

# Hydroinformatik I - WiSe 2019/2020

## V11: BigData

Prof. Dr.-Ing. habil. Olaf Kolditz

<sup>1</sup>Helmholtz Centre for Environmental Research – UFZ, Leipzig

<sup>2</sup>Technische Universität Dresden – TUD, Dresden

<sup>3</sup>Center for Advanced Water Research – CAWR

Dresden, 10.01.2020

## Semester-Fahrplan

WiSe 2019/2020: Hydroinformatik I, Freitag (3. DS) 11:10-12:40, HÜL/S186/H					
No	KW	Datum	ID	Vorlesung	Dozent
1a	42	19.10.2019	BHYWI-08-01	Hydroinformatik - Einführung	Kolditz
1b	42	19.10.2019	BHYWI-08-02	Compiler (Installation)	Kolditz
2	43	25.10.2019	BHYWI-08-03	Datentypen	Kolditz
3	44	01.11.2019	BHYWI-08-05	Hausaufgaben	Kolditz
5	45	08.11.2019	BHYWI-08-04	Klassen	Kolditz
4	46	15.11.2019	BHYWI-08-06	Programmieren in Nat-Ing	Kalbacher
6	47	22.11.2019	BHYWI-08-07	Input-Output (I/O)	Kolditz
7	48	29.11.2019	BHYWI-08-08	Strings - Textverarbeitung	Kolditz
8	49	06.12.2019	BHYWI-08-09	Hydrologische Modellierung	HA
9	50	13.12.2019	BHYWI-08-10	Hydrologische Modellierung	Kolditz
10	51	20.12.2019	BHYWI-08-11	Pointer & Container	Kolditz
12	2	10.01.2020	BHYWI-08-12	BigData & Water 4.0	Kolditz
13	3	17.01.2020	BHYWI-08-13	Neuronale Netzwerke	Kolditz
14	4	24.01.2020	BHYWI-08-14	ANN / Bayes'sche Netzwerke	Kolditz
15	5	31.01.2020	BHYWI-08-15	BN / Maschinelles Lernen	Kolditz
16	6	07.02.2020	BHYWI-08-16	Klausurvorbereitung	Kolditz

# C++ Programmierkurs: Abschluss

All-in-One: Synthese-Aufgabe

## Übung: E84\_STD

- ▶ Klasse: CStudent
- ▶ I/O: Read/Write
- ▶ Container: vector/list
- ▶ Sortieren von Datensätzen: list

# C++ Programmierkurs: Abschluss

## E84.STD

```
int main()
{
    //-----
    // 1 File handling
    ifstream input_file; // ifstream instance
    input_file.open("HydroInformatics2017.txt");
    ofstream output_file;
    output_file.open("out.txt");
    if(!input_file.good()           // Check is file existing
    {
        cout << "! Error: input file could not be opened" << endl;
        return 0;
    }
    //-----
    // 2 Read data base
    STDRead(input_file);
    //-----
    // 3 Sort data base
    STDSort();
    //-----
    // 4 Write data base
    STDWrite(output_file);
    //-----
    return 0;
}
```

# C++ Programmierkurs: Abschluss

E84.STD

## Aufgaben:

- ▶ Ergänzen von Datensätzen
- ▶ Funktioniert die alphabetische Sortierung?
- ▶ Warum gibt es nach der Programmausführung einen Datensatz mehr (in der Ausgabedatei)?

## Compiler-Optionen

- ▶ `g++ main.cpp student.cpp -o my.exe`
- ▶ IDEs ...
- ▶ Eingabefile (richtiges Verzeichnis)

# Big Data

# Big Data - Übersicht

1. Definition
2. Forschungsbedarf
3. Beispiele aus der Umwelt
4. Wasser 4.0

# Big Data Definition(en)



# Big Data

## Definition: Wikipedia

Der aus dem englischen Sprachraum stammende Begriff Big Data (von englisch big ‚groß‘ und data ‚Daten‘) bezeichnet Datenmengen, welche

- ▶ zu groß,
- ▶ zu komplex,
- ▶ zu schnelllebig oder
- ▶ zu schwach strukturiert

sind, um sie mit manuellen und herkömmlichen Methoden der Datenverarbeitung auszuwerten.[1] Im deutschsprachigen Raum ist der traditionellere Begriff Massendaten gebräuchlich. „Big Data“ wird häufig als Sammelbegriff für digitale Technologien verwendet, die in technischer Hinsicht für eine neue Ära digitaler Kommunikation und Verarbeitung und in sozialer Hinsicht für einen gesellschaftlichen Umbruch verantwortlich gemacht werden.[2] Er steht dabei grundsätzlich für große digitale Datenmengen, aber auch für deren Analyse, Nutzung, Sammlung, Verwertung und Vermarktung.[3]

