



Konzeption der Datenbankschnittstelle

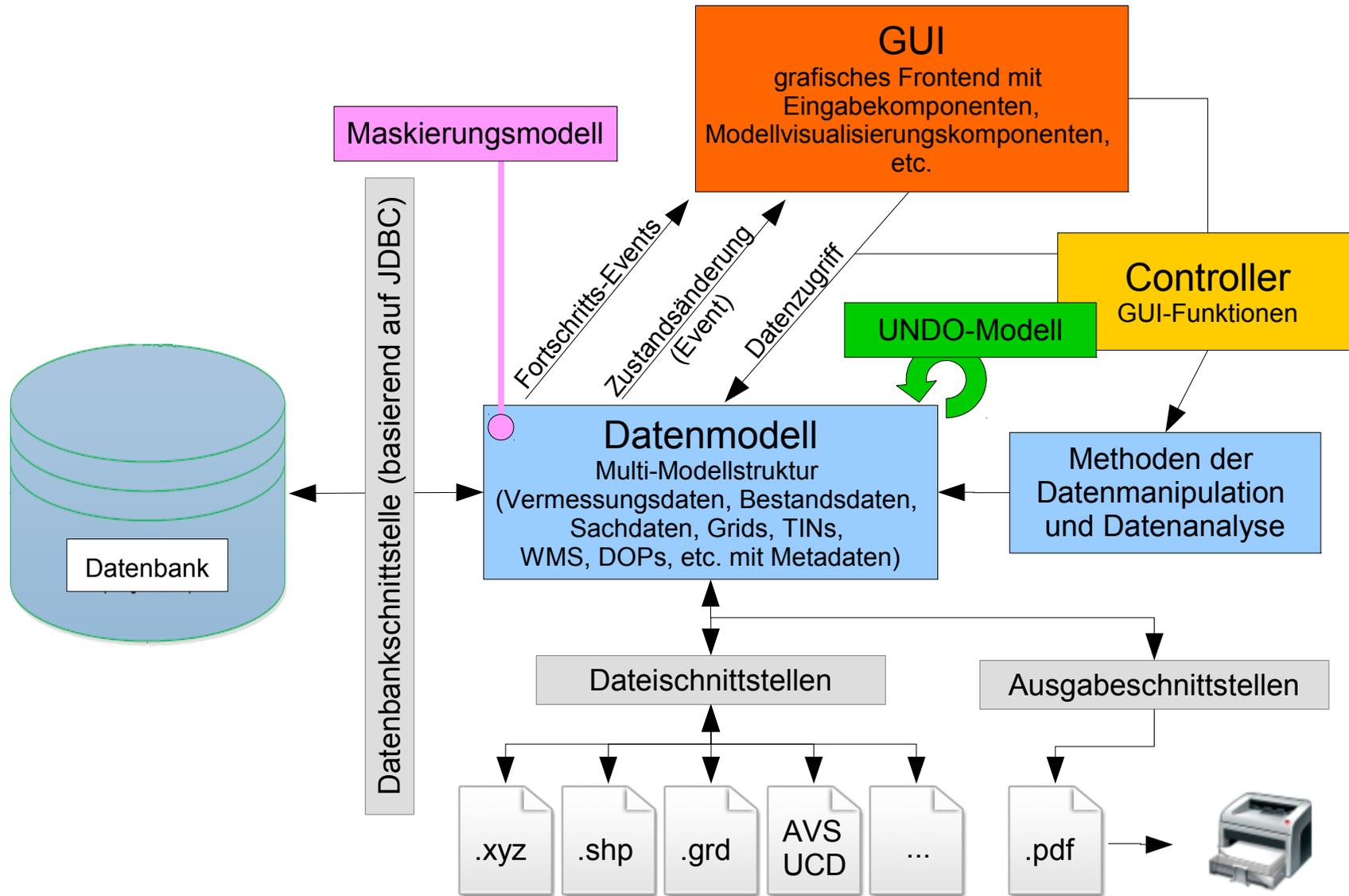
Gismo Workshop am 2.3.2017 im BSH, Hamburg

Dipl.-Ing. Christoph Lippert
smile consult GmbH

Inhalt

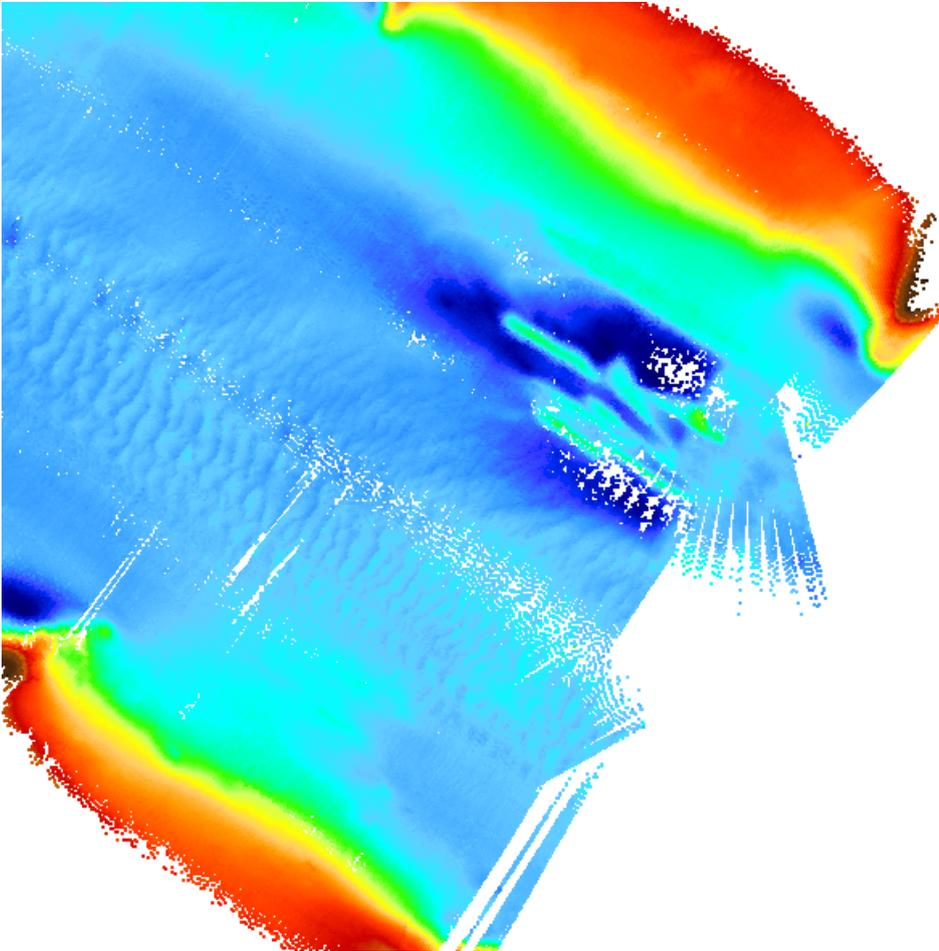
- Grundlegendes Softwaredesign
 - Datenmodell der Klassenbibliothek
- Konzeption und Realisierung der Datenbankschnittstelle
 - Dynamisches Tabellenkonzept
 - Konzeption der räumlichen Suche
 - Rechte- und Rollenkonzept
- Besonderheiten der Umsetzungen für die DBMS MySQL, PostgreSQL, Oracle

