PIKOBYTES
DATA-DRIVEN SOFTWARE AND CONSULTING.

Data-Hub Boden

Überblick zur Systemkonzeption

02.12.2025

DataHub Boden – Mission Statement

Der DataHub Boden soll die zentrale Anlaufstelle für einen **institutionsübergreifenden Zugang zu den verschiedenen Monitoringprogrammen des Bundes und der Länder sowie verschiedener Forschungseinrichtungen** sein. Er soll die Inwertsetzung von Daten aus dem Bodenmonitoring erhöhen, indem er verfügbare Daten aus bestehenden Bodenmonitoringaktivitäten zusammenführt (**Integration**) und diese Daten **einem größeren Kreis an Personen und Institutionen zur weiteren Nutzung zur Verfügung stellt** (**Dissemination**). Darüber hinaus soll die **Erstellung messnetzübergereifender Auswertungen** mit geeigneten Werkzeugen erleichtert werden.

Gliederung des Konzeptes

- Anwendungsfälle und User Stories (vom Workshop 06/2025)
- Datenarten
- Nutzergruppen nach Funktionen, Zugangsbeschränkungen
- Nutzungsvereinbarungen und Datenweitergabe
- Datenlieferungen an den DataHub
- Datenbereitstellung und Präsentation für Datennutzer
- Methodendatenbank
- Exemplarische Spezifikation von Anwendungsszenarien

Anwendungsfälle und User Stories

Warum?

- Systemkonzeption: Keine Anforderung ohne Anwendungsfall

- Erfassung von Systemkontext und Hauptaufgaben
- Identifikation relevanter Datenarten, benötigter Funktionalität, prioritärer Datenquellen, Nutzergruppen
- Spezifikation möglicher Anwendungsszenarien

Datenarten im DataHub Boden

Datenarten im DataHub

- Daten aus Beprobungen
- Ortsfeste in-situ Messdaten
- Statische räumlich aufgelöste Kontextdaten mit Bezug zum Boden
- Bohrlochdaten zur Bodenphysik
- Räumlich und zeitlich aufgelöste Daten aus der Fernerkundung
- Räumlich und zeitlich aufgelöste Daten aus der Modellierung

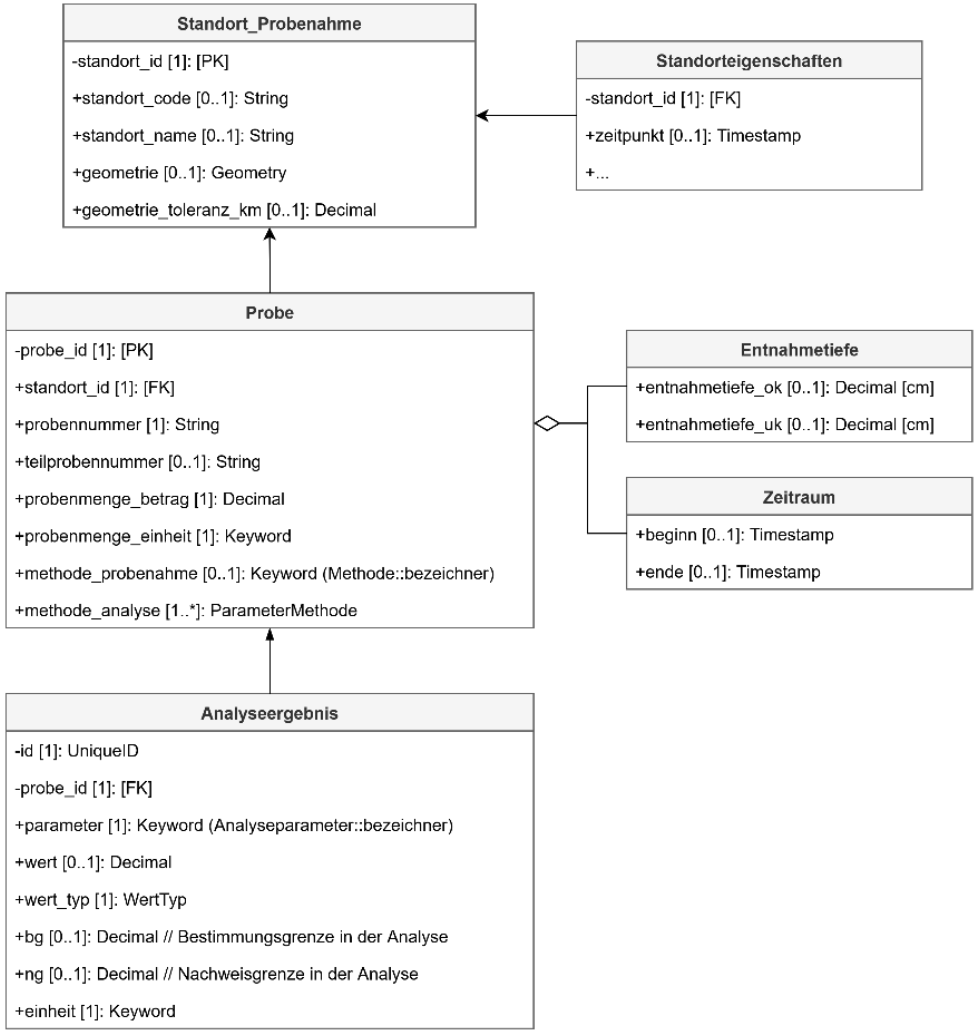
Datenarten

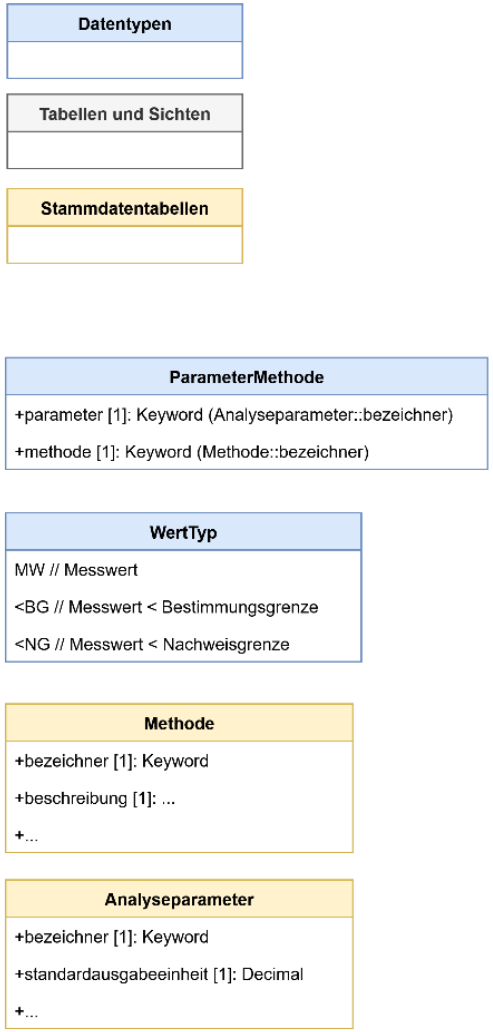
Daten aus Beprobungen

Datenobjekt	Beschreibung
Analyseergebnisse	Kernsatz an auswertungsrelevanten Parametern
Informationen zur Probe	Informationen zur Probenahmemethodik und Analyse-/Labormethodik, Begleitparameter
Standort	Geografischer Standort der Probenahme (Punkt bzw. Fläche)
Standortbeschaffenheit und -umgebung	„Standortsteckbrief“ zu den Umweltbedingungen am Standort selbst und um den Standort herum. Bei langjährig genutzten Standorten ggf. mit Änderungen behaftet.
Zeitpunkt der Probenahme	Angabe eines möglichst exakten Zeitpunktes / Zeitraums

Datenarten

Daten aus Beprobungen





- Zunächst High-Level
- Orientierung an bestehenden Datenbanken
- Detailspezifikation und Validierung im Zuge der Umsetzung

- Agile Umsetzung, erforderliche Anpassungen im Zuge erster Datenübernahmen

Datenarten

Daten aus Beprobungen – **Standortinformationen**

- Ideal: **exakte geografische Lage**, ggf. mit bekannten **Lageunsicherheiten** bei historischen Daten
- Je nach Verfügbarkeit / Bereitstellung: generalisierte bzw. **verschleierte Koordinaten**
- Wenn nicht anders möglich: **Standortübergreifende Aggregation**

