAGE:

Korreliert die Anzahl der Teilungen in der Drahtfertigung mit Fehlermerkmalen in der Warenausgangsprüfung?

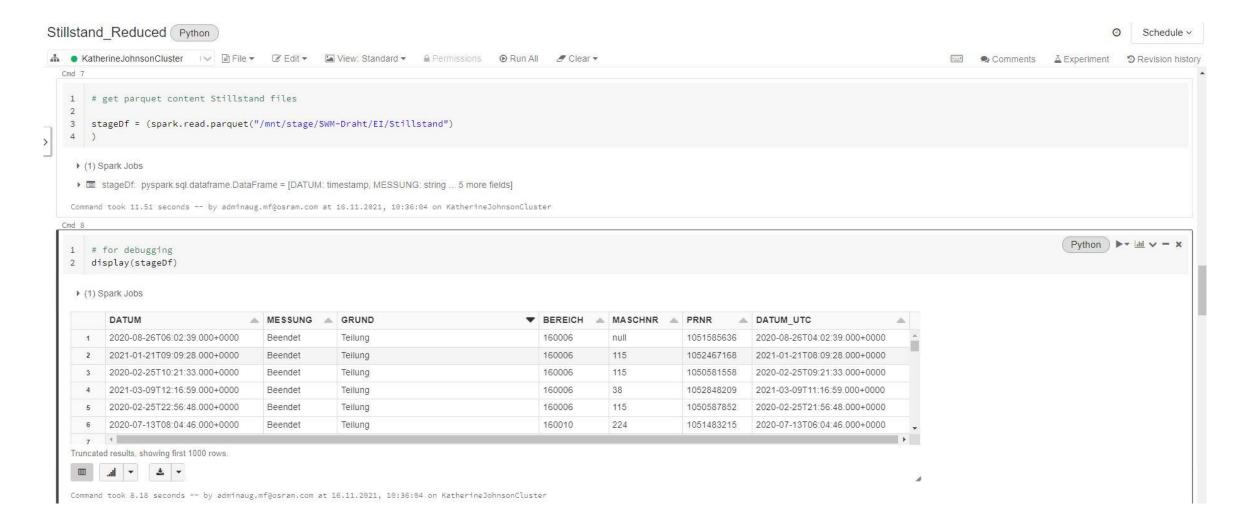
#### HERAUSFORDERUNG:

Eine neue Datenquelle ist noch nicht im Datalake, sondern nur per Excel zugänglich.