Grundlegendes Softwaredesign



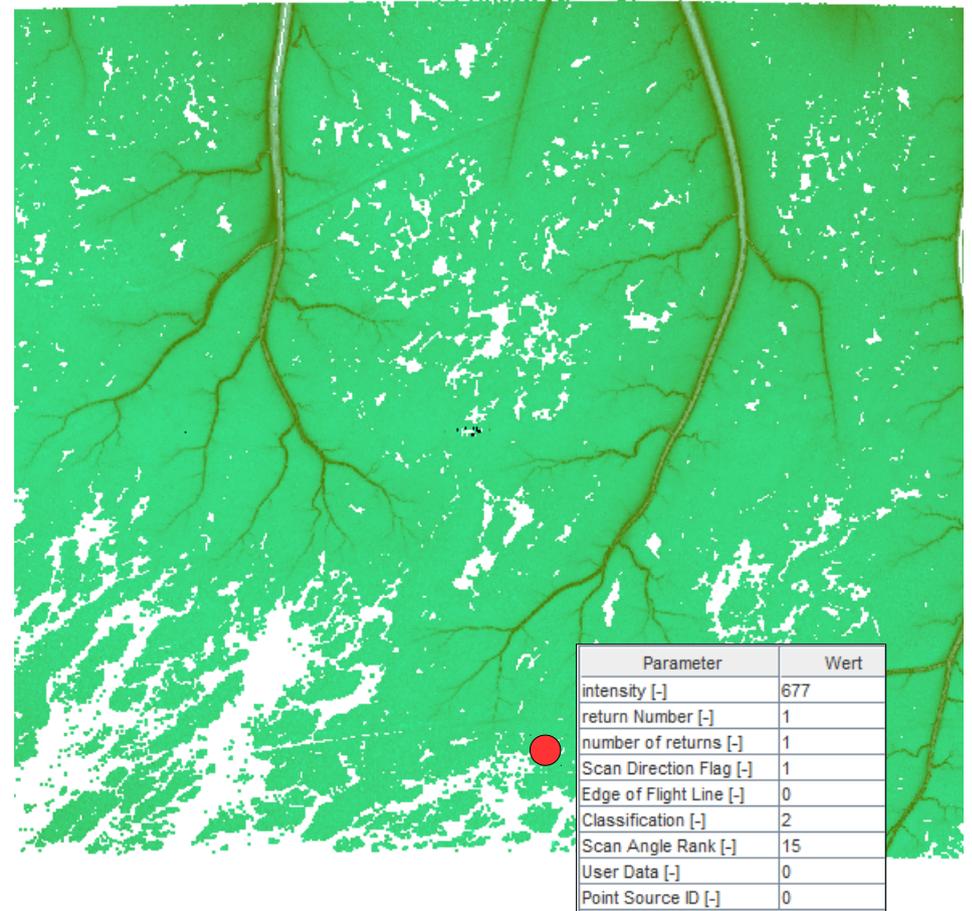
Datenmodell

unstrukturierte Daten



Fächerecholot-Daten

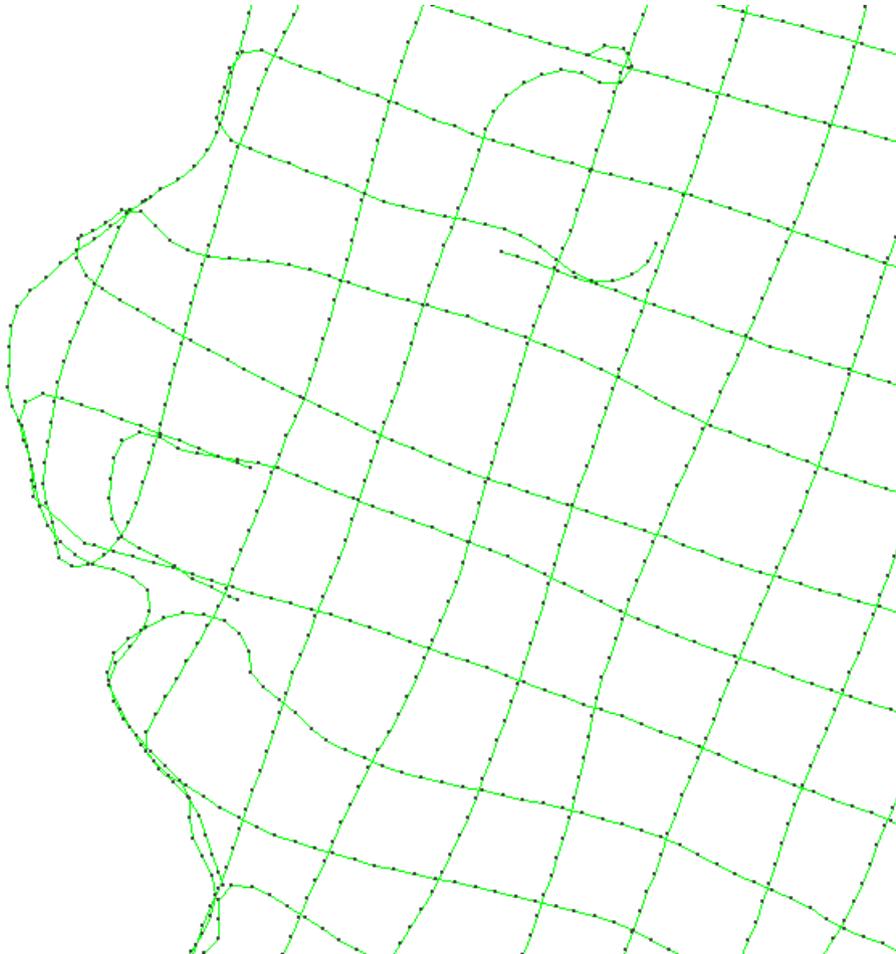
unstrukturierte Daten



Lidar-Daten mit Attributen

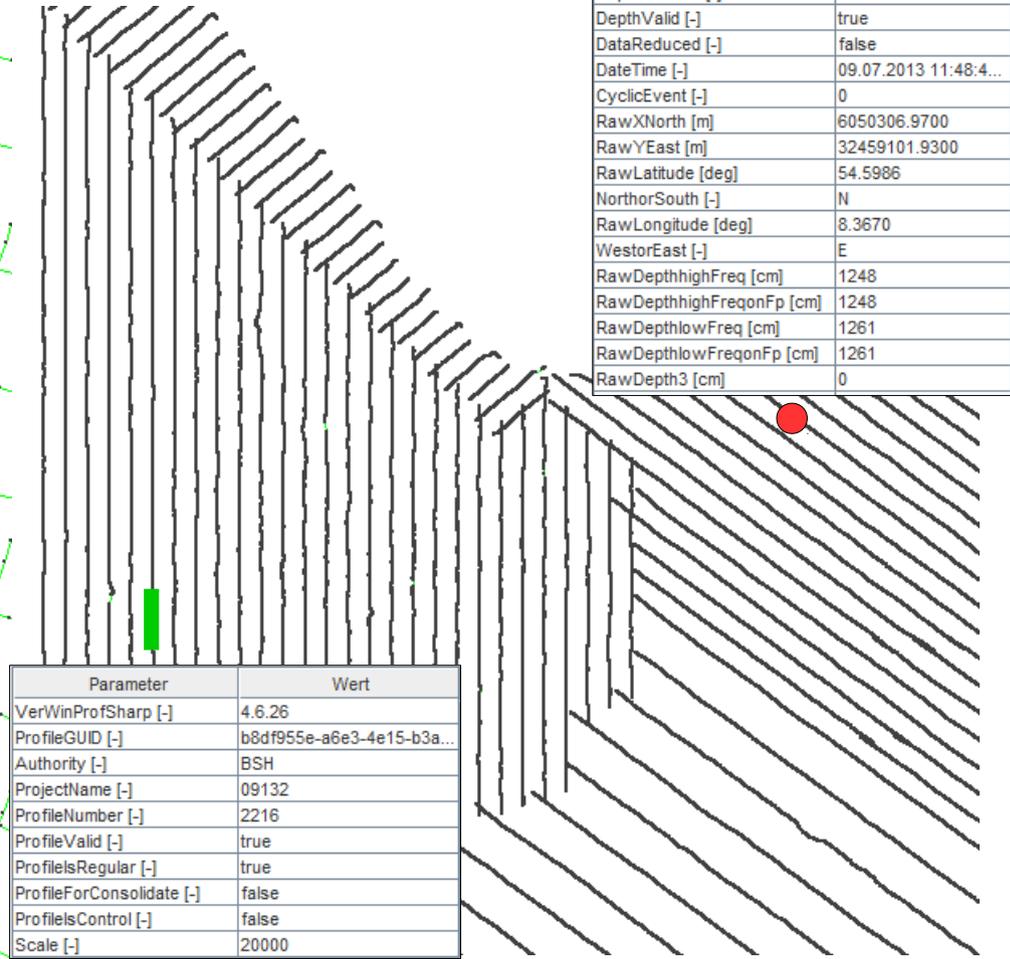
Datenmodell

unstrukturierte Daten



Kreuzprofile

unstrukturierte Daten

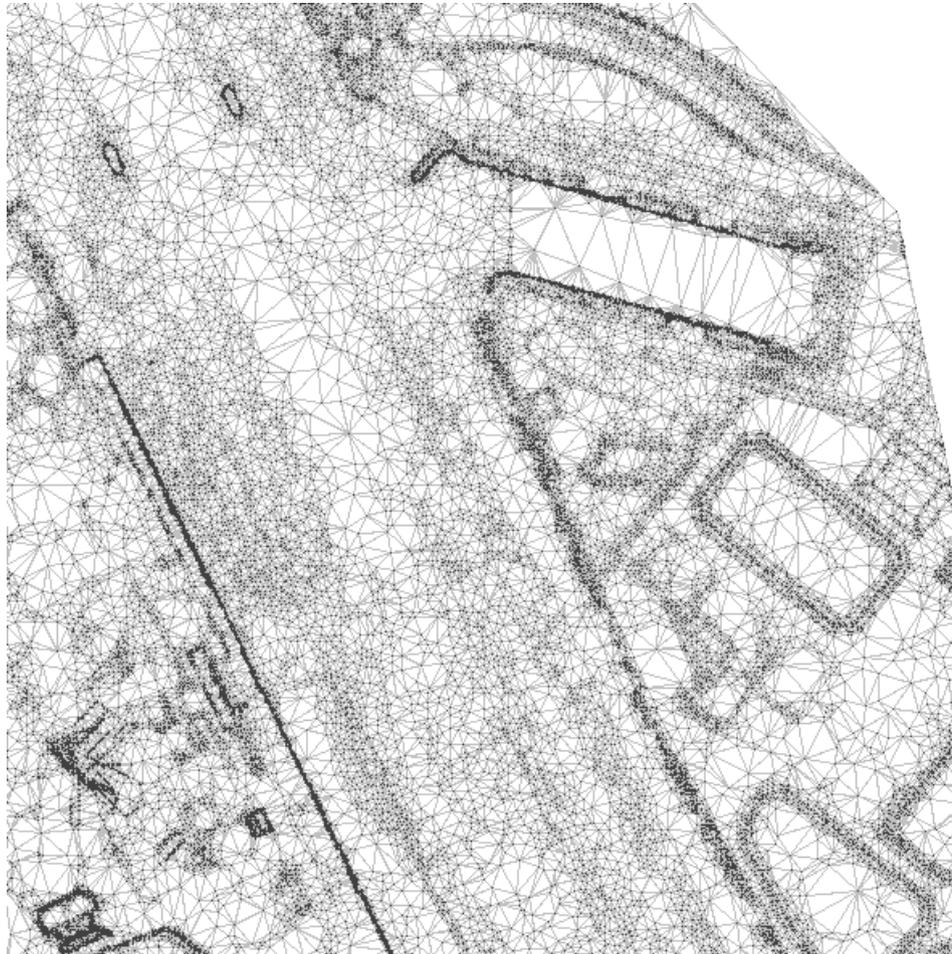


Linienpeilungen mit Attributen an den Vermessungspunkten und den Peillinien

Parameter	Wert
DepthNumber [-]	1733
DepthValid [-]	true
DataReduced [-]	false
DateTime [-]	09.07.2013 11:48:4...
CyclicEvent [-]	0
RawXNorth [m]	6050306.9700
RawYEast [m]	32459101.9300
RawLatitude [deg]	54.5986
NorthorSouth [-]	N
RawLongitude [deg]	8.3670
WestorEast [-]	E
RawDepthhighFreq [cm]	1248
RawDepthhighFreqonFp [cm]	1248
RawDepthlowFreq [cm]	1261
RawDepthlowFreqonFp [cm]	1261
RawDepth3 [cm]	0