Source: [https://de.wikipedia.org/wiki/Big\\_Data](https://de.wikipedia.org/wiki/Big_Data)

# Big Data

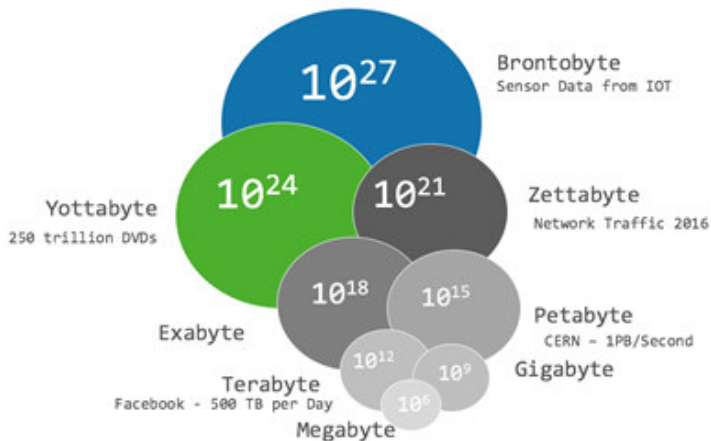
Definition: Wikipedia

In der Definition von Big Data bezieht sich das „Big“ auf die drei Dimensionen volume (Umfang, Datenvolumen), velocity (Geschwindigkeit, mit der die Datenmengen generiert und transferiert werden) sowie variety (Bandbreite der Datentypen und -quellen).[4] Erweitert wird diese Definition um die zwei V's value und validity, welche für einen unternehmerischen Mehrwert und die Sicherstellung der Datenqualität stehen.[5] Der Begriff „Big Data“ unterliegt als Schlagwort einem kontinuierlichen Wandel; so wird mit ihm ergänzend auch oft der Komplex der Technologien beschrieben, die zum Sammeln und Auswerten dieser Datenmengen verwendet werden.[6][7] Die gesammelten Daten können dabei aus verschiedensten Quellen stammen (Auswahl):

Source: [https://de.wikipedia.org/wiki/Big\\_Data](https://de.wikipedia.org/wiki/Big_Data)

# Big Data

## Definition



Source: <http://api.ning.com>

# Big Data

## Definition

### Big Data (Definition)

- **Volume** (Datenmenge -> Fernerkundung)
  - **Variety** (Heterogenität -> multivariante Daten ...)
  - **Velocity** (Geschwindigkeit -> in-situ Visualisierung ...)
  - **Veracity** (Wahrheitsgehalt -> Unsicherheiten ...)
  - **Visualization** (in-situ, VISLAB 2.0)
- **Value**