- Für wiederkehrend genutzte Standorte (z.B. BDF): stabile Pseudonyme bzw. **Standortkennungen**

- **Angaben zur Standortbeschaffenheit und -umgebung**

Datenarten

Ortsfeste in-situ Messdaten

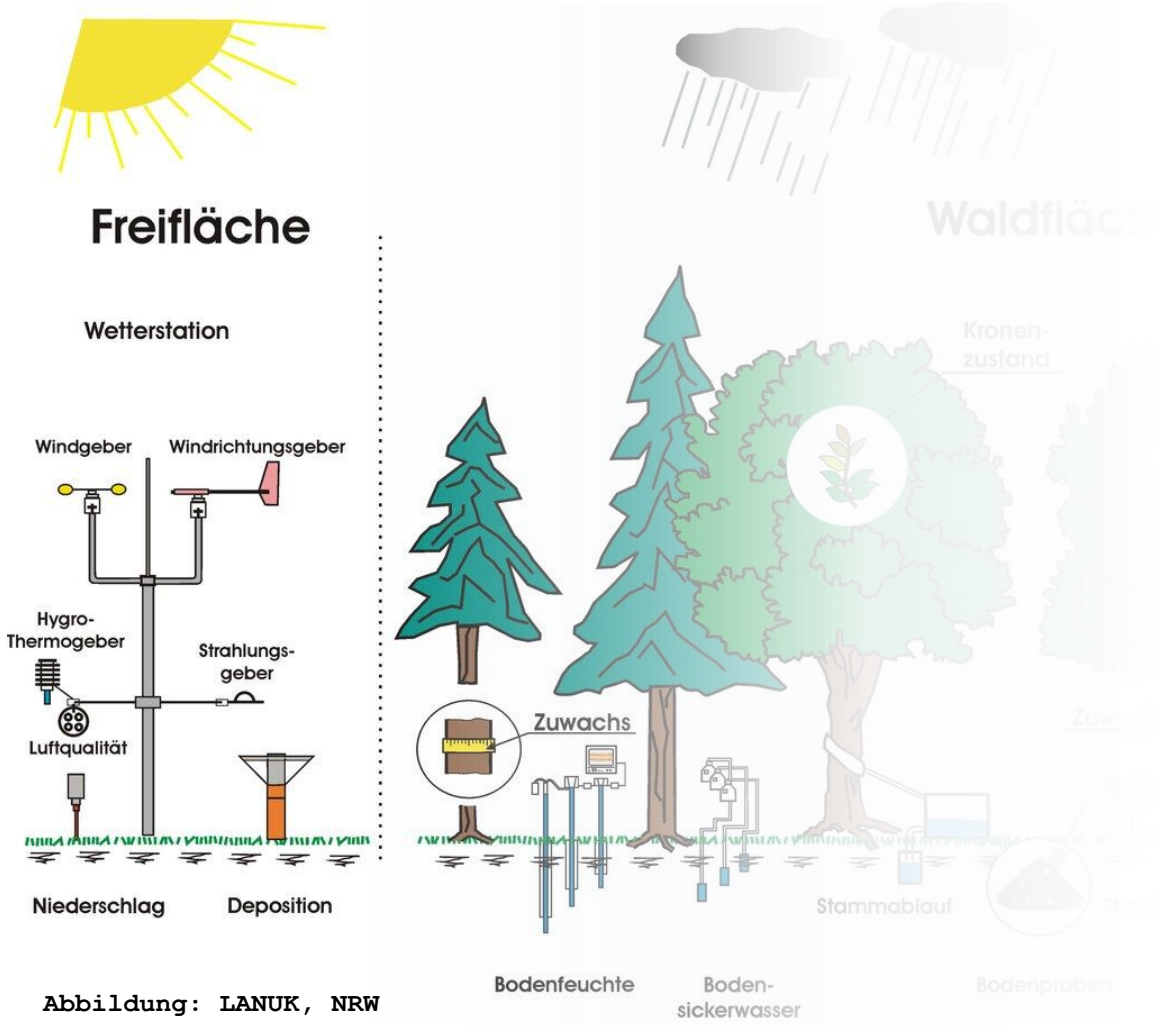


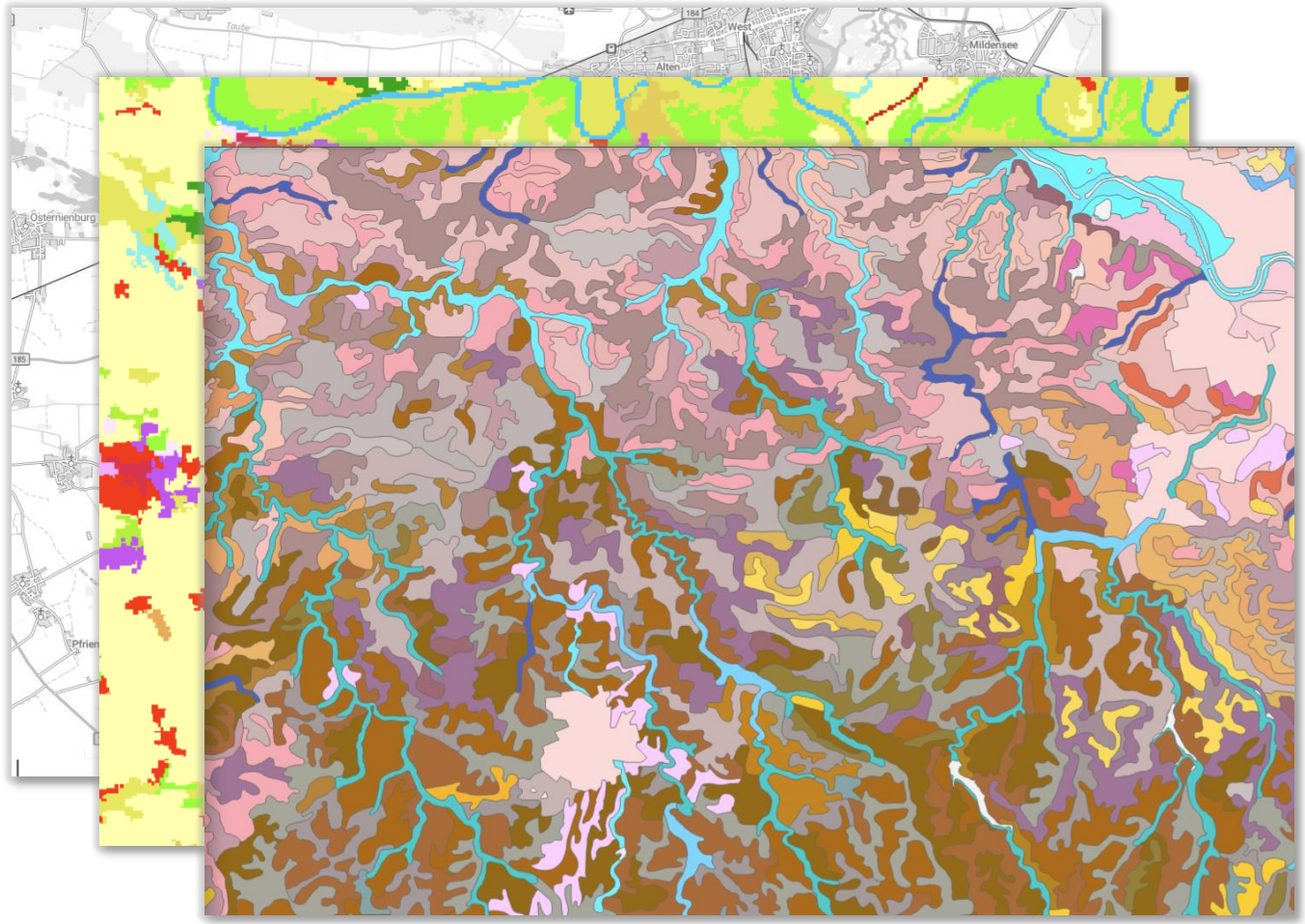
Abbildung: LANUK, NRW

- Daten zum Mess-Standort
- Zeitreihendaten der aufgenommenen Messwerte
- Daten zur Messtechnik und Wartung