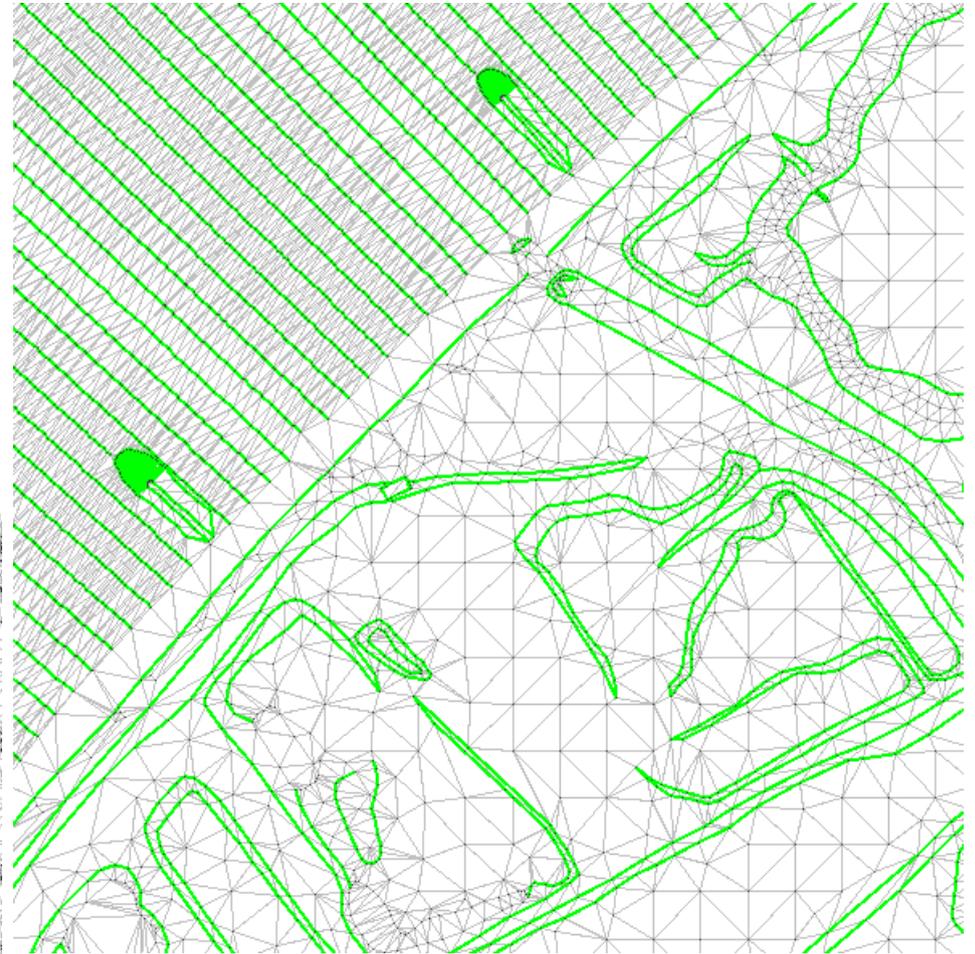
Datenmodell

unstrukturierte Daten



TIN

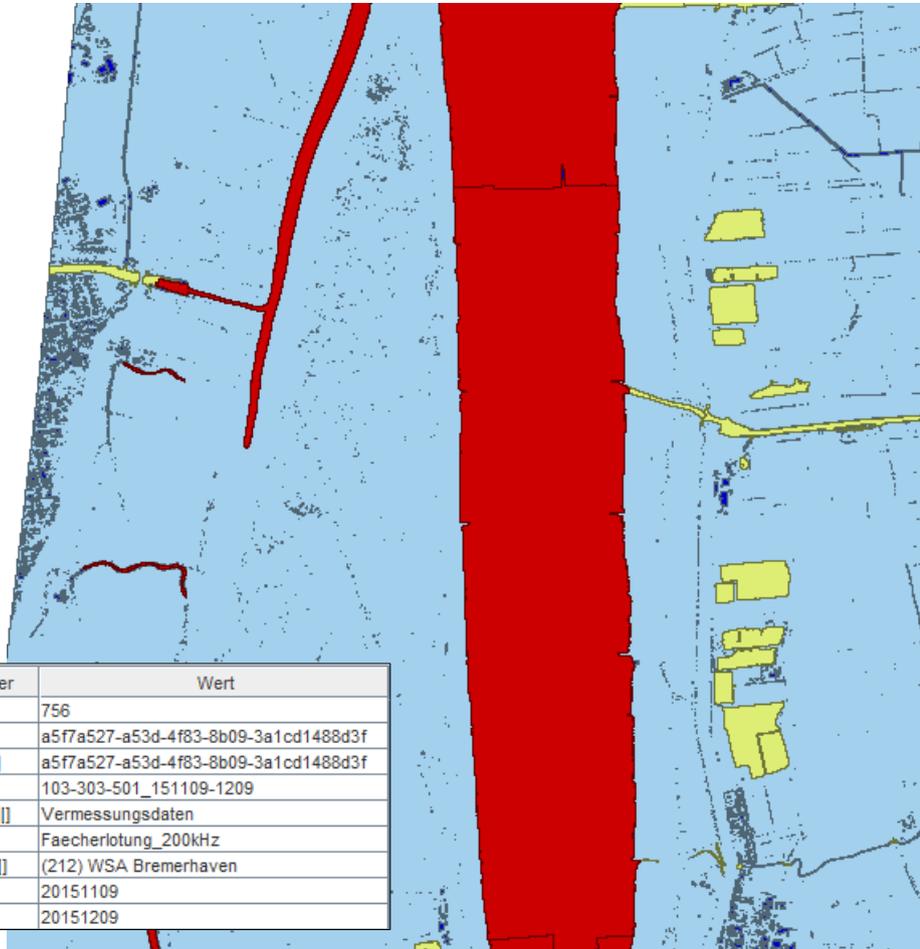
unstrukturierte Daten



TIN mit Strukturkanten

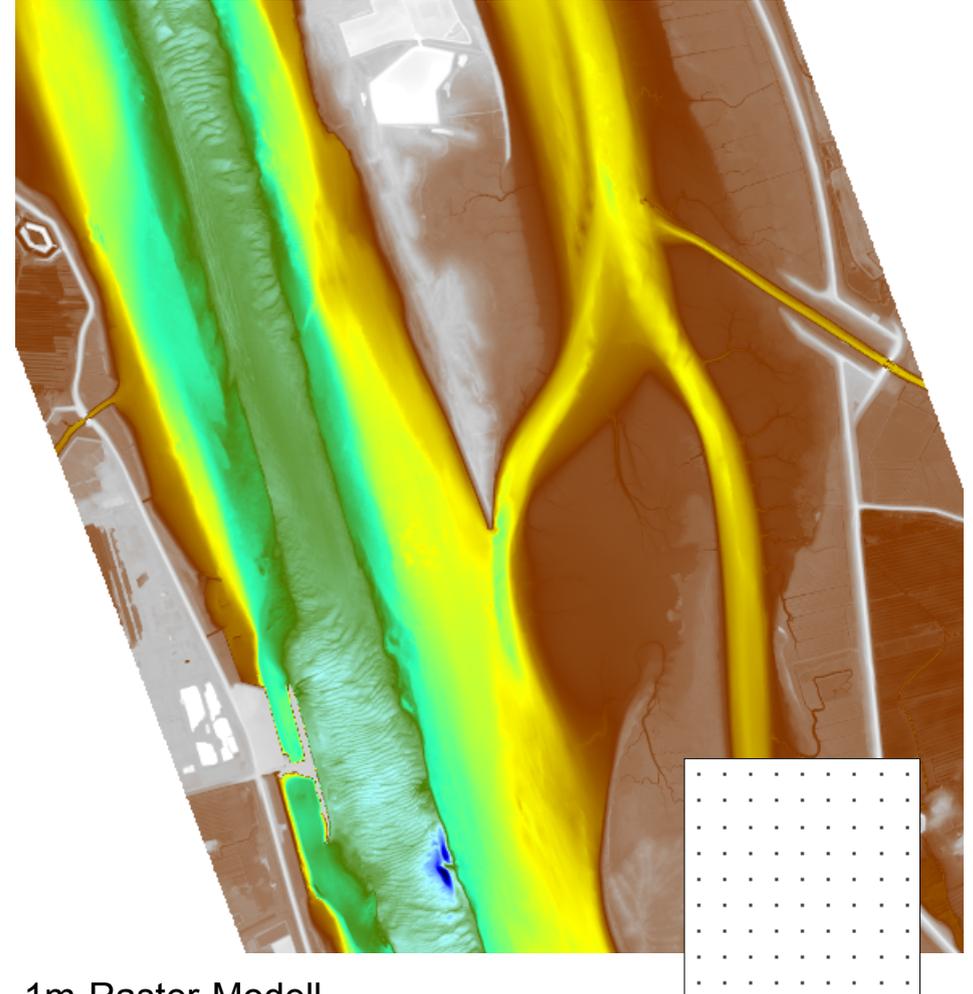
Datenmodell

unstrukturierte Daten



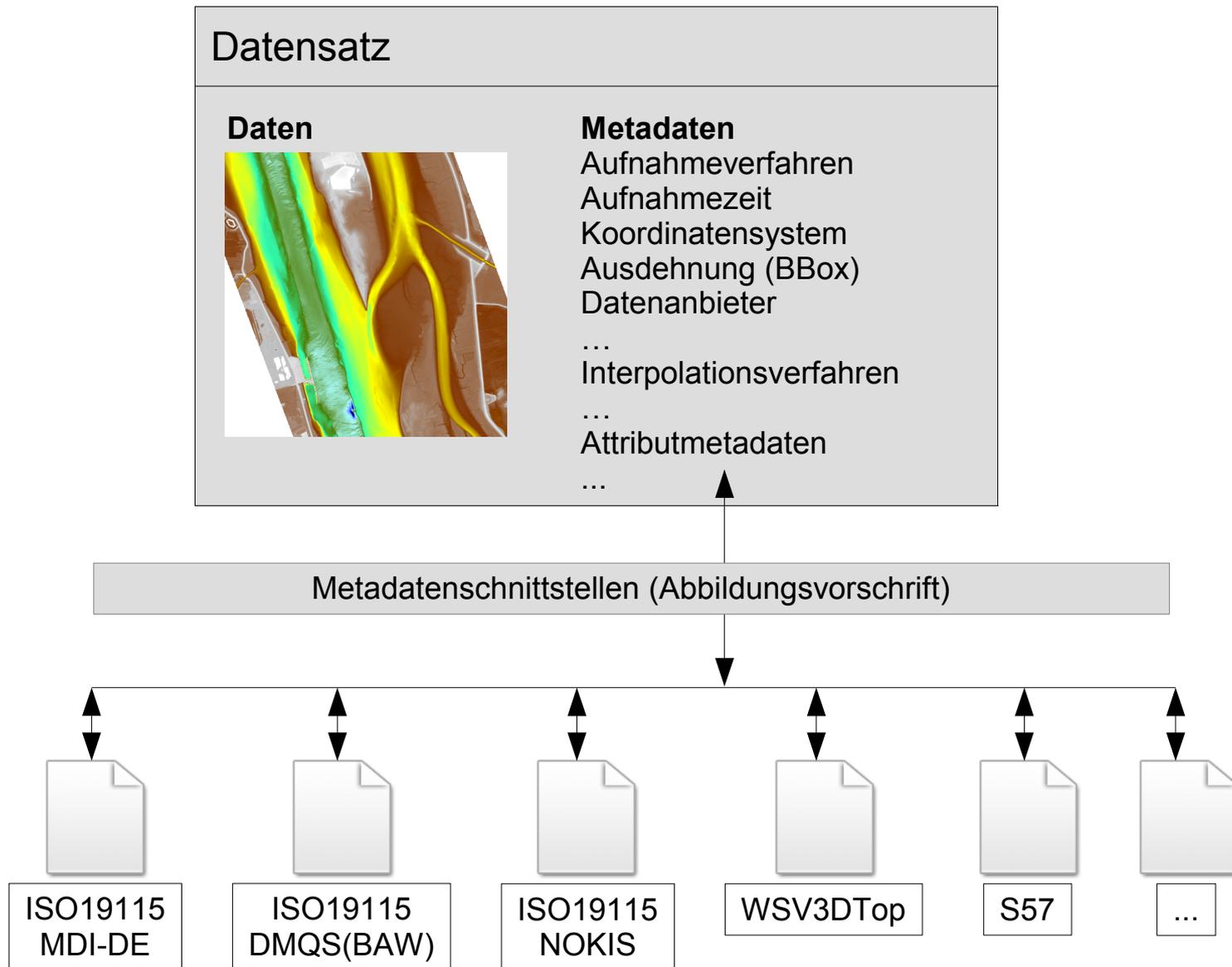
Datenquellenkarte (Polygonflächen mit Attributen)

strukturierte Daten



1m-Raster-Modell

Datenmodell



Designanforderungen an das Datenbankmodell

- das Datenbankmodell soll die Datenarten und Modelltypen des Datenmodells ohne Informationsverlust abbilden :
 - Speicherung strukturierter und unstrukturierter Daten
 - Speicherung unstrukturierter Daten als beliebige Kombination der Geometrietypen Punkt, Polygon und Element
 - beliebige Attribute (einfache und komplexe Datentypen) an den Geometrietypen
- das Datenbankmodell soll die flexible Metadatenstruktur des Datenmodells abbilden :
 - Übernahme von Metadaten aus unterschiedlichen Metadatenstandards bzw. Metadatenformaten
 - Definition von Prozeßmetadaten zur Steuerung von datenbankgestützten Methoden (z.B. Interpolationsverfahren)
 - individuelle Definition von Metadaten ermöglichen

Designanforderungen an das Datenbankmodell

- die Datenbankschnittstelle soll den selektiven Zugriff
 - auf einzelne Datensätze des Datenbestandes
 - auf Teildatenbereiche innerhalb der Datensätzeunterstützen
- ein selektiver Datenzugriff erlaubt intelligente Zugriffs- bzw. Nachladestrategien
- der selektive Datenzugriff ist Grundlage für die Umsetzung datenbankgestützter Methoden, z.B.
 - Generierung von Digitalen Geländemodellen durch Interpolation der Modellstützstellen auf einem Bestand an Vermessungsdatensätzen
 - Generierung bathymetrischer Zeitreihen
 - Generierung von Profilschnitten
 - ...