Data Management

Visual Analytics

# Big Data

## Definition

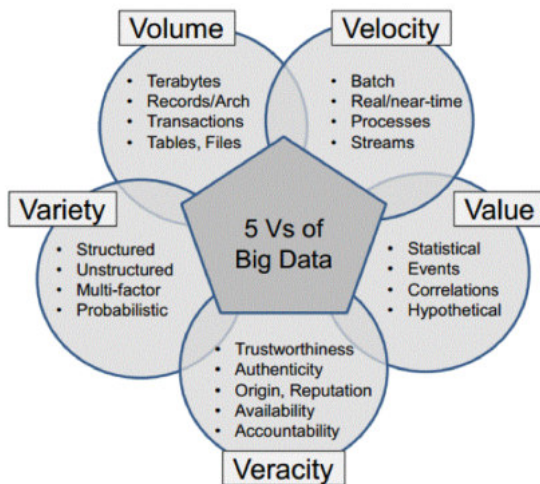


Source <https://www.i-scoop.eu/big-data-action-value-context/>



# Big Data

## Definition



# Big Data Forschung

# Big Data

## Forschung

### Forschungsbedarf (aus Sicht der Visualisierung)

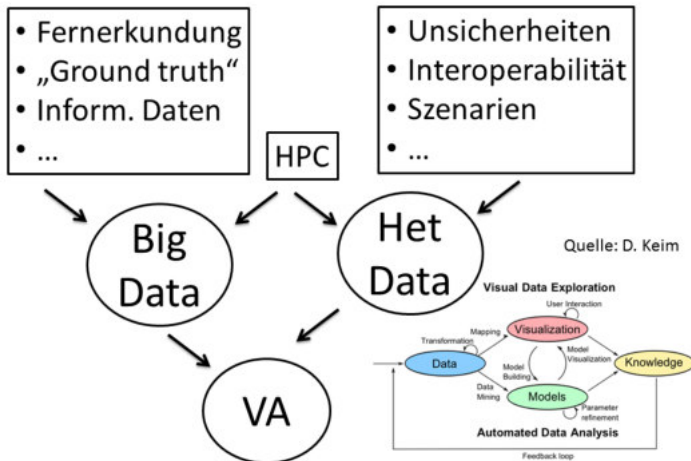
- Challenge: **Big Data**, z.B. aus Fernerkundung (Hyperspektraldaten), hochaufgelöste Geophysik und Simulationsergebnisse (High-Performance-Computing)
- Methodik: **Visual Analytics** (neue Erkenntnisse gewinnen aus großen und heterogenen Datenmengen)
- Methodik: Interaktive **Workflows**: automatisierte Datenanalyse, Qualitätsmanagement (Validierung), visuelle Datenexploration, Mustererkennung, in-situ Visualisierung
- Sichtbarkeit: EUROVIS Workshop „EnvirVis“ (EES TI)



# Big Data

Forschung

## Visual Analytics (VA)



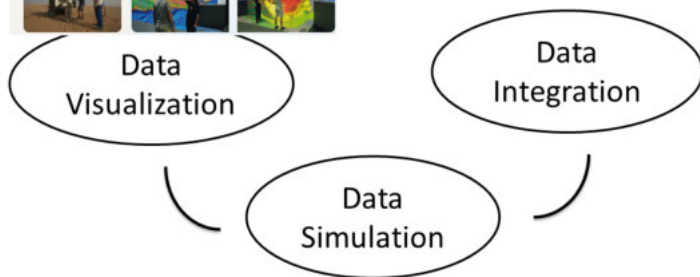
# Big Data

## Forschung

### Interaktive Workflows



- **In-situ Visualisierung**  
(Echtzeitvisualisierung  
modellierter Daten,  
VISLAB 2.0)



# Big Data

Forschung: Umweltinformationssysteme

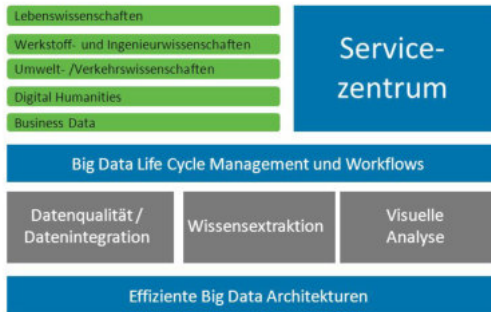
...

# Big Data

Forschung: ScaDS >> ScaDS.AI



## GROBSTRUKTUR DES ZENTRUMS



www.scads.de 12

Source [www.scads.de](http://www.scads.de)

# Big Data

## Beispiele aus der Umwelt

# Big Data

## Behörden

The screenshot shows the navigation bar of the Umwelt Bundesamt (UBA) website. The bar includes the UBA logo and icons for 'Das UBA', 'Themen', 'Presse', 'Publikationen' (highlighted in blue), 'Tipps', and 'Daten'. Below the navigation bar, a breadcrumb trail reads 'Publikationen > Umweltinformationssysteme'. The main content area features a cover image of the publication 'Umweltinformationssysteme 58/2015' with the subtitle 'Big Data - Open Data - Data Variety'. The article title is 'Umweltinformationssysteme Big Data – Open Data – Data Variety' under the category 'Nachhaltigkeit | Strategien | Internationales'. The text discusses the challenges of handling large data volumes in environmental information systems and the need for clear conceptual alignment and appropriate IT tools.