- Gängige Qualitätslevel innerhalb der Zeitreihen:
  (1) plausibilisiert
  (2) technisch geprüft
  (3) ungeprüfte Rohdaten

Datenarten

Statische räumlich aufgelöste Kontextdaten mit Bezug zum Boden

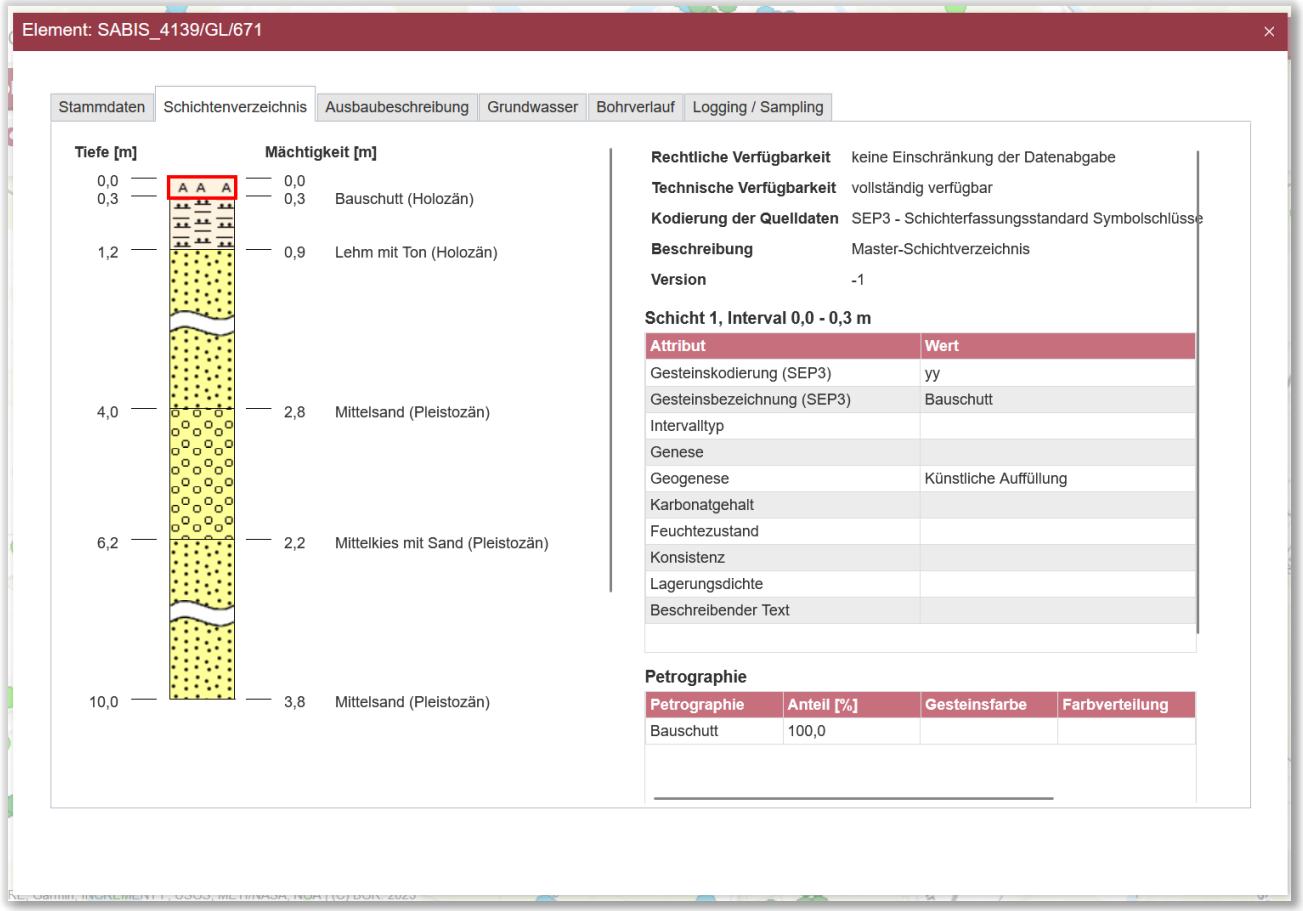


- Topografische Basiskarte
- Corine Land Cover
- BÜK200 / BÜK250 des BGR
- DGM Deutschland
- …

- Gelegentlich neue Versionen, aber keine echten Zeitreihen

Datenarten

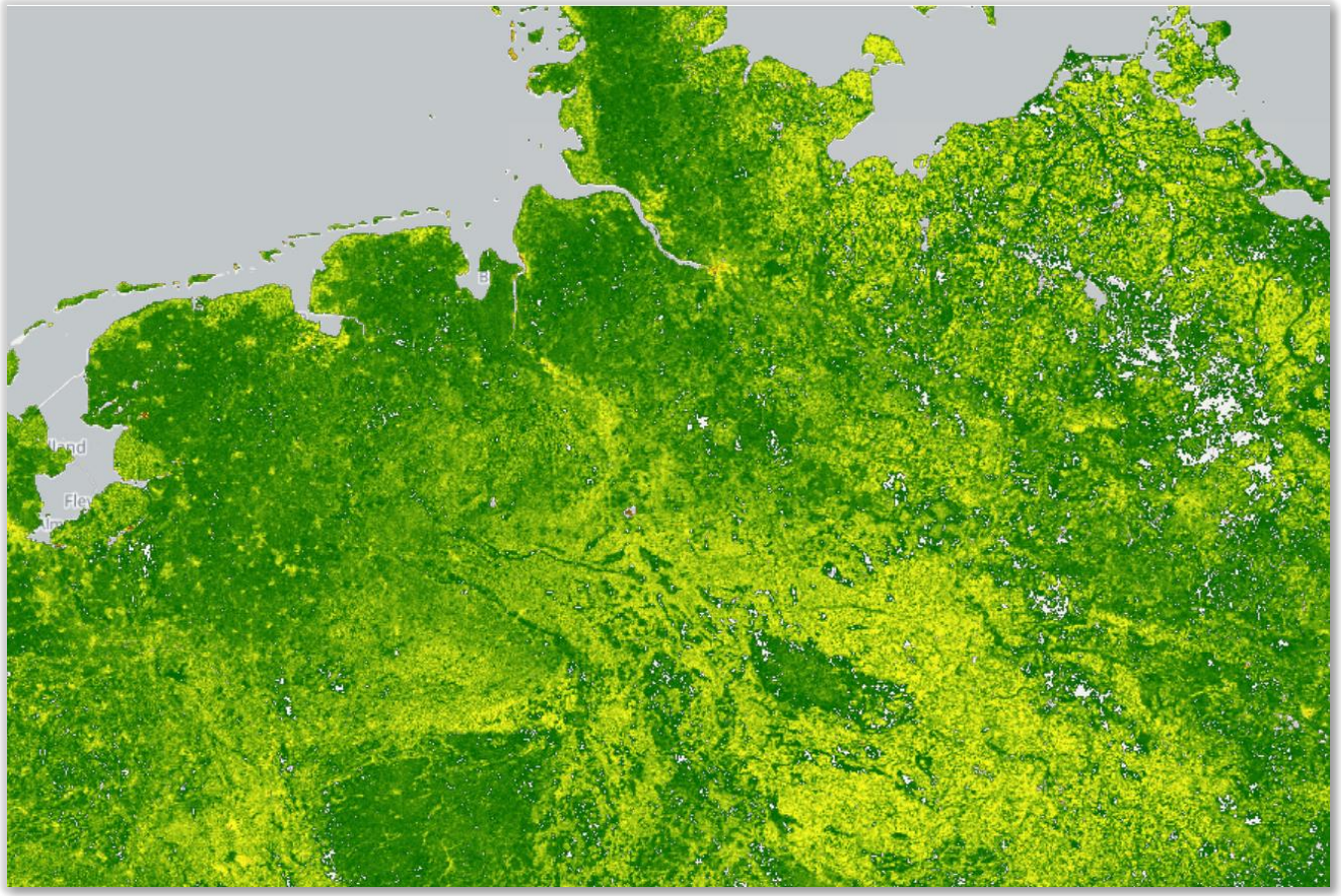
Bohrlochdaten zur Bodenphysik



- Vor allem die Schichtenverzeichnisse
- Anzeige und Meldepflicht für geologische Untersuchungen nach Geologiedatengesetz (GeoIDG)
- Deutschlandweit verfügbar und stetig wachsend