#### Erste Schritte: Extraktion der benötigten Daten via Databricks





#### Filtern der benötigten Daten und Export der Daten

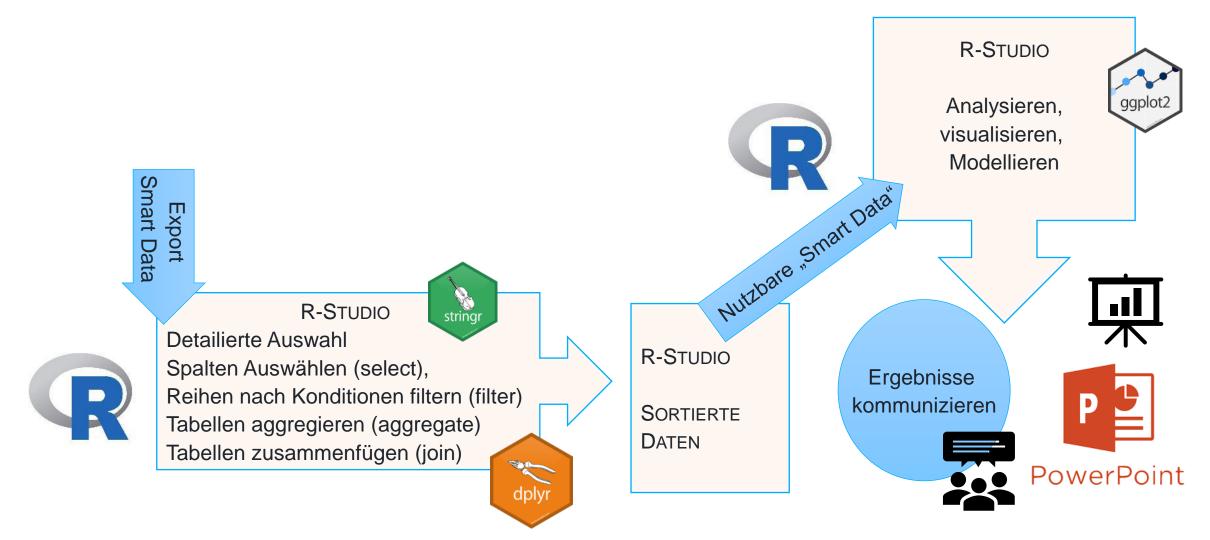


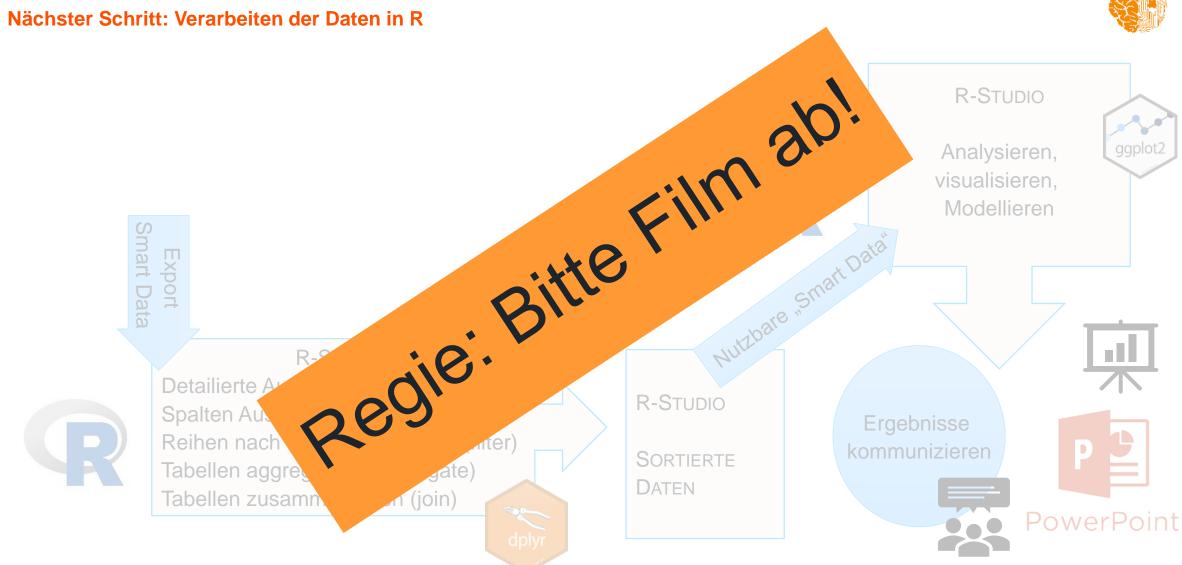
	DATUM	MESSUNG A	GRUND ▼	BEREICH A	MASCHNR A	PRNR 🔺	DATUM_UTC	$\triangle$
1	2020-08-26T06:02:39.000+0000	Beendet	Teilung	160006	null	1051585636	2020-08-26T04:02:39.000+0000	
2	2021-01-21T09:09:28.000+0000	Beendet	Teilung	160006	115	1052467168	2021-01-21T08:09:28.000+0000	
3	2020-02-25T10:21:33.000+0000	Beendet	Teilung	160006	115	1050581558	2020-02-25T09:21:33.000+0000	
4	2021-03-09T12:16:59.000+0000	Beendet	Teilung	160006	38	1052848209	2021-03-09T11:16:59.000+0000	
5	2020-02-25T22:56:48.000+0000	Beendet	Teilung	160006	115	1050587852	2020-02-25T21:56:48.000+0000	
6	2020-07-13T08:04:46.000+0000	Beendet	Teilung	160010	224	1051483215	2020-07-13T06:04:46.000+0000	
7	2020-12-01T22:47:05.000+0000	Beendet	Teilung	160006	115	1052215004	2020-12-01T21:47:05.000+0000	
uncat m	d took 8.18 seconds by adminaug.	.mf@osram.com at	16.11.2021, 10:36:04 on KatherineJohnsonClust	er				
		mf@osram.com at .	16.11.2021, 10:36:04 on KatherineJohnsonClust	er				
em omman	d took 8.18 seconds by adminaug. educedDf = stageDf.filter(coloreducedDf: pyspark.sql.dataframe.Dat	("GRUND") == "1 aFrame = [DATUM	Teilung").filter(col("DATUM_UTC") > '20 timestamp, MESSUNG: string 5 more fields]	020-01-01T00:	00:00.000+0000	'')		
em omman	d took 8.18 seconds by adminaug. educedDf = stageDf.filter(coloreducedDf: pyspark.sql.dataframe.Dat	("GRUND") == "1 aFrame = [DATUM	reilung").filter(col("DATUM_UTC") > '20 timestamp, MESSUNG: string 5 more fields]	020-01-01T00:	00:00.000+0006	(')		



