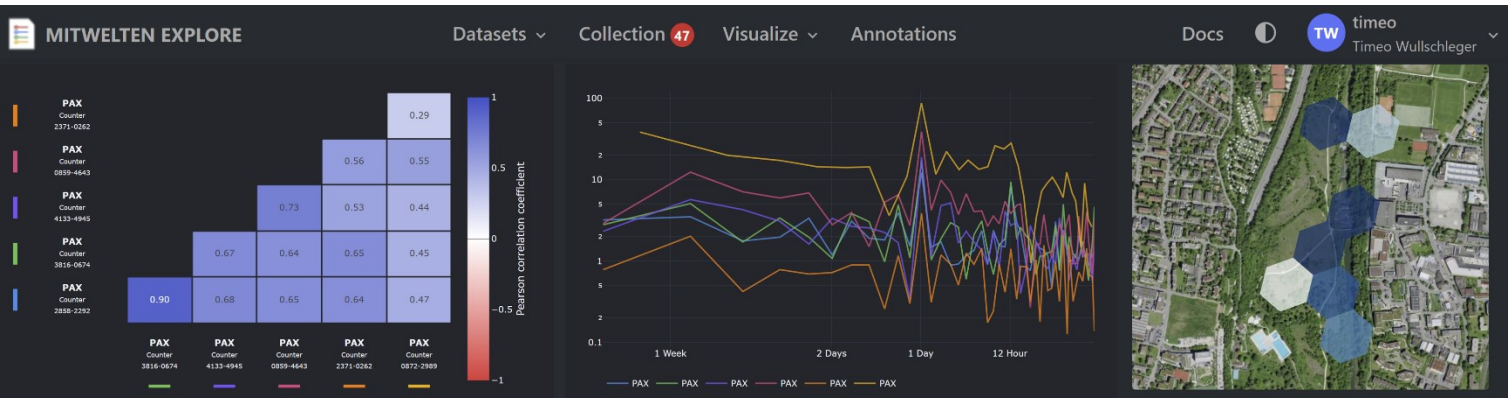


## Projektdokumentation P9

### Visual Exploration for Urban Biodiversity Monitoring



Autor

Timeo Wullschleger  
timeo.wullschleger@fhnw.ch

Betreuung

Prof. Thomas Amberg

Experte

Oleg Lavrovsky

Ort, Datum

Windisch, 08.09.2023

## Abstract

In an increasingly urbanizing world, the preservation of biodiversity in urban habitats is becoming more and more crucial. This Project is focused on the analysis of biodiversity monitoring data to provide a better understanding of it. In the research project Mitwelten, automated data collection systems were used to collect and analyze various datasets, including observations of birds, pollinators, and human presence. The data that has been collected continuously over long periods provides an interesting basis for the exploration of biodiversity. For the interpretation of the datasets, the challenge is to find the right view of the data to identify important phenomena and correlations.

The goal of this work was to develop an interactive application, called "Mitwelten Explore," to visually explore datasets, compare them to each other, and identify patterns, trends, and correlations. In the first step, the fundamentals of UX/UI design, data visualization, and data management were elaborated. In the second phase, it was analyzed how the existing data can be filtered to explore and compare subsets of it. For the investigation of further variables influencing biodiversity, additional data from external sources, including meteorological data, were obtained. To operate the application, a clear workflow was developed, which is comparable to the operation of an online store. To enable researchers to share their findings with others, a concept for writing annotations was developed. In the final phase of the project, the application was implemented using an iterative development principle. In addition to the Explore tool, the data collected in the Mitwelten project was made available via an API to enable further analysis with third-party tools.

The developed application provides an efficient and user-friendly solution for the exploration and analysis of ecologically relevant data and contributes to an improved understanding of biodiversity in urban terrains.

## Inhaltsverzeichnis

1	Einleitung	5
1.1	Ausgangslage	5
1.2	Aufgabenstellung	5
1.3	Strukturierung der Dokumentation	5
2	Theoretische Grundlagen	6
2.1	Biodiversität	6
2.2	UX / UI Design	7
2.3	Datenvisualisierung	8
2.4	Datenmanagement	15
2.5	Access Management	16
3	Analyse	18
3.1	Anforderungen an das Tool	18
3.2	Daten	19
3.3	Datensets Explore Applikation	25
3.4	Bedienstruktur und Workflow der Applikation	26
3.5	Dashboard Layout	27
3.6	Datenmanagement	30
3.7	Datenanalyse	32
3.8	Plotly Dash	34
4	Implementierung	38
4.1	Architektur	38
4.2	Navigation	40
4.3	Datenauswahl	40
4.4	Collection	46
4.5	Dashboards	47
4.6	Annotationen	61
4.7	Anleitung und Dokumentation im Tool	64
4.8	Landing Page	65
4.9	Datenvisualisierung	68
4.10	User Management	72
4.11	Datenmanagement	73
4.12	Datenbezug für externe Tools	77
4.13	Mögliche Erweiterungen des Tools	78
5	Fazit	79
6	Literaturverzeichnis	80
7	Abbildungsverzeichnis	83

8	Tabellenverzeichnis	85
9	Codeausschnittverzeichnis	85
10	Ehrlichkeitserklärung	86
11	Anhang	87
11.1	Repositories	87
11.2	Datenbezug Mitwelten Data API	87



# 1 Einleitung

## 1.1 Ausgangslage

Das interdisziplinäre SNF Forschungsprojekt Mitwelten befasst sich mit dem Monitoring der Biodiversität auf urbanem Terrain. Für das Monitoring wurden verschiedene Datenerfassungssysteme entwickelt und in drei Fallstudien im Raum Basel eingesetzt.