Datenarten

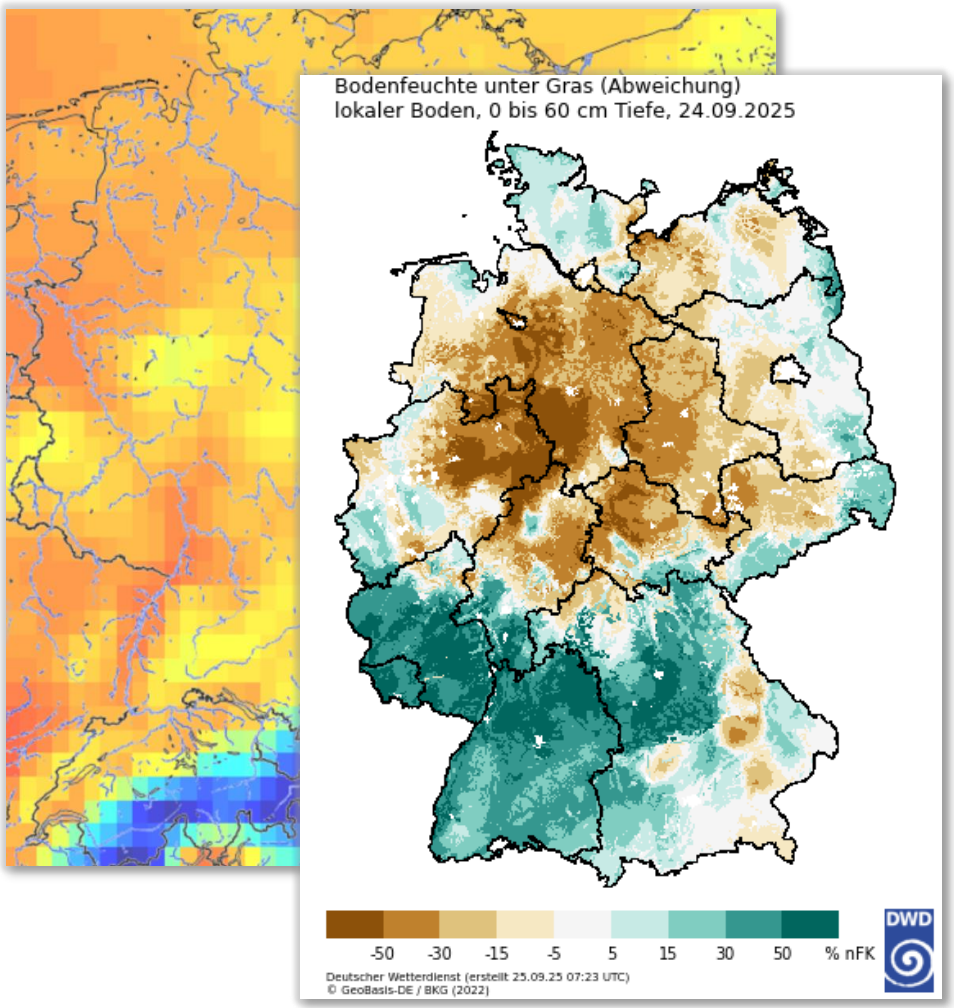
Räumlich und zeitlich aufgelöste Daten aus der Fernerkundung



- Prominent: NDVI
- Weitere Zeitreihen / Rasterdaten via Copernicus verfügbar
- Zeitlich aufgelöste Kontextdaten für bodenbezogene Auswertungen

Datenarten

Räumlich und zeitlich aufgelöste Daten aus der Modellierung



- ERA5-Daten aus Klimareanalysen
- Dürremonitor UFZ
- Bodenfeuchteprodukte des DWD
- …

- Zeitlich aufgelöste Kontextdaten für bodenbezogene Auswertungen

Funktionsbezogene Nutzergruppen

Funktionsbezogene Nutzergruppen

Fachaufsicht DataHub	„Kernteam“ **Vollzugriff** auf den Gesamtdatenbestand – **ausschließlich für Administration** und operativen Betrieb, **keine Datenauswertung**
Privilegierte Datenauswertung	Personenkreis der „Datenauswerter“ ist zahlenmäßig sehr groß und sehr heterogen. • Datenrecherchen • Einsicht in vorgefertigte Reports und ad-hoc-Auswertung • Datenabruf für komplexe Analysen in der eigenen Arbeitsumgebung außerhalb des DHB Für bestimmte Nutzer ist **privilegierter Zugriff auf nicht-öffentliche Daten**, die für die jeweilige Auswerteaufgabe erforderlich sind, notwendig. Ermöglichung durch **Club-Modell / dynamische Rollenmodelle** – dazu später mehr.

Funktionsbezogene Nutzergruppen

Öffentlichkeit	Im Unterschied zur Funktionsgruppe „Privilegierte Datenauswertung“ erfolgt der Zugriff auf den Datenbestand des DHB unprivilegiert, d.h. es können nur solche Daten und Auswertungen eingesehen und bezogen werden, die für die Öffentlichkeit uneingeschränkt freigegeben sind.
Datenbereitsteller	Liefern Daten auf unterschiedlichen Wegen Daten an den DHB. Gelieferte Daten werden dann im DHB vorgehalten und stehen den übrigen Nutzergruppen in • unterschiedlichem Detailgrad, • mit unterschiedlichen Nutzungsbestimmungen und Vertraulichkeitseinstufungen, und • in unterschiedlicher Form zur Verfügung. Datenbereitsteller sollen außerdem überprüfen können, welche Personengruppen auf die von ihnen bereitgestellten Daten zugreifen können und – soweit mit vertretbarem Aufwand feststellbar – wie häufig diese Daten aufgerufen und genutzt werden. Oftmals personelle Überlappung mit der Gruppe **Privilegierte Datenauswertung**

Nutzungsvereinbarungen und Datenweitergabe

Datenweitergabe bodenbezogener Daten

Grundprinzipien

- Freiwilligkeit der Datenbereitsteller

- Öffentlichkeit und aktive Informationspflicht (UIG)
- Ausnahmen aufgrund öffentlicher und privater Belange
  - Geschützte Bereitstellung / vertrauliche Angaben werden einem nur einem begrenzten Personenkreis zur Verfügung gestellt
  - Verfremdung / Generalisierung von Standortangaben

  - Betrifft nach aktuellem Stand ausschließlich Daten aus Beprobungen

Datenweitergabe bodenbezogener Daten

Grundlagen

- Umweltinformationsgesetz (UIG)
- Nutzung von Daten innerhalb der Verwaltung (Land, Bund)
- Besondere Nutzungsvereinbarungen mit Mittelbarer Verwaltung, Forschungseinrichtungen, Unternehmen, …

- Berücksichtigung mündlicher und schriftlicher Zusagen zwischen Datenbereitstellern und Eigentümern beprobter Flächen

Datenweitergabe bodenbezogener Daten

„Rest der Welt“

Mittelbare Verwaltung, Forschungseinrichtungen, einfache Bürger, Unternehmen, …

- *Status quo*:
  - Einzelvereinbarungen – je nach Ausgestaltung: restriktiv, behaftet mit Unsicherheiten, arbeitsintensiv
  - Nach UIG rechtlich nicht haltbare Ungleichbehandlung von Forschung und Öffentlichkeit

Datenweitergabe bodenbezogener Daten

Club-Modell

Mittelbare Verwaltung, Forschungseinrichtungen, einfache Bürger, Unternehmen, …

- Vorschlag **Club-Modell für geschützte Daten**
  - Ähnlichkeit zu Kooperationsverträgen (z.B. in Forschungsprojekten)
  - Vertrag regelt, was mit den Daten geschehen darf und was nicht
  - Mitglieder können beitreten und austreten
  - Soll Nutzungsvereinbarungen zentralisieren

  - … muss an konkreten Fällen entwickelt und erprobt werden – Testszenarien!