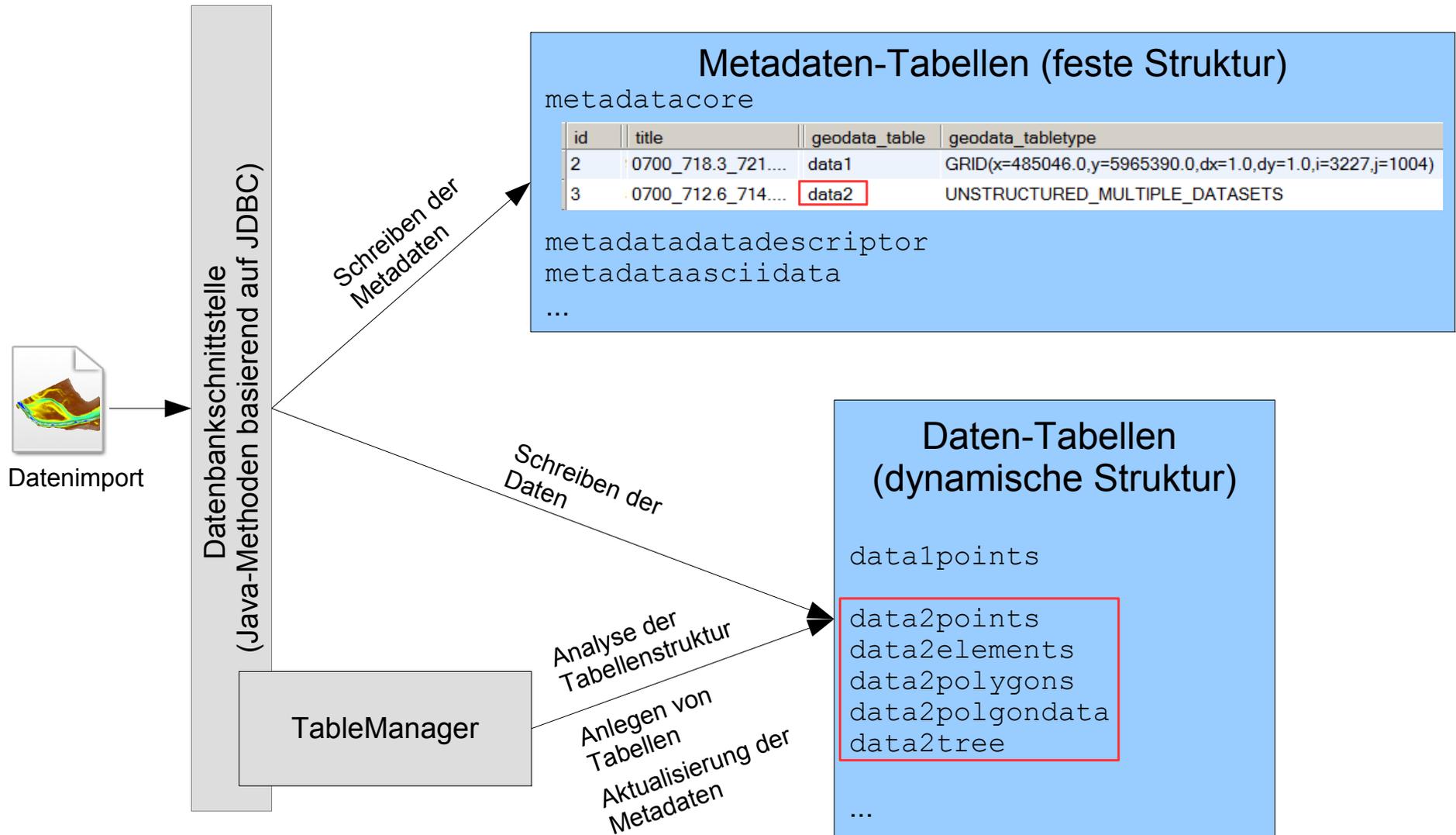
Designanforderungen an das Datenbankmodell

- Unterstützung gängiger Datenbanksysteme (PostgreSQL, Oracle, MySQL, etc.)
 - Nutzung bestehender DBMS
- produkt-unabhängige Umsetzung :
 - Verwendung von JDBC (Java Database Connectivity), integriert in der API des Java Development Kit (ab 1.1)
 - JDBC stellt eine standardisierte Schnittstelle für relationale Datenbanken dar
 - Datenbankhersteller stellen JDBC-konforme Treiber für ihre Systeme zur Verfügung

Voruntersuchung zu räumlichen Datentypen

- alle gängigen Datenbanksysteme unterstützen räumliche Datentypen
 - über räumliche Datentypen können geometrische Operationen als Datenbankabfragen formuliert werden
- verfügbare Geometrietypen orientieren sich i.d.R. eng an der SimpleFeature Spezifikation der OGC
 - die spezielle Struktur der unstrukturierten Datenarten des Datenmodells kann nicht geeignet abgebildet werden
- Durchführung eines Anwendungsbeispiels für ein TIN durch Speicherung der Dreieckselemente als Geometrietyp `polygon` (MySQL)
 - Datenbankgröße für 3,6 Mio. Punkte, 7,1 Dreiecke: 2 GB (Faktor DB/JBF = 9,7)
 - Räumliche Suche, 200000 Punkt-In-Polygon-Tests: 7,4 msec/Suche
 - (Approx. Interpolationszeit in DGM-W-Projekten, z.B. für 330 Kacheln mit 1 Mio. Stützstellen: ca. 28 Tage)

Dynamisches Tabellenkonzept



Tabellenstruktur für einen unstrukturierten Datensatz (triangulierte Linienpeilungen mit Attributen an den Vermessungspunkten und den Peillinien, aus SDB)

Tabelle data2points (Basis-Tabellenname + points)

id	md_id	nr	x	y	z	status	spi	depthnumber	depthva	datared	datetime	cyclice	rawxnorth	rawyeast	rawlatitude
1	1	0	529685.99	6075004.97	1.705	1	0	161	1	0	2012-09-19 07...	0	6075004.97	529685.99	54.82130881
2	1	1	529855.03	6075145.61	1.205	1	0	290	1	0	2012-09-19 06...	0	6075145.61	529855.03	54.82256262
3	1	2	529612.79	6074869.85	1.405	1	0	252	1	0	2012-09-19 07...	0	6074869.85	529612.79	54.8200989

Tabelle data2elements (Basis-Tabellenname + elements)

id	md_id	nr	n0	n1	n2	n3	status	spi
1	3	0	4206	3992	3983	3983	128	110
2	3	1	3934	4206	3935	3935	128	110
3	3	2	3886	4230	3920	3920	128	110

Tabelle data2polygons (Basis-Tabellenname + polygons)

id	md_id	nr	poly_edge_nr	p0	p1	spi
1	3	0	0	427	963	8
2	3	0	1	963	1463	10
3	3	0	2	1463	2016	33

Punktnummern !

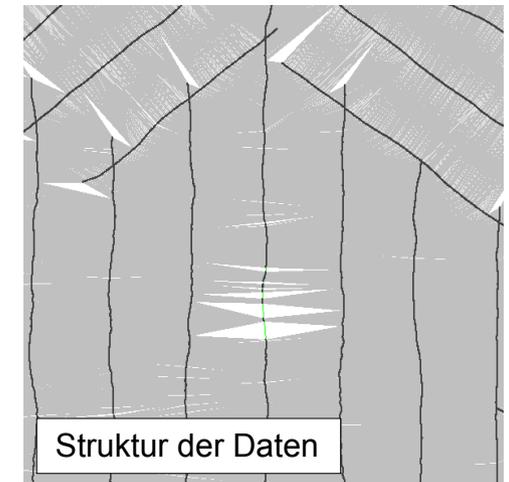


Tabelle data2polygondata (Basis-Tabellenname + polygondata)

id	poly_id	md_id	verwinprofsha	profileguid	authority	projectname	profilenumber	profilevalid	profileisregula	profileforconsc	profileiscontrol	scale
1	0	1	3.11.13	93b21fa9-4...	BSH	0312	1	1	0	1	0	20000
2	1	1	3.11.13	e5b6988e-b...	BSH	0312	2	1	0	1	0	20000
3	2	1	3.11.13	16fcdd52-c0...	BSH	0312	3	1	0	1	0	20000

Tabelle data2tree (Basis-Tabellenname + tree)

id	md_id	spi	depth	minx	miny	maxx	maxy	pointcount	edgecount	polygonedgecount
2	3	0	6	529694.5585540001	6075010.048557	529723.481446	6075038.911443001	356	0	0
3	3	2	7	529690.578996	6075027.6091665	529710.6610040001	6075044.2808335	96	0	0
4	3	3	7	529706.188919	6075037.549109999	529727.8110810001	6075055.3508899985	51	0	0

Tabellenstruktur für einen strukturierten Datensatz (1m-Raster-Modell)

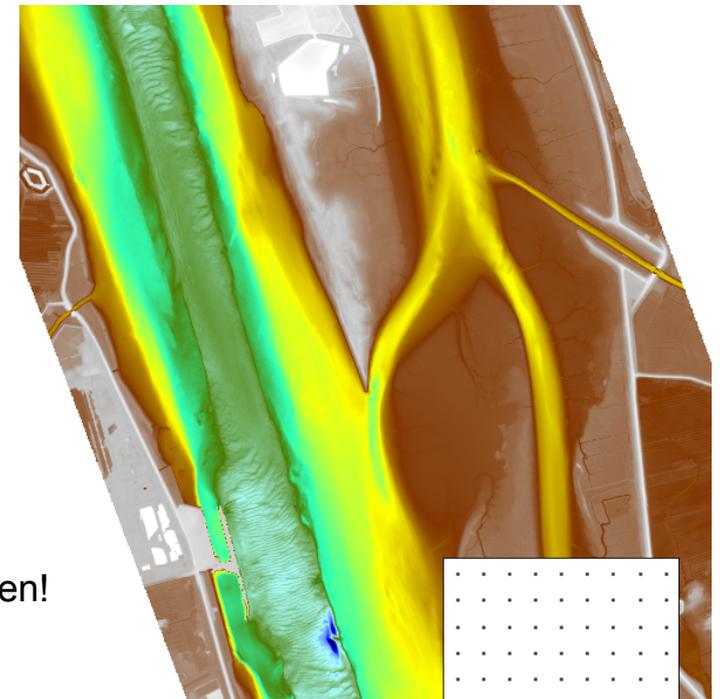
Tabelle `datapoints` (Basis-Tabellenname + `points`)

nr	z	status
0	-0.76	1
1	-0.75	1
2	-0.76	1

Strukturinformation des Rasters
(Referenzkoordinate, Auflösung, Anzahl Zeilen und Spalten)
steht in den Metadaten:

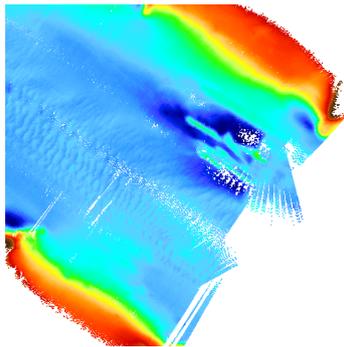
`GRID(x=470000.5,y=5929000.5,dx=1.0,dy=1.0,i=1000,j=1000)`

Strukturierte Daten sollten wenn möglich nie unstrukturiert gespeichert werden!

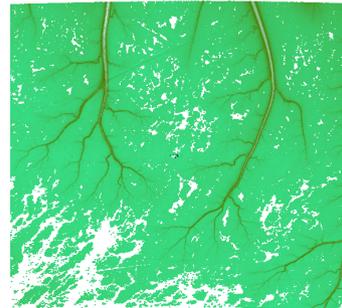


Struktur der Daten

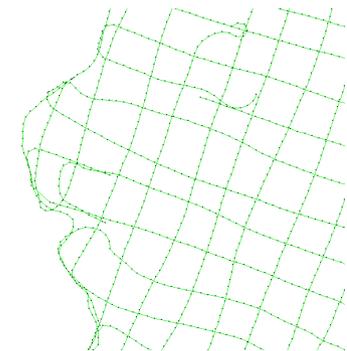
Tabellenstrukturen für unterschiedliche Datenarten / Modelltypen



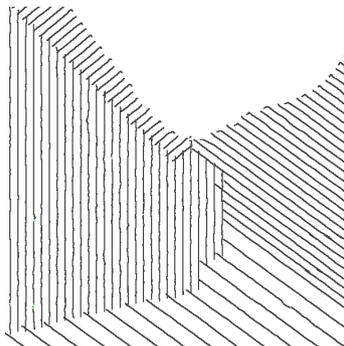
Fächerecholot-Daten
data?points
data?tree



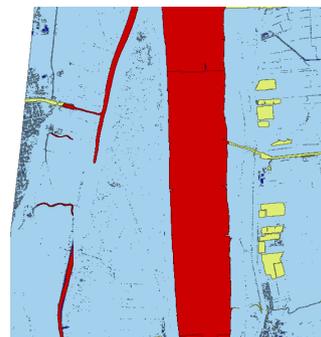
Lidar-Daten mit Attributen
data?points
data?tree



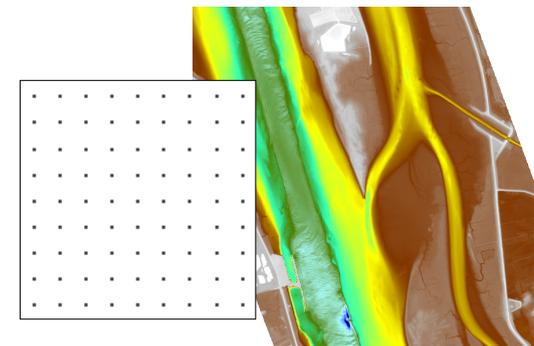
Kreuzprofile
data?points
data?polygons
data?polygondata
data?tree



Linienpeilungen mit Attributen an den Vermessungspunkten und den Peillinien
data?points
data?polygons
data?polygondata
data?tree



Datenquellenkarte (Polygonflächen mit Attributen)
data?points
data?polygons
data?polygondata
data?tree



1m-Raster-Modell
data?points

Suchbaum für unstrukturierte Daten

Zielgröße: ca. 200 Geometrien pro Blatt

Geometrien innerhalb eines Blattes bilden kleinste Zugriffseinheit beim Datenbankzugriff auf die Daten

Binärpfad zum Blatt des Baumes

Bounding-Boxen der Blätter des R-Trees

id	md_id	spi	depth	minx	miny	maxx	maxy	pointcount	edgecount	polygonedgecount
2	3	0	6	529694.5585540001	6075010.048557	529723.481446	6075038.911443001	356	0	0
3	3	2	7	529690.578996	6075027.6091665	529710.6610040001	6075044.2808335	96	0	0
4	3	3	7	529706.188919	6075037.549109999	529727.8110810001	6075055.3508899985	51	0	0

Tabelle data2tree (Basis-Tabellenname + tree)

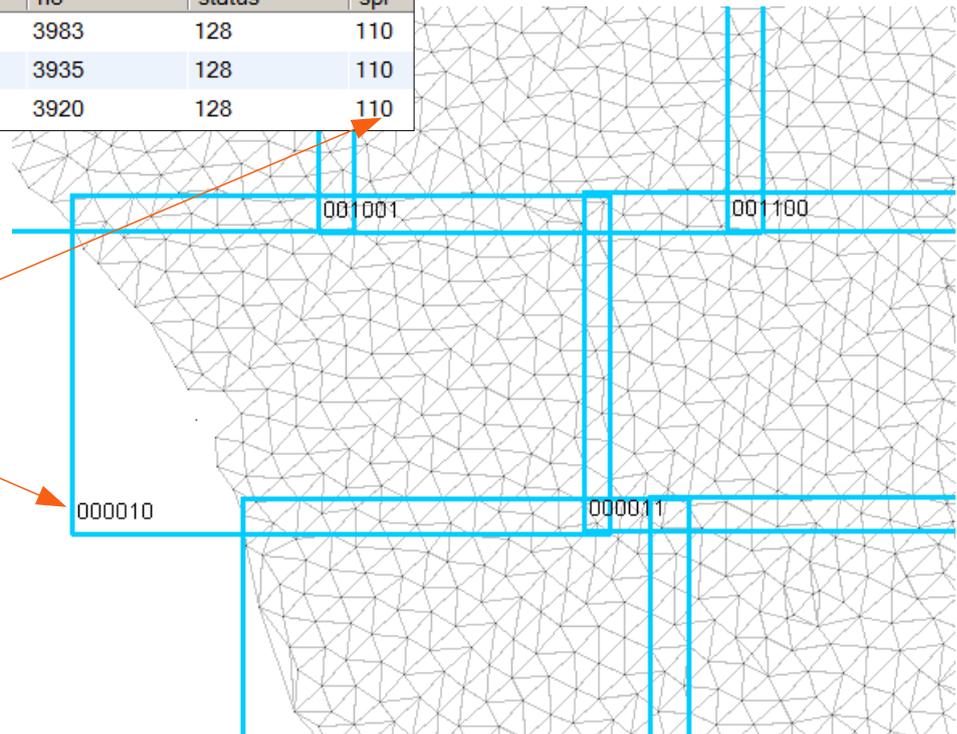
Suchbaum für unstrukturierte Daten – Räumlicher Index

Tabelle data2elements (Basis-Tabellenname + elements)

id	md_id	nr	n0	n1	n2	n3	status	spi
1	3	0	4206	3992	3983	3983	128	110
2	3	1	3934	4206	3935	3935	128	110
3	3	2	3886	4230	3920	3920	128	110

die Geometrien „kennen ihr Blatt im R-Tree“,
SELECT-Statement zur Extraktion der Geometrien
eines Blattes sind einfach formulierbar!

