## Big Data in Meteorology

Operational Data Distribution, Processing and Archival at the German Meteorological Service

#### Outline

- 1. The German Meteorological Service
- 2. The "5 Vs"
  - a. Volume
  - b. Velocity
  - c. Variety
  - d. Veracity
  - e. Value
- 3. Outlook

# The German Meteorological Service

Your single source for weather and climate

Government organization

Basic and applied research

Operational support for specialized customers



#### **DWD law**

Provision of meteorological services

Safeguarding of aviation and shipping

Warnings about dangerous weather phenomena

Recording, monitoring, evaluation, and forecasting of the atmosphere and its interactions with other systems

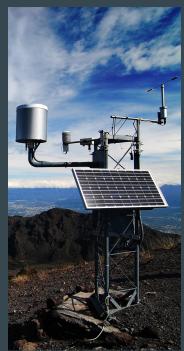
Monitoring for radioactive trace elements and the forecasting of their dispersion



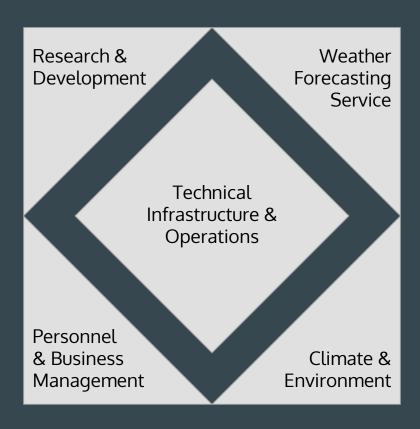








#### **Business areas**



## Volume

Yesterday's weather isn't old news

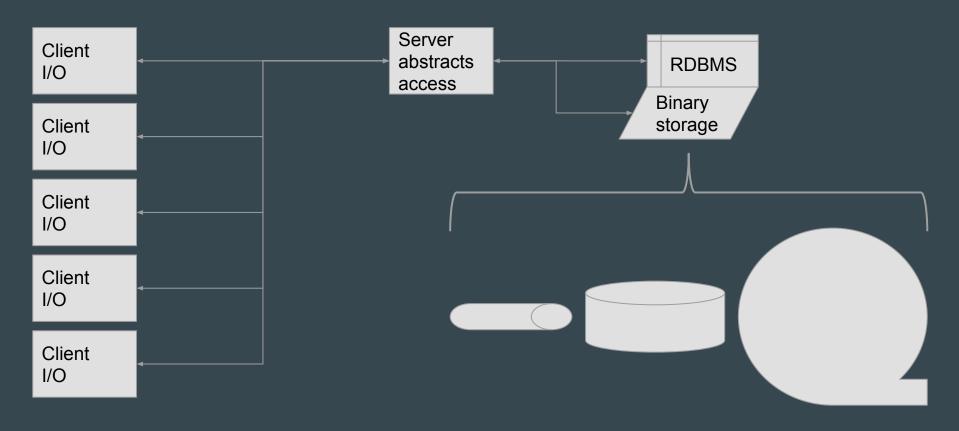
Binary data warehouse approach

Multiple storage media

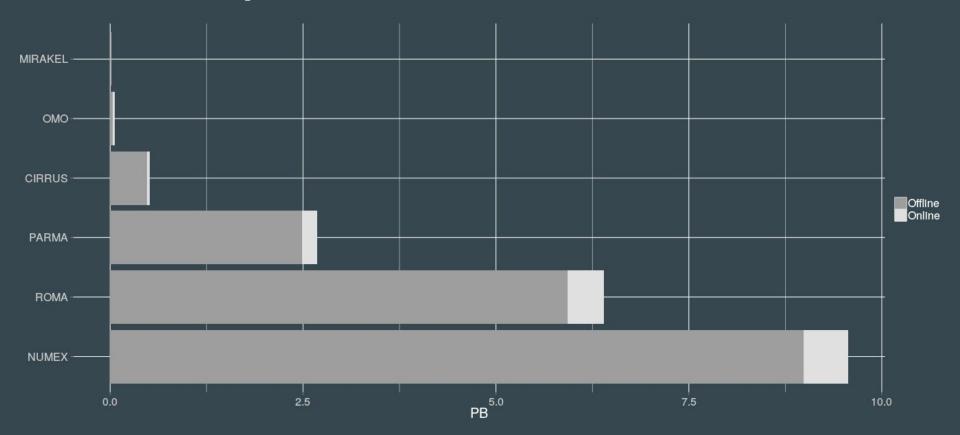
Work is done on local copies