Datenlieferungen an den DataHub

Datenlieferungen an den DataHub

Lösungsmöglichkeiten

- **PULL – DataHub holt Daten** regelmäßig oder episodisch aus einer Datenbank / Schnittstelle des Datenbereitstellers ab
  - Ideal, wenn Datenbereitsteller schon über eine gut strukturierte, maschinenlesbare Datenhaltung verfügen
  - Nach einem ersten Stimmungsbild auch die Präferenz der Datenbereitsteller

- **PUSH – Datenbereitsteller sendet aktiv Daten** regelmäßig oder episodisch an den DataHub
  - Alternativer Bereitstellungsweg, insbesondere wenn
    (1) Daten beim Bereitsteller nicht maschinenlesbar vorliegen
    (2) Der Datenbereitsteller das Verfahren PULL technisch nicht realisieren kann

Datenlieferungen an den DataHub

Lösungsmöglichkeiten

- **Versionierung** abgerufener Daten
  - Datenbereitsteller liefern i.d.R. den „Stand nach bestem Wissen“
  - Auch ältere Daten werden ggf. noch nachkorrigiert

- Identifikation (und Bereinigung) von **Dubletten**
  - Datenintegration aus mehreren Quellen birgt immer die Gefahr von Dubletten
  - Erster Schritt: Dublettenrisiko für Datenquellen abschätzen und dokumentieren
  - Zweiter Schritt: Einfache Werkzeuge zur Dublettendetektion
  - Danach Fahren „auf Sicht“ – weitere Maßnahmen / Heuristiken auf empirischer Grundlage

Datenbereitstellung und Präsentation für Datennutzer

Datenbereitstellung für die Nutzer

Übersicht

- Datenkatalog und katalogbasierte Recherchefunktionen
- Kartenviewer
- Datenexport

Datenbereitstellung für die Nutzer

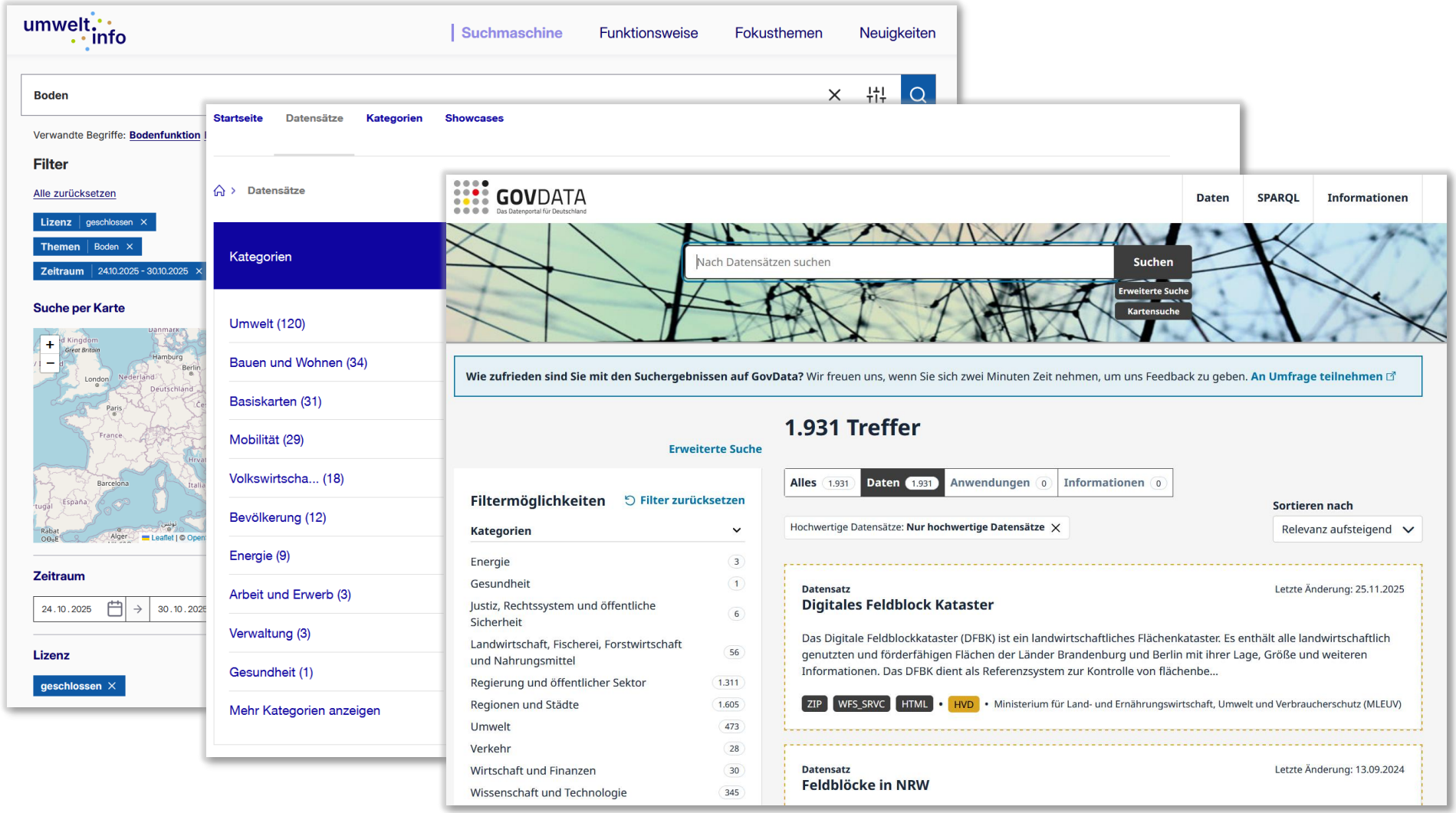
Datenkatalog und katalogbasierte Recherchefunktionen

- Allgemeine Suchfunktionen
  - Kategorien
  - Tags
  - Stichwortsuche
  - Volltextsuche
  - Zeitbezug
- Räumliche Suche
- Datenexploration

- Darstellung von Inhalten und Suchergebnissen
  - Ergebnislisten
  - Detailbeschreibung Datensatz
  - Absprung Datenexport
  - Visualisierung / Karte
- (Berücksichtigung von Zugriffsrechten)