**Umwelt Bundesamt**

Das UBA Themen Presse **Publikationen** Tipps Daten

Publikationen > Umweltinformationssysteme

Nachhaltigkeit | Strategien | Internationales

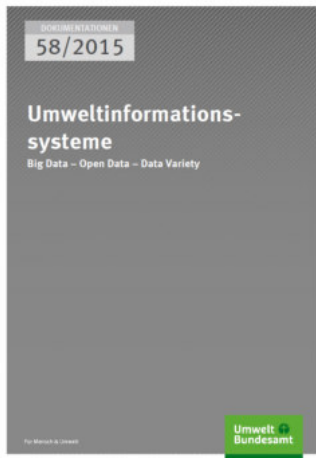
**Umweltinformationssysteme**  
Big Data – Open Data – Data Variety

Mit dem Zuwachs an Daten und Informationen in Datenbanken und Informationssystemen steht die angewandte Umweltinformatik ambitionierten Aufgaben gegenüber. Die Aufbereitung dieser großen Datenmengen für verschiedene Zielgruppen erfordert eine klare konzeptionelle Ausrichtung der Anwendungsentwicklungen. Die Nutzung von Methoden und Werkzeugen der Informatik ist ein Weg, Produkte und Dienste aus diesen Daten zu generieren. Die Bereitstellung dieser Produkte für die verschiedenen Nutzergruppen wie der wissenschaftlichen Community der Modellierer, der Fachnutzer in Umweltbehörden oder eine App auf mobilen Endgeräten für die Öffentlichkeit spiegeln die breite Vielfalt von Datenangeboten wider.

Source: UBA

# Big Data

## Behörden



Source: UBA

## Themen (Workshop 2015)

- ▶ Umwelt-Sensordaten (SOS Web Services)
- ▶ Metadaten
- ▶ Geodateninfrastrukturen
- ▶ Cloud Computing für die Kalibrierung von Hochwassersimulationen
- ▶ Geovisualisierung
- ▶ Crowdsourcing
- ▶ ...

# Big Data

Forschung: Fernerkundung

## Erfassung und Auswertung von Hyperspektraldaten

### Früherkennung von Waldschäden

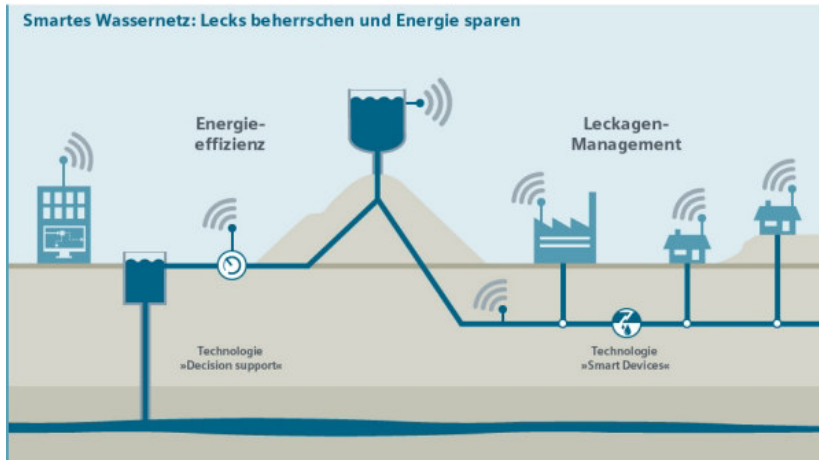
- Vorbereitung schneller, sicherer und umweltschonender Maßnahmen zur Früherkennung und Prävention von Kalamitätsfällen im Forstbereich
- Hochdimensionale funktionale und hochaufgelöste 3D-Rasterdaten aus Befliegungen von Waldgebieten
- Eine Vielzahl von Einflussfaktoren überlagert das hochdimensionale Signal
- Kalibrierung der Spektraldaten durch Methoden des maschinellen Lernens
- Algorithmen zur Separierung einzelner Baumkronen und Ableitung deren hyperspektraler Signatur
- Aussagen zur Baumart, Vitalität sowie zu biotischen und abiotischen Stressfaktoren





# Big Data

## Industrie: Von Big to Smart Data



Source: Siemens

<https://www.siemens.com/innovation/de/home/pictures-of-the-future/digitalisierung-und-software/von-big-data-zu-smart-data-projekt-icewater.html>

# Big Data Wasser 4.0

# Big Data

## Industrie 4.0

Industrie 4.0 ist ein Begriff, der auf die Forschungsunion der deutschen Bundesregierung und ein gleichnamiges Projekt in der Hightech-Strategie der Bundesregierung zurückgeht; zudem bezeichnet er ebenfalls eine Forschungsplattform.[1][2][3] Die industrielle Produktion soll mit moderner Informations- und Kommunikationstechnik verzahnt werden.[4] Technische Grundlage hierfür sind intelligente und digital vernetzte Systeme. Mit ihrer Hilfe soll eine weitestgehend selbstorganisierte Produktion möglich werden: Menschen, Maschinen, Anlagen, Logistik und Produkte kommunizieren und kooperieren in der Industrie 4.0 direkt miteinander.[4] Durch die Vernetzung soll es möglich werden, nicht mehr nur einen Produktionsschritt, sondern eine ganze Wertschöpfungskette zu optimieren. Das Netz soll zudem alle Phasen des Lebenszyklus des Produktes einschließen – von der Idee eines Produkts über die Entwicklung, Fertigung, Nutzung und Wartung bis hin zum Recycling.[4]