Binärpfad zum
Blatt des Baumes

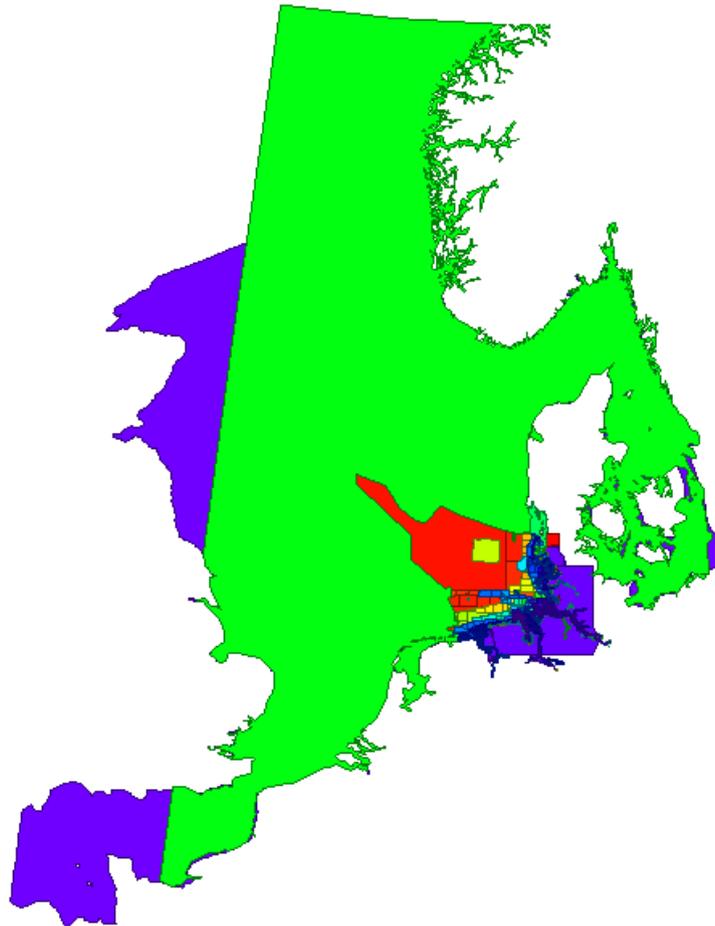


id	md_id	spi	depth	minx	miny	maxx	maxy	pointcount	edgecount	polygonedgecount
2	3	0	6	529694.5585540001	6075010.048557	529723.481446	6075038.911443001	356	0	0
3	3	2	7	529690.578996	6075027.6091665	529710.6610040001	6075044.2808335	96	0	0
4	3	3	7	529706.188919	6075037.549109999	529727.8110810001	6075055.3508899985	51	0	0

Tabelle data2tree (Basis-Tabellenname + tree)

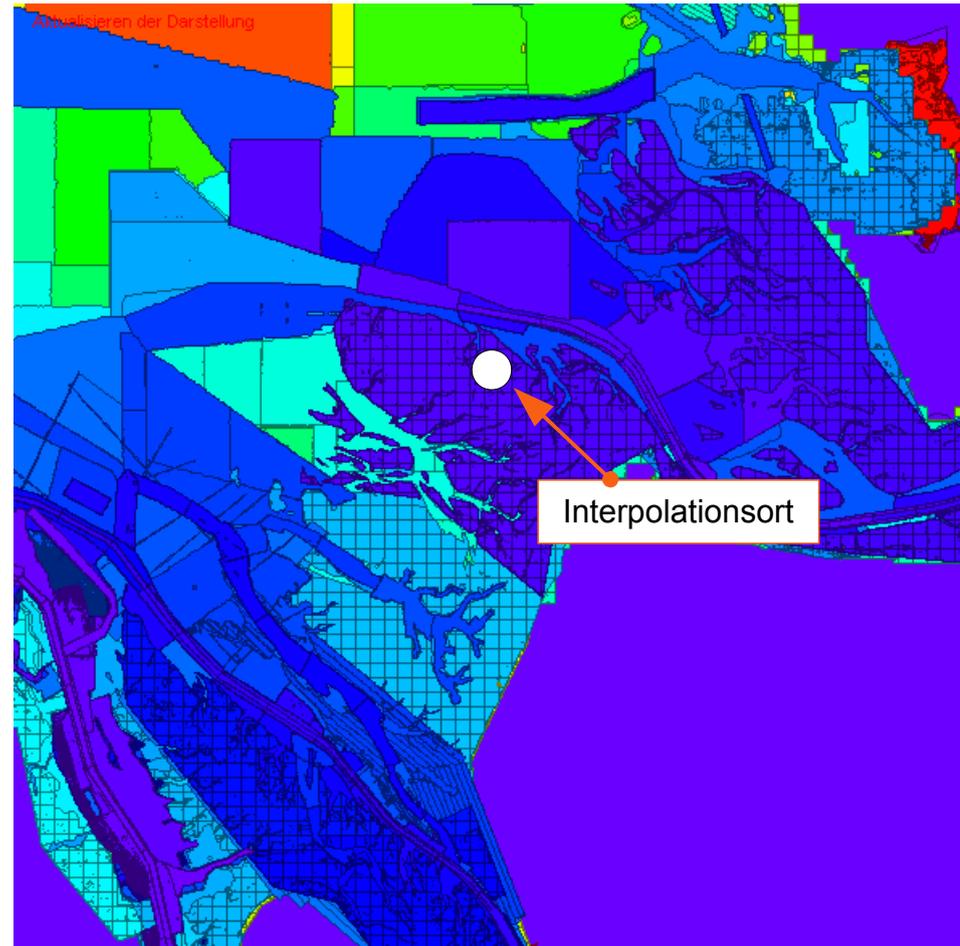
Konzeption der räumlichen Suche

(beispielhaftes Anwendungsszenario)