#### Proprietary data access with SKY



### Data volume per database



## Velocity

The future's most interesting now

Users on a live update cycle

- Forecasters
- External users
- Georedundant backup

Streaming formats are used

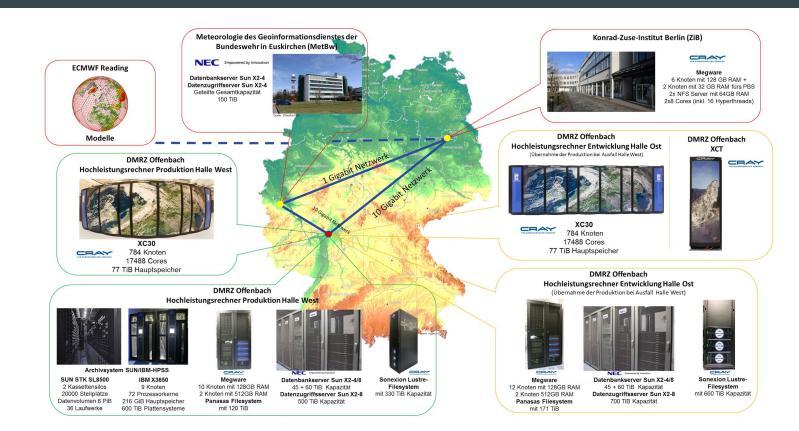


#### Daily model runs

- 3 nested models
- 8 runs each
- 20 member ensembles



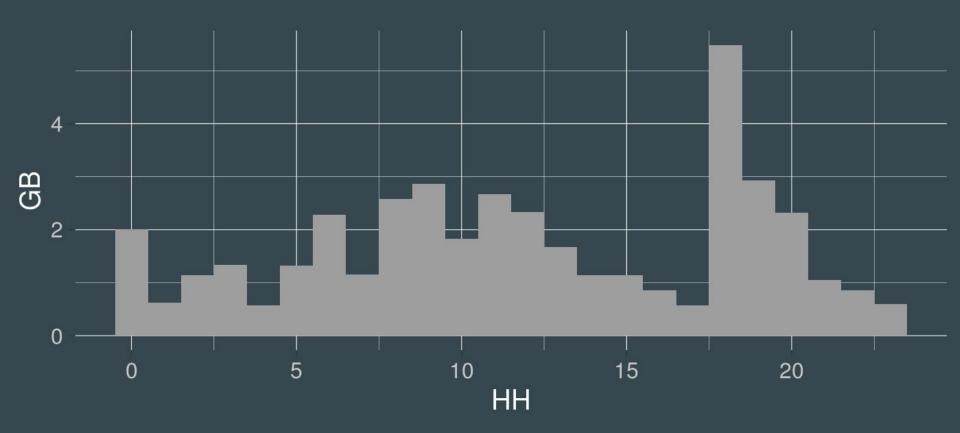
#### **Computing infrastructure**



#### Daily data growth for binary DBs



## Transaction log generation for MIRAKEL



## Veracity

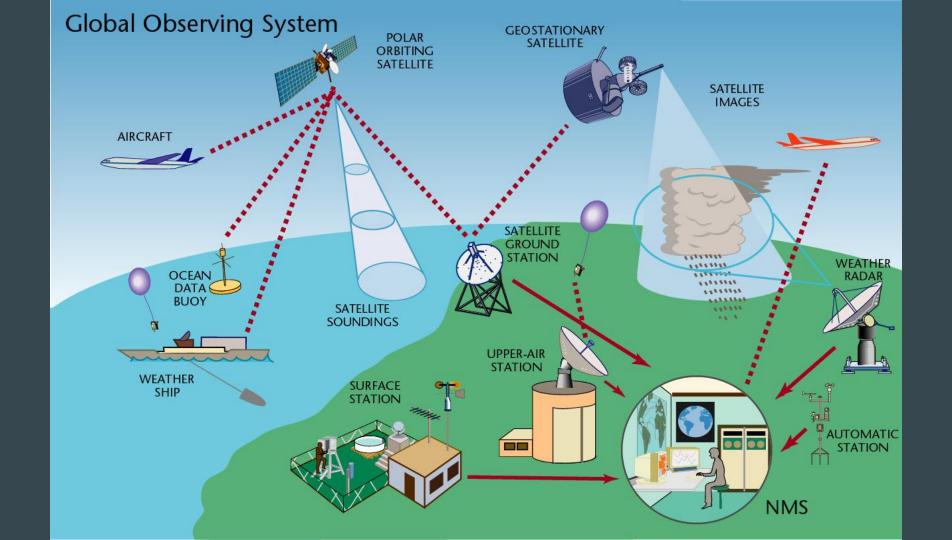
The search for truth

Observation errors

**Encoding errors** 

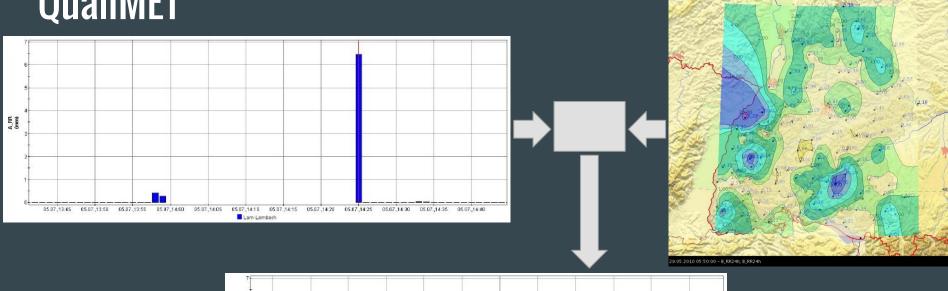
Local vocabularies

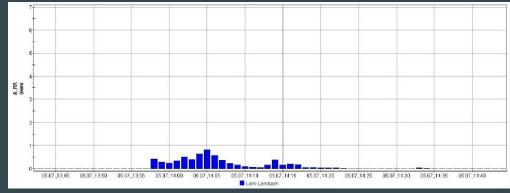




## Something's bound to go wrong

## QualiMET





## **Transmission monitoring**

												-		-																												-
TerminZeitpunkt/ Datenart	<u> </u>	78 079	<u>081</u>	<u>082</u>	087	105	106	107	08 1	09 1	11	1 112	2 113	114	115	116	<u>117</u>	118	119	120	121	122	123 1	124	25 12	6 127	128	129	<u>130</u>	131	132	33 13	4 135	136	137	138 1	39 14	10 14	1 142	143	144 ]	.45 146
Mail-Überwachu	ıng								12054			20/04/2											-	10000	200					-	10054						112	2000			kein	e Mail-
12 UTC 20160223 14:14		•			•		•									•				0		•	0							•						0					0	
<b>09 UTC</b> 20160223	•												•				•		•									0														
<b>86 UTC</b> 20160223					•															•																						
<b>03 UTC</b> 20160223 05:14													•				•	0		<u> </u>								0														
<b>00 UTC</b> 20160223 02:14				0										0		•				•																					0	
G21 UTC 20160222 23:14	•	•											•	•			•			<u> </u>								•														
G18 UTC 20160222 20:14							9						0	0		•		•		•														•							9	
G15 UTC 20160222 17:14		-		•									_	0	•				•										•					•		0						
G12 UTC 20160222 14:14		0												0					•																						0	
G09 UTC 20160222 11:14	•				•				9	) (			•	0	0	0	•		0								0	0		•				0								
G06 UTC 20160222 08:14														0	0			•	0			•						0														
G03 UTC 20160222 05:14					•					) (			0	0	•	•	•	•	•	•	•							0		•				•		9						
G00 UTC 20160222 02:14		0			•	•								•		•		•	0	•	•	•	•				0	0	•					•	•	9			9			