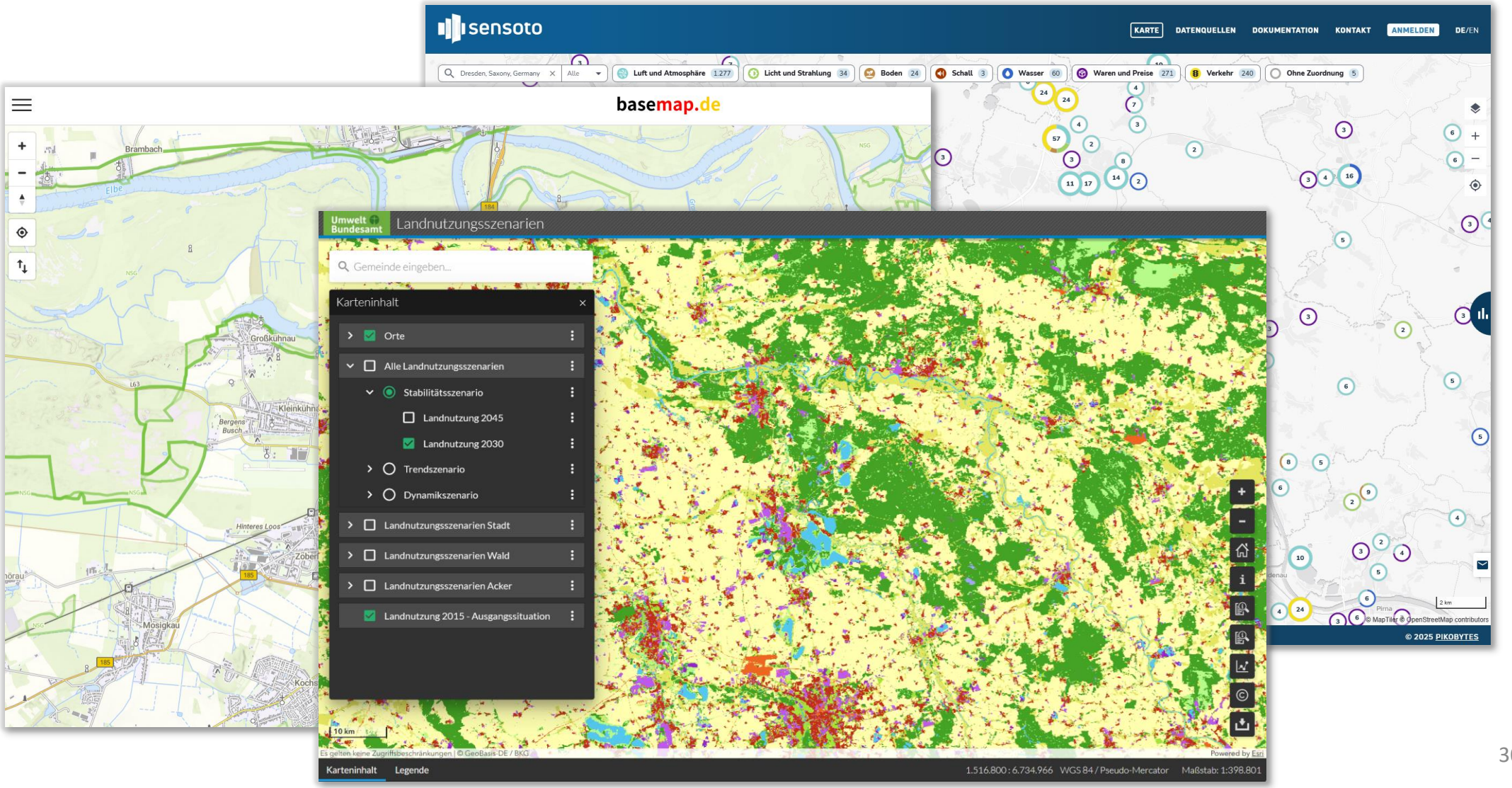
Datenbereitstellung für die Nutzer

Datenkatalog und katalogbasierte Recherchefunktionen



Datenbereitstellung für die Nutzer

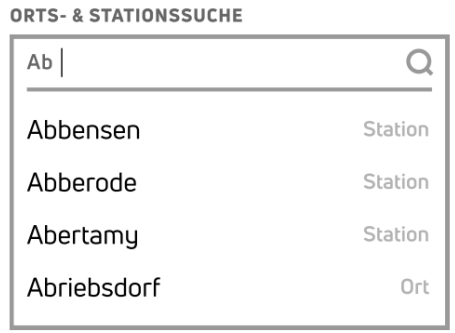
Kartenviewer



Datenbereitstellung für die Nutzer

Kartenviewer

- Primäre Karteninteraktionen (Zoom, Pan), BaseMap, schnelles Nachladen
- **Ortsnamenssuche**
  Allgemeine Ortsnamen, Namen / Kennungen von Beprobungsstandorten
- Legenden, Transparenzregler
- intuitive Nutzerführung und effiziente Bedienung Gelegenheitsnutzer, „Dauernutzer“
- Filterfunktionen
  - Zeit
  - Attributwert, Phänomen / Parameter



Datenbereitstellung für die Nutzer

Kartenviewer

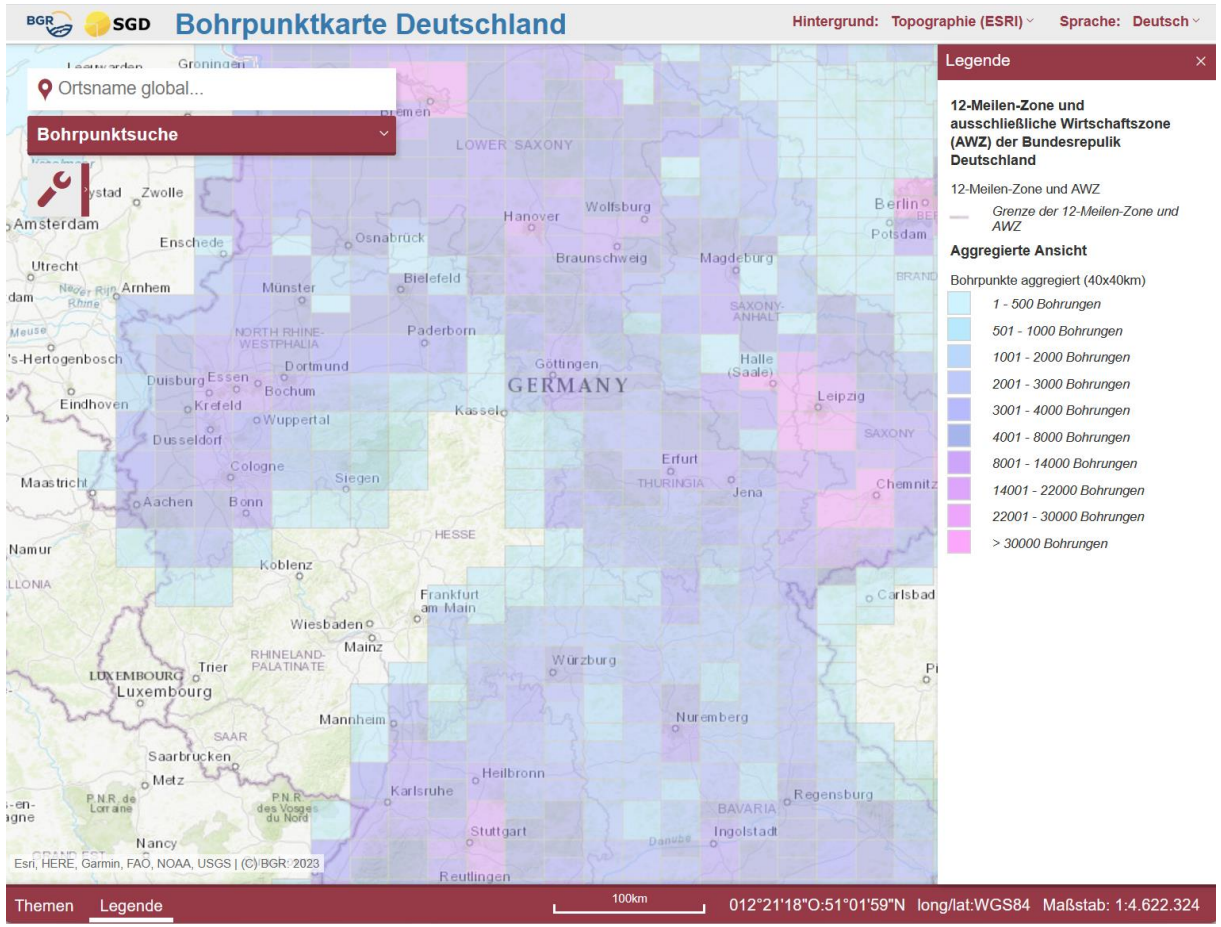
Abbildung: UBA, Bodendaten in Deutschland

Lagedarstellung von Einzelstandorten (Bodenproben, ortsfeste in-situ-Messungen)