Nächster Schritt: Verarbeiten der Daten in R









### **Inhalt**

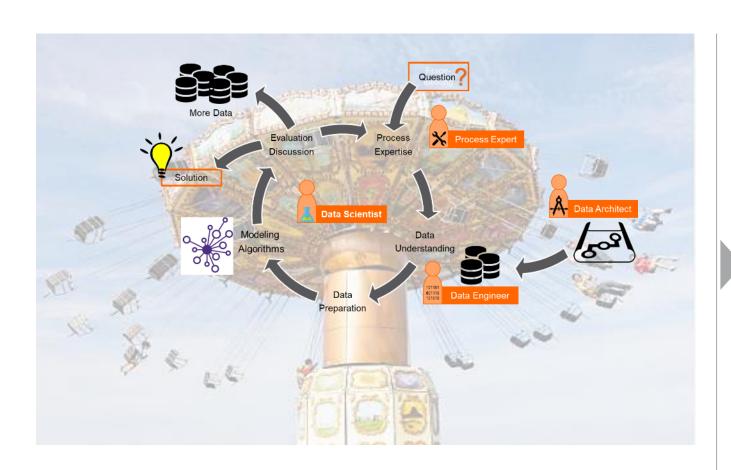


1.	Ingo Hild, Dr. Bernd Eberhard: Einführung  • Data Analytics @ Schwabmünchen  • Wertstrom / ETL & Cloud	~ 5 Min.
2.	Ingo Hild, Dr. Bernd Eberhard: Daten & Produktionsverfahren  • Entwicklung & Implementierung eines Backtrackingsystems entlang des Wolfram-Wertstroms	~ 10 Min.
3.	<ul> <li>Dr. Jürgen Almanstötter: Algorithmen &amp; Datenanalyse</li> <li>Technische Prozessanalyse mit Deep-Learning</li> <li>Dynamic Time warping (DTW): Ein Algorithmus zur Prozessoptimierung</li> </ul>	~ 15 Min.
4.	<ul> <li>Martina Flörchinger: Scripting &amp; Ad-hoc Analyse</li> <li>Ad-hoc Analyse</li> <li>Tutorial: Data wrangling toolbox – von Big Data zu neuen Einsichten</li> </ul>	~ 15 Min.
5.	Ingo Hild, Dr. Bernd Eberhard: Wrap Up	~ 5 Min.