Source: Wikipedia

# Big Data

## Industrie 4.0

Mit der Bezeichnung "Industrie 4.0" soll das Ziel zum Ausdruck gebracht werden, eine vierte industrielle Revolution einzuleiten.

1. Die erste industrielle Revolution bestand in der Mechanisierung mit Wasser- und Dampfkraft, darauf folgte
2. die zweite industrielle Revolution: Massenfertigung mit Hilfe von Fließbändern und elektrischer Energie,
3. daran anschließend die dritte industrielle Revolution oder digitale Revolution mit Einsatz von Elektronik und IT (v. a. die speicherprogrammierbare Steuerung) zur Automatisierung der Produktion.

Mit dem Ausdruck „4.0“ wird Bezug genommen auf die bei Software-Produkten übliche Versionsbezeichnung, die bei größeren Änderungen von einer neuen Version spricht, die erste Ziffer der Versionsnummer um Eins erhöht und gleichzeitig die zweite Ziffer auf Null zurücksetzt.

Source: Wikipedia

# Big Data

## Wasser 4.0

### GWP-Arbeitskreis Wasser 4.0



© Siemens

» [GWP-Broschüre "Wasser 4.0" herunterladen in Deutsch und Englisch](#) »

Source: German Water Partnership

# Big Data

## Wasser 4.0

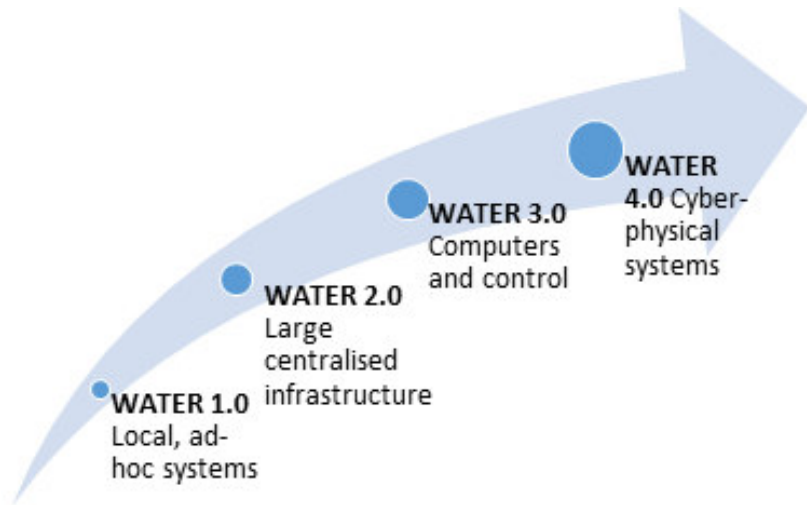
Wasser 4.0 stellt die Digitalisierung und Automatisierung in den Mittelpunkt einer Strategie für eine ressourceneffiziente, flexible und wettbewerbsfähige Wasserwirtschaft.

Source: German Water Partnership

<http://www.germanwaterpartnership.de/de/arbeitskreise/wasser-40/index.htm>

# Big Data

## Wasser 4.0

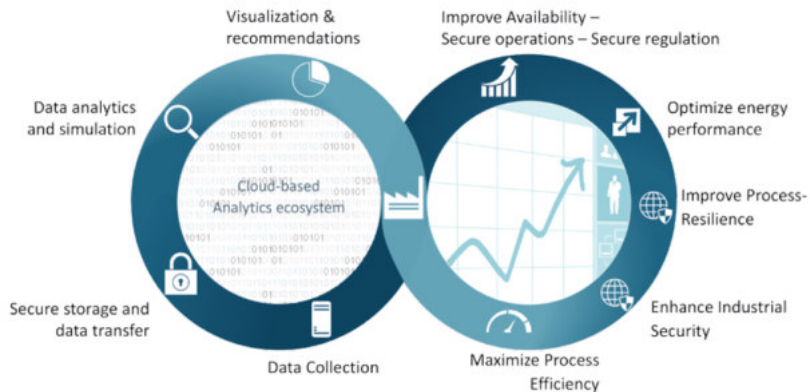


# Big Data

Wasser 4.0

From Data...