- Punkte (hier beispielhaft)
- Ggf. kleine Flächen

Datenbereitstellung für die Nutzer

Kartenviewer

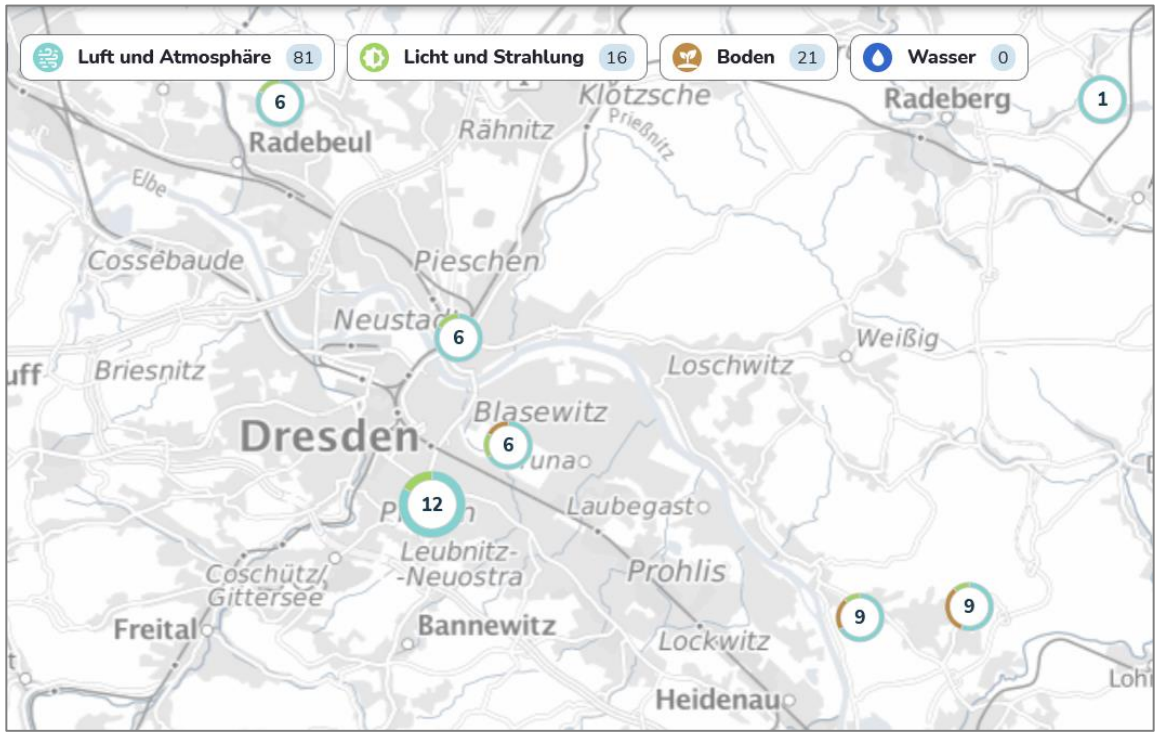


Lagedarstellung von Einzelstandorten (Bodenproben, ortsfeste in-situ-Messungen)

- Aggregation auf Rasterzellen (beispielhaft)
- Auch geeignet zur Darstellung von Erhebungskulissen

Datenbereitstellung für die Nutzer

Kartenviewer



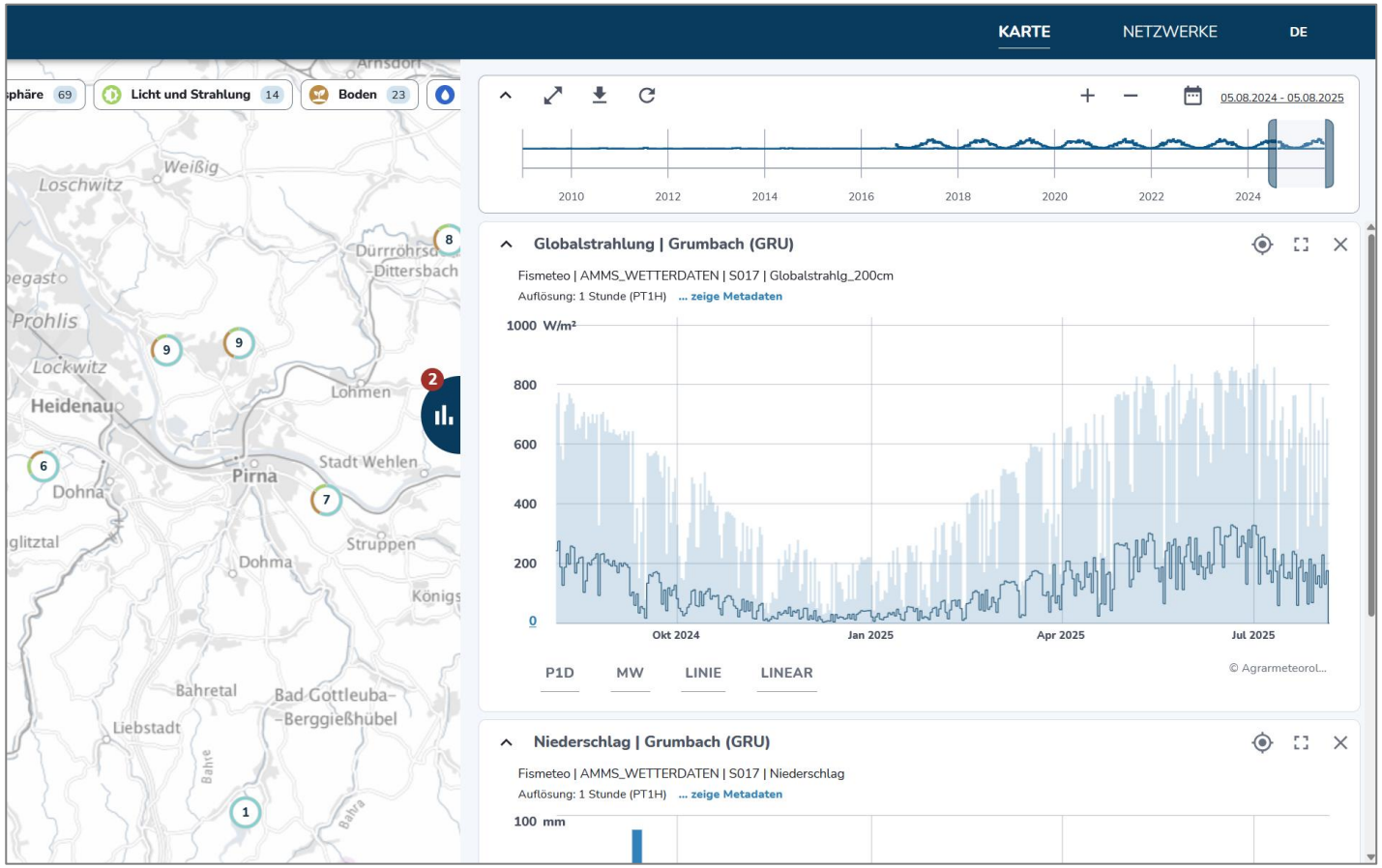
Lagedarstellung von Einzelstandorten (Bodenproben, ortsfeste in-situ-Messungen)

- geclusterte Darstellung von Standorten
- angereichert mit Attributinformationen (hier: Zuordnung zu bestimmten Datenthemen)

Abbildung: FIS Meteorologie, Sachsen

Datenbereitstellung für die Nutzer

Kartenviewer



- Kombi-Darstellung Diagramm und Karte
- Insbesondere für in-situ-Zeitreihen zu Niederschlag, Bodenfeuchte, ...