### Erkenntnis #1

#### Data Analytics ist ein fortlaufender Prozess





### **Ergebnis:**



daher: schnelles und häufiges Durchlaufen des Cycles notwendig

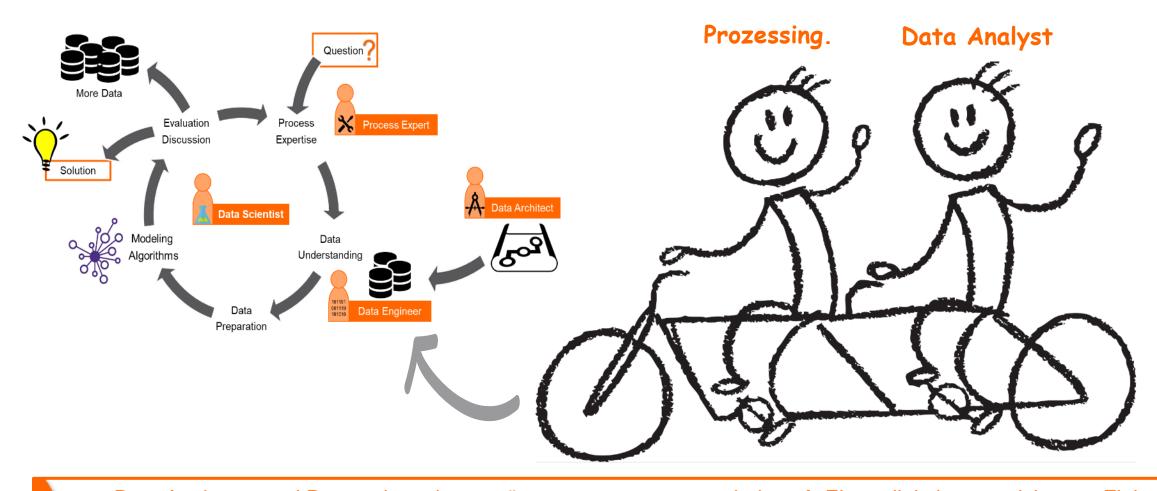
Schnelles und häufiges Durchlaufen des Crisp Cycles für erfolgreiche Anwendung der Data Analytics notwendig



### Erkenntnis #2

#### Zusammenarbeit zwischen Fachexperten und Data-Analysten erforderlich





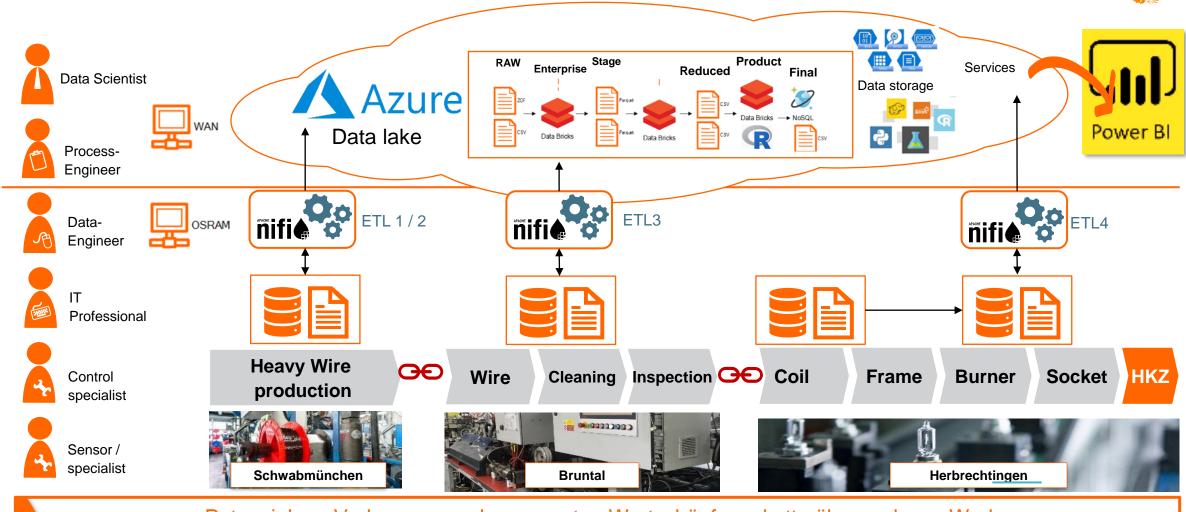
Data Analysten und Prozessingenieure müssen eng zusammenarbeiten → Einer allein kommt nicht zum Ziel



### Holistic Traceability @ AM



**Vom Pulver bis zur fertigen Lampe** 



Potenzial zur Verbesserung der gesamten Wertschöpfungskette über mehrere Werke