#### **Encoding monitoring**

40 pair(s) from 88 Message(s) checked and 14 comparison(s) with problems

 $FM12 = 40 Message(s) \longleftrightarrow FM94 = 48 Message(s)$ 

SV	IIiii	Datum	UTC	iX	h	VV	N	DDD	FF	TT	TD	RH	QFE	QFF	geopH	a	pp	ww	W1	W2	Nh	CI	Cm	Ch
FM12	40153	22.02.2016	00:00	1	0	0	9	220	12.3	7.1	6.9		894.3			7	-2.3	62	9	6		3		
FM94	40153	22.02.2016	00:00	1	30	50		220	12.4	7.1	6.9	99	894.3		1352	7	-2.3	62	9	6				
FM12	40153	22.02.2016	06:00	1	0	800	9	270	5.1	6.9	6.8		897.0		1377	2	2.2	50	6	5				
FM94	40153	22.02.2016	06:00	1	30	800		270	5.1	6.9	6.8	99	897.0		1377	2	2.2	50	6	5				
FM12	40153	22.02.2016	09:00	1	200	3000	8	320	4.1	7.2	6.8		899.0		1395	2	2.0	60	9	6	8	4	-1	-1
FM94	40153	22.02.2016	09:00	1	250	3000	8	320	4.1	7.2	6.8	97	899.0		1395	2	2.0	60	9	6	8	4		
		22.02.2016		1	0	800	9	330	6.7	7.0	6.6		899.4		1399	2	0.5	60	9	5				, S
FM94	40153	22.02.2016	12:00	1	30	800		330	6.7	7.0	6.6	97	899.4		1399	2	0.5	60	9	5		3		
FM12	40153	22.02.2016	15:00	1	0	600	9	290	4.1	7.3	7.0		901.5		1418	2	2.1	60	6	5				
FM94	40153	22.02.2016	15:00	1	30	600		290	4.1	7.3	7.0	98	901.5		1418	2	2.1	60	6	5				
FM12	40153	22.02.2016	18:00	1	300	4000	7	310	3.6	6.7	6.4		903.8		1438	2	2.3	61	6	5	5	4	6	-1
		22.02.2016		$\overline{}$	450	4000	7	310	3.6	6.7	6.4	98	903.8		1438	2	2.3	61	6	5	5	4	6	
FM12	40153	22.02.2016	21:00	1	2500	10000	7	340	6.7	6.6	5.9		905.1	,	1450	=	1.3	1	6	2	7	0	3	-1
FM94	40153	22.02.2016	21:00	1	2550	10000	7	340	6.7	6.6	5.9	96	905.1		1450	2	1.3	1	6	2	7	0	3	
š <u> </u>		22.02.2016			600	2000	8	260	6.2	13.6	11.6		1000.5	1004.2		7	-2.0	95	9	7	8	3	-1	-1
FM94	40179	22.02.2016	00:00	1	800	2000	8	260	6.2	13.6	11.6	88	1000.5	1004.2		7	-2.0	95	9	7	8	3		
FM12	40179	22.02.2016	09:00	1	600	15000	7	230	2.1	13.1	11.4		1004.4	1008.1		2	1.8	2	6	2	7	8	-1	-1
FM94	40179	22.02.2016	09:00	1	800	15000	7	230	2.1	13.1	11.4	90	1004.4	1008.1		2	1.8	2	6	2	7	8		