...to Value



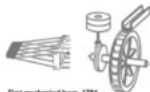


# Big Data

## Wasser 4.0

### 1<sup>st</sup> Industrial Revolution

Mechanization of work, powered by water and steam



First mechanical loom, 1784

### 2<sup>nd</sup> Industrial Revolution

Specialized mass production of goods, powered by electricity



First production line, slaughterhouse Chickell, 1870

### 3<sup>rd</sup> Industrial Revolution

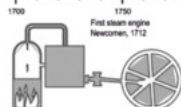
Electronics and information technology facilitate automated production; ICT gives rise to computerization



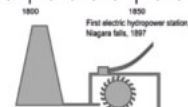
First programmable logic controller (PLC), Modicon 084, 1969

### 4<sup>th</sup> Industrial Revolution

Intelligent devices in intelligent global networks provide permanent availability and analysis of data and information; Merging of physical and virtual worlds into Cyber-Physical Systems (CPS); Internet of Things (and services)



First steam engine  
Newcomen, 1712



First electric hydropower station,  
Niagara falls, 1897



First computer models  
European Hydrological System (SHE), 1977  
FEFLOW, 1979



### 1<sup>st</sup> Revolution in Water Management

Utilisation of steel to handle high water pressure (steam boilers, hydraulic steelwork)

### 2<sup>nd</sup> Revolution in Water Management

Pumps and turbines use and generate electricity

### 3<sup>rd</sup> Revolution in Water Management

IT capable of physical-numerical modelling of water systems; Integration of field sensors into IT systems

### 4<sup>th</sup> Revolution in Water Management

Interfacing of real and virtual water systems (CPS); Real-time and forecasting models reduce risks and costs; Distribution and collection concepts include internet-based networking through to the end user (Smart sensing)

# Big Data

## Wasser 4.0

Engineering 5 (2019) 828–832



ELSEVIER

Contents lists available at [ScienceDirect](#)

# Engineering

journal homepage: [www.elsevier.com/locate/eng](http://www.elsevier.com/locate/eng)



### Views & Comments

## Environmental Information Systems: Paving the Path for Digitally Facilitated Water Management (Water 4.0)



Olaf Kolditz<sup>a,b,e</sup>, Karsten Rink<sup>a</sup>, Erik Nixdorf<sup>a</sup>, Thomas Fischer<sup>a</sup>, Lars Bilke<sup>a</sup>, Dmitri Naumov<sup>a</sup>, Zhenliang Liao<sup>c,e</sup>, Tianxiang Yue<sup>d,e</sup>

<sup>a</sup>Department of Environmental Informatics, Helmholtz Center for Environmental Research (UFZ), Leipzig 04318, Germany

<sup>b</sup>Applied Environmental Systems Analysis, Technische Universität Dresden, Dresden 01069, Germany

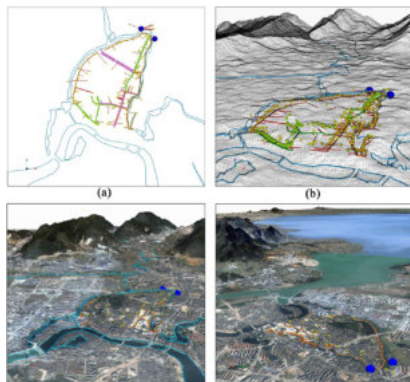
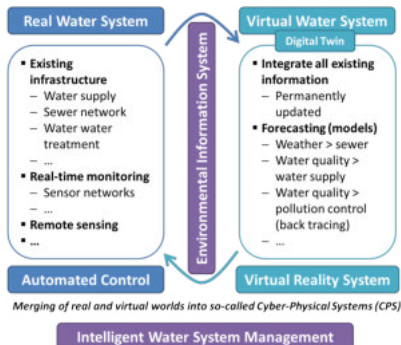
<sup>c</sup>UN Environment–Tongji Institute of Environment for Sustainable Development & College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China

<sup>d</sup>Department for Ecological and Environmental Informatics, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 1000101, China

<sup>e</sup>Sino-German Research Center for Environmental Information Science (RCEIS), Leipzig 04318, Germany

# Big Data

## Wasser 4.0



More details and examples: [www.ufz.de/vislav](http://www.ufz.de/vislav)