Abbildung: sensoto.io

Datenbereitstellung für die Nutzer

Kartenviewer

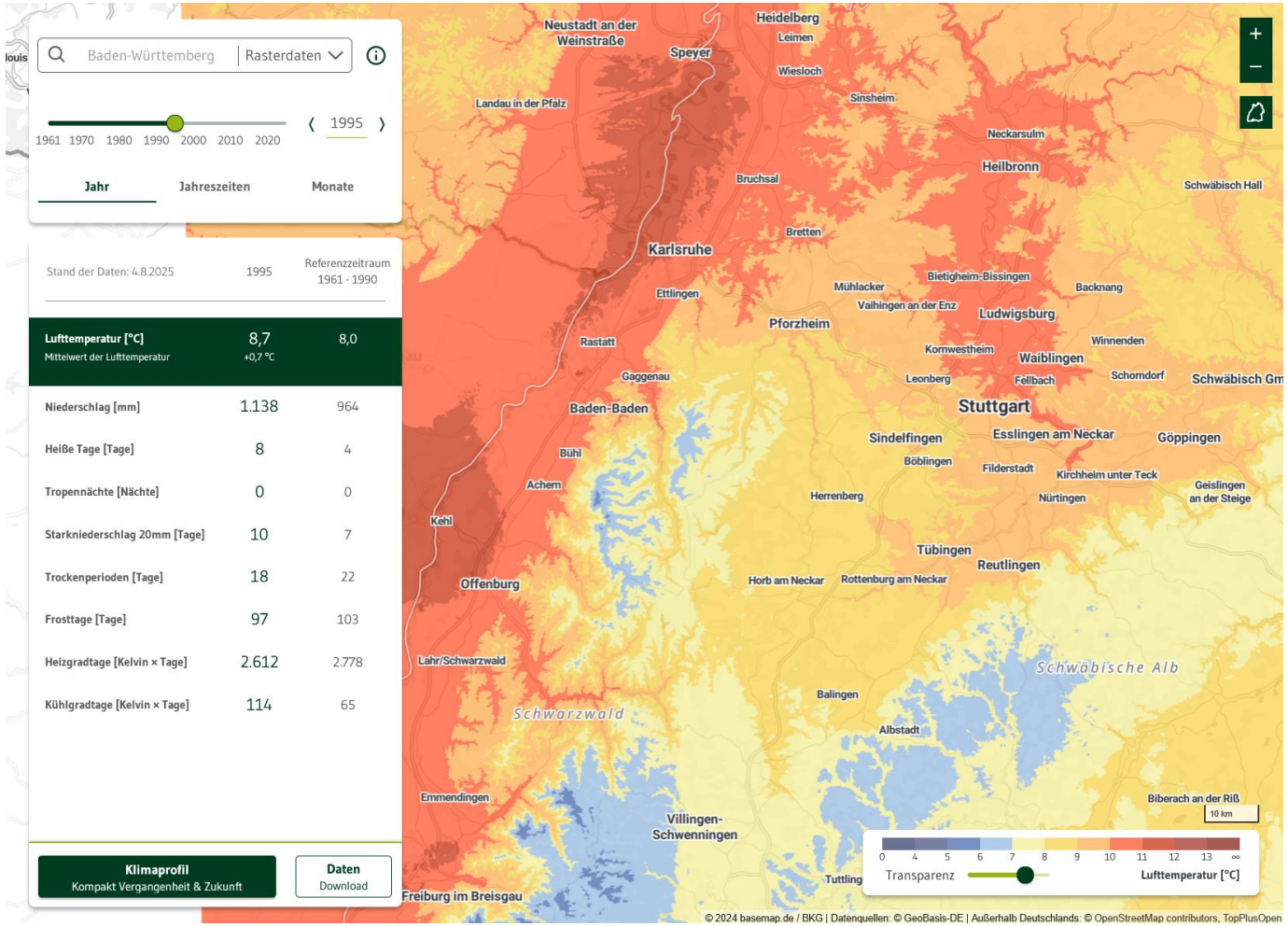


Abbildung: Klimaatlas Baden-Württemberg

- Interaktive Darstellung flächenhafter räumlich und **zeitlich aufgelöster** Daten (Beispielansicht)

Datenbereitstellung für die Nutzer

Detailansichten und Datenexport

- Detailansichten müssen für einzelne Teildatenbestände definiert werden
- Kombinierte Darstellungen / Speziallösungen: Abwägung zwischen Bedienkomfort und Funktionalität und Aufwand

- Datenexport sollte eng am gespeicherten Datenformat erfolgen und in CSV oder JSON maschinenlesbar codiert werden

Methodendatenbank

Methodendatenbank

Für Probenahme- und Analysemethoden

- Methodendatenbank und Methodenschlüssel für Methoden der **Probenahme** und **Analyse**
- Sinnvolle Detailtiefe
- „Fallback“-Methoden für historische Daten, die nur unzureichend tief dokumentiert sind

- Partnerschaftlich mit Datenbereitstellern

Erste Anwendungs-szenarien

Szenario 1

Integration von Daten zum Bodenbiodiversitätsmonitoring im DHB

- Öffentliche und geschützte Daten
- Ableitung von Indikatoren für EU-weites Soil Monitoring
- Erprobung des Club-Modells
- Entwicklung und Erprobung von Verfahren zur Datenlieferung
- Initiale Methodendatenbank
- Schrittweise Zusammenstellung von Erhebungskulissen

- Datenarten
  - Beprobungsdaten zur Regenwurmfauna
  - Beprobungsdaten aus bodenkundlichen und bodenchemischen Untersuchungen

Szenario 2

Begleitparameter Bodenfeuchte

- Integration wichtiger Kontextdaten für die Auswertung von Bodenproben
- Kooperation mit DWD und ISMN in der arbeitsteiligen Zusammenführung und Qualitätssicherung von Bodenfeuchtedaten

- Datenarten
  - Ortsfeste in-situ Messdaten

Szenario 3

Bereitstellung eines Kernensembles bodenbezogener Flächendaten

- Integration wichtiger Kontextdaten für die Auswertung von Bodenproben
- Erprobung und Umsetzung verschiedener Verfahren zur Einbindung externer Datendienste

- Datenarten
  - Flächenhaft aufgelöste Kontextdaten

Szenario 4

Bereitstellung eines deutschlandweiten Datensatzes zum NDVI

- Erprobung zur Integration zeitlich aufgelöster, flächendeckender Rasterdaten
- Klärung offener Frage zur langfristigen Verfügbarkeit relevanter Copernicus-Dienste und Datenprodukte
- Abwägung von Kosten und Nutzen, d.h. gestuftes Vorgehen
- Umgang mit großen Datenmengen

- Datenarten
  - Räumlich und zeitlich aufgelöste Daten aus der Fernerkundung
    Hier: Sentinel-2 oder Copernicus-Produkte

Recap

Übersicht zum Arbeitsstand

- Anwendungsfälle und User Stories
- Nutzergruppen nach Funktionen, Zugangsbeschränkungen
- Datenarten
- Nutzungsvereinbarungen, Datenweitergabe, Club-Modell
- Datenlieferungen an den DataHub
- Datenbereitstellung und Präsentation für Datennutzer
- Methodendatenbank
- Szenarien

PIKOBYTES GmbH
Zittauer Straße 16
01099 Dresden

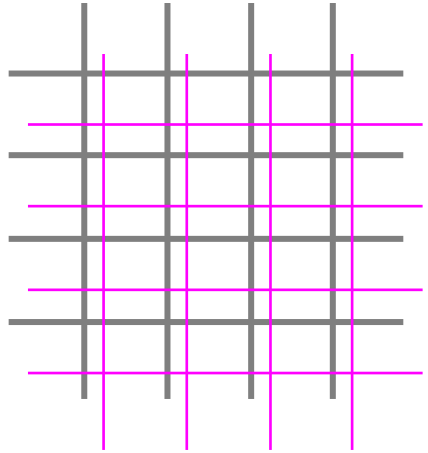
info@pikobytes.de

https://pikobytes.de
https://sensoto.io

Datenarten

Daten aus Beprobungen – **Standortinformationen**

Standortverschleierung – Addition von Lageunsicherheiten



- Bsp: „Abschneiden auf 1 km“

- Welches Koordinatensystem?
- Welcher Koordinatenursprung für das Gitter?
- Finden im Zuge der Bereitstellung weitere Koordinatentransformationen statt?

- Resultierende Genauigkeit ist i.A. deutlich schlechter als 1km

Datenarten

Daten aus Beprobungen – **Standortinformationen**

Standortverschleierung – Nachträgliche Ergänzung von Angaben zur Standortbeschaffenheit und -umgebung

- Bsp: Standort verschleiert, aber Angaben zur Landbedeckung (Corine Landcover, CLC) vorhanden
- Welcher CLC-Datensatz wurde genutzt? (CLC 2000, 2006, 2012, 2018, …)
- Analyse erfordert ggf. einen anderen CLC-Datensatz als Grundlage
- CLC-Datensatz passt, aber Analyse benötigt den NDVI zum Zeitpunkt der Datenaufnahme …

Räumliche Auflösungen

STO (verschleiert): **1km**

CLC: **100m**

NDVI: **10m**

Datenarten

Daten aus Beprobungen – **Standortinformationen**

Standortverschleierung – Fazit

- Nutzung verschleierter Daten führt zu Addition von Unsicherheiten
- Effektive Lageunsicherheit i.d.R. schlechter als vom Datenbereitsteller ursprünglich gewollt
- „Workarounds“ für den Umgang mit verschleierten Daten technisch machbar aber:
  - Methodisch anspruchsvoll, teilw. anspruchsvoller als die Analyse selbst
  - In der Umsetzung teuer
  - Analyseergebnisse aufgrund der Unsicherheiten leichter in Zweifel zu ziehen

Datenarten

Daten aus Beprobungen – **Standortinformationen**

Standortverschleierung – vielleicht auch direkt im DataHub?

- DataHub enthält exakte Standortinformationen (zumindest für neure Daten)
- Verschleierung / Generalisierung von Standortinformationen für Öffentlichkeit / weniger privilegierte Nutzer wird nach einheitlichem Verfahren durch den DataHub bewerkstelligt
- Nachträgliche Ergänzung von Standortinformationen kann im DHB stattfinden und erfordert keine nachträgliche Zusatzarbeit beim Datenbereitsteller

- Zeitersparnis, höhere Datenqualität, geringere Kosten, schnellere Ergebnisse – erfordert aber auch mehr Vertrauen! (und wird fixiert in entsprechenden Vereinbarungen)