## **Variety**

#iwokeuplikethis

Heterogeneous data received

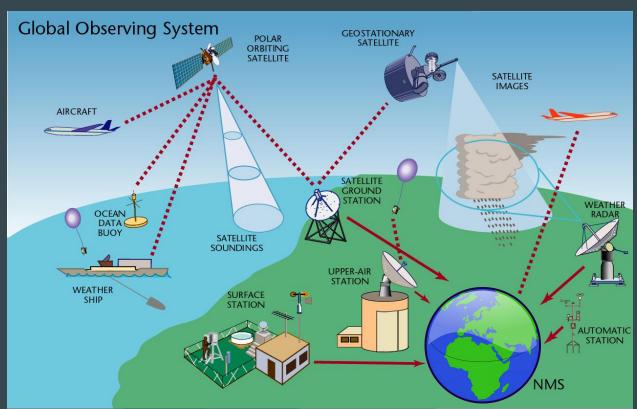
Formatting more difficult with new partners

Diverse user needs



#### **Standardized ingress**

Now: 1 format / data source, homogenization on arrival

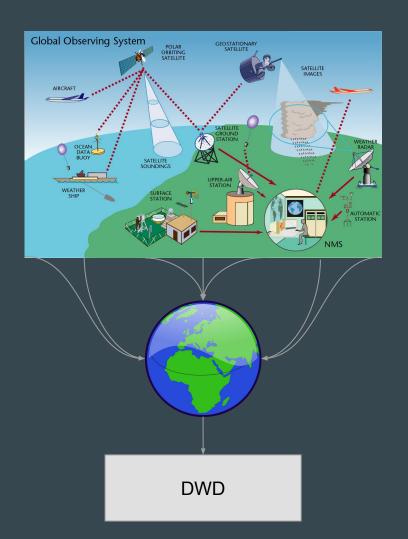


#### **Standardized ingress**

Now: 1 format / data source, homogenization on arrival

Future: Homogenization before

transmission

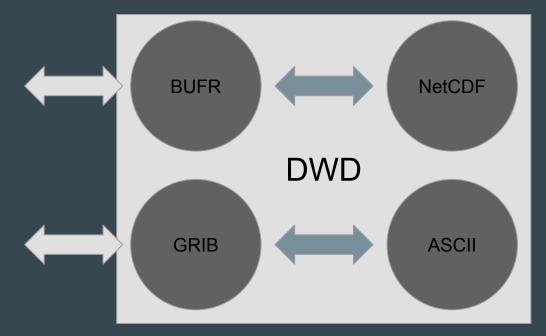


## Standardized internal flow

Now: internal conversion

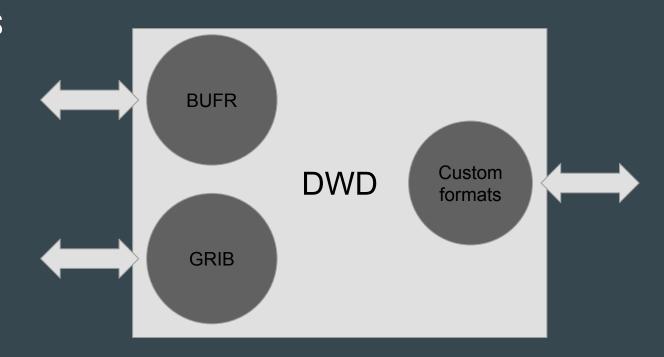
Future: Perhaps less I/O, better

streaming formats?



#### **Standardized egress**

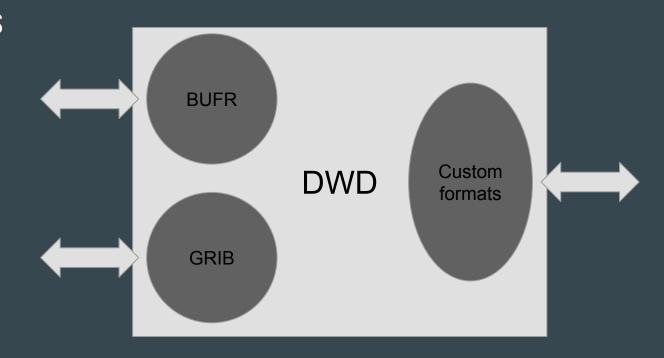
Current: standardized outputs + some specialized formats (fax, phone, etc.)