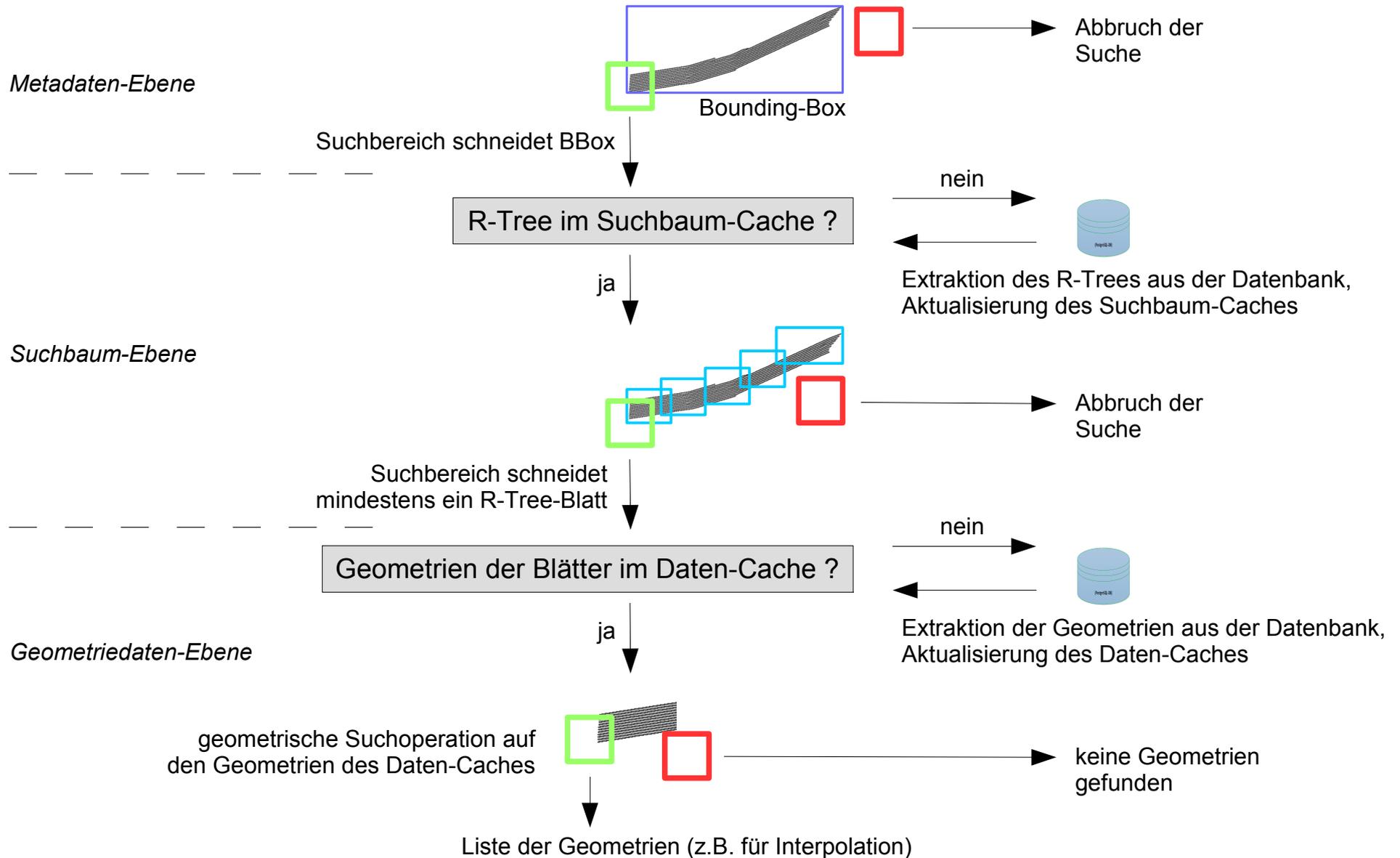


63771 Datensätze
36,1 Mrd. Datenpunkte
3,3 Mrd. Dreiecke
2,6 Mio. Strukturpolygone

Aufgabenstellung:
Interpolation des Höhenwertes auf dem aktuellsten Datensatz



Konzeption der räumlichen Suche



Tabellenstruktur für Metadaten

Tabelle metadatacore

id	metadata_id	title	temporalexten	temporalextent_	srid	minx	maxx	miny	maxy	minz	maxz
1	7ff39375-5c4...	DGMWA001_324705930	2016-09-16 ...	2016-09-16 ...	25832	470000.5	470999.5	5930000.5	5930999.5	-16.388751...	9.48099994...
2	80559ea1-ac...	DGMWGREN_324705930	2016-09-16 ...	2016-09-16 ...	25832	470000	471000	5930000	5931000	-99999999	-99999999
3	2e21b356-04...	DGMWA001_324715930	2016-09-16 ...	2016-09-16 ...	25832	471000.5	471999.5	5930000.5	5930999.5	-19.523790...	8.45375251...
...											
pointcount	polygoncount	stored_ec	elementcount	hullpolygoncount	hullpoly	active	interpolation_function	layertype	geodata_table	geodata_table	
972155	0	0	0	47	BLOB	Y	LINEAR_INTERPOLATION...	24	data0	GRID(x=470...	
22786	488	0	0	2	BLOB	Y	POLYGON_ZONE_INTER...	0	data1	UNSTRUCT...	
906095	0	0	0	46	BLOB	Y	LINEAR_INTERPOLATION...	24	data2	GRID(x=471...	

Core-Metadatentabelle enthält ...

Informationen zur Identifikation des Datensatzes (Titel, UUID)

Zeitliche und Räumliche Ausdehnung

Koordinatensystem-Information (EPSG-Code)

Datensatzgrößen (Datenpunkte, Elemente, ...)

Daten zur grafischen Präsentation der räumlichen Ausdehnung (Hüllpolygone)

Interpolationsmethodik und Aktivitätsflag

Struktur und Speicherort der Daten (Layertyp, Tabellentyp, Basistabellenname)

Tabelle metadatadatadescriptor

id	md_id	content_type	table_name	table_field	geometry	description	short_descript	unit	unit_si	param_values
51	1	ATTRIBUTE50		corrdepthlo...	POINT	CorrDepthlow...		cm		java.lang.Integer
52	1	ATTRIBUTE51		corraltitudeg...	POINT	CorrAltitudeG...		m		java.lang.Double
53	1	ATTRIBUTE0	data0polygo...	verwinprofs...	POLYGON	VerWinProfS...		-		java.lang.String
54	1	ATTRIBUTE1	data0polygo...	profileguid	POLYGON	ProfileGUID		-		java.lang.String

Metadatentabelle enthält ...

Beschreibende Informationen zur Z-Koordinate und den Attributen

Tabellenstruktur für Metadaten

Tabelle metadataasciidata

id	md_id	elementset	content_type	keyword	text
15	1	4	text/plain	line_dist_max	50m
13	1	4	text/plain	surveyor	Krüger
11	1	4	text/plain	line_dist_plan	50m
12	1	4	text/plain	data_status	in Bearbeitung
10	1	4	text/plain	import_filetype	WinProfil
8	1	4	text/plain	ship	K4,KTest
9	1	4	text/plain	project	0312

Tabelle metadatadatetimedata

id	md_id	elementset	keyword	date_
5	1	4	data_lastmodified	2015-10-05 11:17:46
4	1	4	data_statuslastmod...	2015-10-05 10:17:34
2	1	4	md_created	2015-10-05 10:17:34
3	1	4	md_lastmodified	2015-10-05 11:17:46
8	2	4	md_created	2015-11-12 12:19:25
9	2	4	md_lastmodified	2015-11-12 12:20:12
10	2	4	data_deliverydate	2015-11-12 12:19:25

Tabelle metadatadoubledata

Tabelle metadataaintdata

Tabelle metadatabinarydata

Metadatatabellen enthalten ...

Elementset-Information zur Filterung der Metadaten nach Extraktionslevel *Brief, Summary, Full*

Gismo-internes Schlüsselwort des Metadaten-Elementes

Wert des Metadaten-Elementes für die Datentypen Ascii (CLOB), Integer, Double, Datum, Binär (BLOB)

Dynamisches Tabellenkonzept (Anmerkungen)

- Tabellenstruktur passt sich an die Struktur der zu speichernden Daten an
- Datenmodell- / Datenbankstruktur erlaubt (mit moderaten Modifikationen) die Unterstützung künftiger Aufnahmeverfahren (z.B. Laserbathymetrie)
- Logik ist leicht erweiterbar für spezielle Lösungsansätze
 - Integration Rasterdatentyp für PostgreSQL
- Metadaten bieten hohe Flexibilität hinsichtlich der Definition individueller, aufgaben-spezifischer Metadaten-Elemente
- das dynamische Tabellenkonzept erfordert einen Datenbank-Login mit entsprechenden Rechten zum Anlegen, Löschen von Tabellen und Indices!

Rechte- und Rollenkonzept

- in der Standard-Konfiguration sind keine Nutzungs- und Zugriffseinschränkungen aktiv
 - Anmeldung an einem Server erfordert die Login-Daten des Datenbank-Accounts
- Konfiguration von Gismo zum automatisierten Verbindungsaufbau zu einem oder mehreren Datenbank-Servern beim Programmstart möglich
 - keine Kenntnis der Login-Daten des Datenbank-Accounts für den Anwender erforderlich
 - keine Zugriffs- bzw. Nutzungseinschränkung für die Datenbank-Operationen
- Konfiguration von Gismo zum automatisierten Verbindungsaufbau und Nutzung einer applikations-eigenen Nutzerverwaltung
 - (derzeit umgesetzte) Rollen: *Nutzer, Bearbeiter, Prüfer, Sachgebietsleiter, Administrator*
 - Beschränkung der Verfügbarkeit der Funktionalität in Gismo für die einzelnen Rollen

Besonderheiten der Unterstützung von MySQL, PostgreSQL, Oracle

- Implementierung des dynamischen Tabellenkonzeptes für alle unterstützten DBMS identisch
 - (derzeit) ausschließliche Nutzung der Methoden von JDBC
- Datentyp-Mapping berücksichtigt die produkt-spezifischen Eigenschaften der unterstützten Datenbanksysteme

Java-Typ	MySQL	PostgreSQL	Oracle
long	BIGINT	BIGINT	NUMBER(19,0)
int	INTEGER	INTEGER	NUMBER(10,0)
short	SMALLINT	SMALLINT	NUMBER(5,0)
byte[]	LONGBLOB	BYTEA	BLOB
boolean	CHAR(1)	CHAR(1)	CHAR(1)
double	DOUBLE	DOUBLE	BINARY_DOUBLE
float	FLOAT	REAL	BINARY_FLOAT
char	CHAR(1)	CHAR(1)	CHAR(1)
char[]	VARCHAR	VARCHAR	VARCHAR2
String	LONGTEXT	TEXT	CLOB
Calendar	DATETIME	TIMESTAMP	TIMESTAMP

Besonderheiten der Unterstützung von MySQL, PostgreSQL, Oracle

- Datenbankgrößen produkt-abhängig abweichend

Beispieldatensatz mit 3,2 Mio. Datenpunkten, 75 Polygonen, 6,4 Mio. Elementen

JBF [MB]	MySQL [MB]	PostgreSQL[MB]	Oracle [MB]
188	628	1066	942

- Import- und Zugriffsperformance produkt-abhängig abweichend

Import in DB: Beispieldatensatz mit 3,2 Mio. Datenpunkten, 75 Polygonen, 6,4 Mio. Elementen

JBF [sec]	MySQL [sec]	PostgreSQL[sec]	Oracle [sec]
-	430	733	-

(Hinweis: davon Generierung Hüllpolygone: 51 sec, Generierung R-Tree: 84 sec)

- nicht alle GUI-Funktionen werden in den jeweiligen DBMS unterstützt
 - Oracle: keine Erstellung von Datenbanken aus Gismo

PostGIS Rasterdatentyp

- prototypische Erweiterungen der Datenbankschnittstelle um die Speicherung von Rastern bei Verwendung von PostgreSQL/PostGIS als Datentyp
`raster`
- Logik zur Auswahl der Speicherstrategie für Raster erfolgt im Tabellen-Manager
- Umfangreiche, serverseitige Bibliothek für geometrische Operationen erlaubt u.a. den Zugriff auf Sub-Raster mit (optional) modifizierter Rasterweite
- Nutzung für „DGM-Produkte“
- Anwendungsbeispiel : 1656 Kacheln mit 1 Mio. Stützstellen
 - DB-Batch-Import : 1 h 12 min
 - Datenbankgröße : 1497 MB



Kontakt

Dipl.-Ing. Christoph Lippert

post: smile consult GmbH
Vahrenwalder Straße 4
30165 Hannover

tel: 0511 / 543617-43

fax: 0511 / 543617-66

mail: lippert@smileconsult.de

web: <http://www.smileconsult.de>

smile:)
consult