#### **Standardized egress**

Current: standardized outputs + some specialized formats (fax, phone, etc.)

Trend: more users, more formats

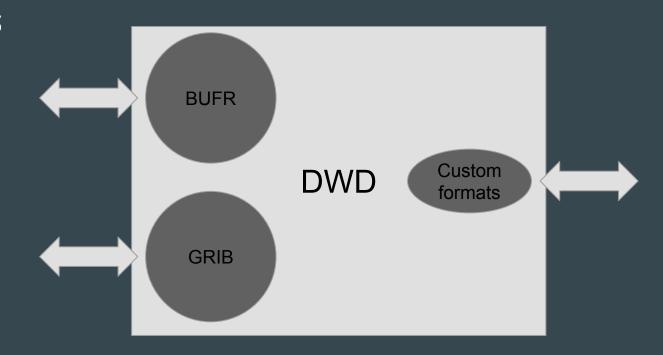


#### **Standardized egress**

Current: standardized outputs + some specialized formats (fax, phone, etc.)

Trend: more users, more formats

Future: Perhaps more standardized output formats (web services, etc.?)



## Value

All data are equal...
But some are more equal than others

Raw observations

Time series

Model outputs

...Integrity checks



Ship logs

En route from Hamburg to Hong Kong

May 17, 1866

	*	90 0						4		-	-	h.J			-9	The state of the s
	0	y am I	Sard de	ban	burge	0 13	Tring The	19.41	ine	ich aus	(der s	Paise	non			Hambarg mach Hongkong.
(X)	NI ON	North Letitude	West	Carry	VY B	Osom.	the Thirmes	n fish	Yay?	Report	Bours a	Moto				Romestunaini
V II	4	49: 110	11: 54.	12:18	31			100					1, 9	Rote 4.		Samskungen Sin Bland Soft fig out THE De ber Juit his blandlight
		19: 13			31	2102		di	110	-10	0		29.9	1-3.		The Control being Still full alle in fligh.
6		11: 51.			31	25 6		,								Me for the first the baration Soft formethe 22: 25 m
	8	18:42	6 11		31	2307	20			10	,		80	3.		but to and I was an lingual Suffer forwish inder
1	12	,			31	11 3	7			10				1	-	· In and ridach and the - long hashauff
11	1	18: 27	7 35		20	11 5			,	10	0	23:00	090	,		Min & Alle bis bloom Till his wifty. My fores their of frances land of the all him Masferlingers. Diting asing afore, and ISO frances
		11 1		1 200		90				10	0		090	3.		M. Milling and Solorans house & 45: W Dist 130 Miles.
	3					97.18			,	10	0					Man Mand allifax. Conta Jour.
8	1	17 9	9 8	,	2.	990	7		12				90	8.		And have a firmed to Mind and Jake, Suffer de me
V	12				2	910	4	93.	1	5	0		1	-	-	Pagegrand grands, Lings in to planty gog fix to garning
19	1	46:50	10: 41	4		9:90 0	1	. Eu.	m 4-	0	v		80,3	5.		1990 And South State of the Conference of the State of th
	12	16:26	11:0	1 19:1	1.	9.150	4		,	0			9	. 0.		Sie Tour Lotten Dest 160 Miles
	3	16: 1	/3:			2.80		9				1				Majora havinge bring Jeff his faith to hardend fife thingen at forth the 1-36 fairne the good father bygging Thomas of and fill and the stand the same of fails of the same of
	12	,			1	1:20		-	1	4	26.		9	or or		dun fogan & Mest
V 20.	1			29:75	60			6	· v w	-	16.	1	8 or			My jest my free he deft sife st ym Ister 2 2.2 hay and the free hand for the stand has been been been been been been been bee
		15: 35												-		a - Walnut House 1840 184 184 Miles Name 12: or Martily
	12	15: 12	15 10					En	in or	5.			9	6		1 50 9m and his world week from and week for for going of from the last on the form
	8	11: 52						29,	X.	11:	21.		9	5		11 Pin fort some Galain Brouffinger Salmarti.
	10			11:75	60											

Historical colonial logs

Chongqing

Jan 1, 1891

,	01	ung	King	7	Rain	)	fe	bua	~ .			d	con			
	94	99	Man	clein	Regen	96	95	Um	lech	Ben.	4	9 4	lear		Ben.	
1	46	46	49	45	7	49	52	50	41	of a.	9 58	59	63	seen 56	Jen.	
2	47	52	5'7	45	-	48	45	47	44		59	56	62	56	-	
3	54	47	58	44	-	43	44	45	43	-	53	53	56	52		
4	43	49	60	43	-	44	45	47	. 43	-	53	52	60	52		
5	50	50	56	49	_	44	42	44	42	-	54	53	58	52	_	
6	50	46	56	46	_	42	43	45	40	-	52	54	60	52	-	
2	48	50	58	47	-	43	43	45	41	-	53	55	62	52	-	
8	50	50	55	50	9	44	43	45	41	-	57	52	60	51	_	
9	50	49	54	48	Ryan Byans	+ 43	39	44	37	-	53	54	57	51	•	
10	47	47	54	47	0 -	43	41	43	36	-	53	51	34	50	•	
11	54	51	58	52	-	40	38	44	36		51	51	54	49	-	
12	50	50	54	50	-	34	42	47	34		54	52	65	49	-	
13	50	48	54	48	_	55	45	55	43	-	50	56	60.	47	-	
14	47	49	58	47		52	49	55	38	-	56	54	57	47	•	
15	48	48	54	48	-	49	49	54	45	-	53	55	60	.50	-	
16	50	50	58	50		47	47	50	45	-	57	55	65	51	-	
17	47	52	58	48	-	47	48	54	45	-	54	56	59	50	-	
18	51	51	54	51		47	47	52	46	-	52	55	60	48	•	
19	48	47	52	47	-	54	50	55	48	-	55	54	55	53		
20	46	45		45	_	50	57	56	47	-	52	50	52	50		
21	45	45	54	45	-	53	53	59	49	-	50	49	50	48		
22	46	47	59	46	. 7	54	58	64	51	-	43	49	50	47		
23	52	52	64	51		60	59	68	48	-	49	48	49	48		
24	51	49	56	49		57	58	61	50	-	48	51	52	47	-	
25	45	48	54	45	-	59	60	66	56	7	52	54	58	52	Pay Tyan	
26	49	49	52	48	-	63	59	74	56	-	54	50	65		-	
27	48	48	58	48		63	62	70	58	-	56	60	70	52	-	
28	46	50	56	45		67	62	69	52	-	63	66	70	56	-	
29	50	49	51	45	-						60	60	66	59		
30	46	44	51	43							58	64		54	-	
31	42	3 44	46	40							60	67	76	54	-	
Mattel	48.4	48.5	55.2	46.9		49.9	49-1	53.9	44.8		54.0	54.9	59.7			

## Outlook

More is more

#### Increasing demands

- Data sources
- Forecast horizons
- Data volume
- Tailored user solutions



#### Risks and opportunities

Increasing data volume Better HSM

User demand for direct access Separation of concerns: ETL, REST

More data sources More standardization

Faster updates Increased scale-out capabilities

Daniel Lee

Deutscher Wetterdienst (DWD)

Data ingress and decoding

Daniel.Lee@dwd.de

Tel: +49 (69) 8062-2706

Fax: +49 (69) 8062-3829

http://www.dwd.de