Das Bird-Datenset repräsentiert die akustische Aktivität von Vogelarten, die anhand ihrer Rufe erkannt wurden. Audiologger an 36 verschiedenen Orten produzierten Aufnahmen mit einer Gesamtdauer von 22'500 Stunden. Für die Auswertung der Aufnahmen wurde ein Machine Learning Algorithmus eingesetzt. Knapp eine Million Rufe von 222 unterschiedlichen Arten wurden mit einer Confidence grösser als 0.6 erkannt. In der Pollinator Studie wurde die Aktivität von Bestäubern erhoben. An 55 Standorten wurden knapp 5 Millionen Bilder automatisiert aufgenommen und mit einer Machine Learning Pipeline ausgewertet. Über 122'000 Bestäuber wurden erkannt und klassifiziert. Eine Detektion von Smartphones wurde eingesetzt, um die Anwesenheit von Menschen an bestimmten Orten zu erheben.

Die über lange Zeitperioden kontinuierlich erfassten Daten bilden eine wertvolle Grundlage für die Erkundung der Biodiversität. Die Herausforderung besteht darin, die richtige Sicht auf die Daten zu entdecken, um wichtige Ereignisse und Zusammenhänge zu erkennen.

## 1.2 Aufgabenstellung

Im Rahmen dieser Projektarbeit soll analysiert werden, wie ein Tool die Exploration von ökologisch relevanten Daten unterstützen kann, um Aussagen über mögliche Einflüsse auf die Biodiversität zu machen. Konkret soll eine interaktive Applikation, «Mitwelten Explore», entwickelt werden, um Teilmengen der Datensets visuell zu erkunden.

In einem ersten Schritt sollen die Grundlagen der Bereiche UX/UI Design, Datenmanagement und Datenvisualisierung recherchiert und dokumentiert werden. Anschliessend sollen Konzepte für das Datenmanagement und die Bedienstruktur der Applikation entworfen werden. In Form einer Recherche sollen zusätzliche und ergänzende Datenquellen gesucht werden. Schlussendlich soll die Applikation in iterativen Schritten implementiert werden.

## 1.3 Strukturierung der Dokumentation

Der Hauptteil dieser Dokumentation ist in drei Abschnitte gegliedert. Im ersten Abschnitt werden die im Rahmen dieser Projektarbeit erarbeiteten Grundlagen dokumentiert. Die konkreten Anforderungen an die Applikation, erarbeitete Konzepte und Informationen zu den erhobenen Daten sind im zweiten Abschnitt dokumentiert. Im dritten Abschnitt wird die Implementierung der Applikation und der abhängigen Systemkomponenten dokumentiert.

## 2 Theoretische Grundlagen

In diesem Kapitel werden relevante Grundlagen dokumentiert, die im Rahmen der Projektarbeit erarbeitet wurden.

### 2.1 Biodiversität

Die biologische Diversität (Biodiversität) steht im Zentrum dieser Arbeit. Sie umfasst folgende Ebenen:

- Die genetische Diversität (Vielfalt innerhalb einer Art)
- Die Artenvielfalt oder Artendiversität (Vielfalt zwischen den Arten)
- Die Diversität von Lebensräumen

Im Rahmen des Projekts Mitwelten liegt der Fokus auf der Artenvielfalt. Diese steht in direkter Abhängigkeit von der Vielfalt der Lebensräume.

Als Artendiversität wird die Anzahl verschiedener Arten bezeichnet, die aktiv in einer spezifischen Umgebung präsent sind. Die Kategorisierung von Lebewesen in Arten wurde ursprünglich nach dem morphologischen Konzept durchgeführt, wobei Farbe, Form und Gestalt der Lebewesen als Kriterien dienten. Später setzte sich das Konzept der Biospezies durch, dass die Paarungsfähigkeit von Individuen miteinbezieht. Für die Erhaltung der Artendiversität spielt neben der Anzahl (Quantität) verschiedener Arten die Verteilung und die Eigenschaften der Arten (Qualität) eine wichtige Rolle.

Ein Ökosystem besteht aus einem Lebensraum (Biotop) und Lebensgemeinschaften (Biozönosen), wobei die Existenz der Biozönosen von den entsprechenden Lebensräumen abhängig ist. Innerhalb eines Ökosystems laufen selbstregulierende Prozesse, wie Stoffkreisläufe, die aus biotischen (lebenden) und abiotischen (nicht lebenden) Komponenten bestehen. Zur Aufrechterhaltung dieser Prozesse ist die Anwesenheit von bestimmten Arten erforderlich, die spezifische Funktionen erfüllen. Die Bedeutung der An- oder Abwesenheit einer spezifischen Art in einem Lebensraum ist abhängig von deren Funktion im Ökosystem. Die Anzahl Individuen, die für die Erfüllung einer artenspezifischen Funktion erforderlich ist, variiert nach Funktion und Art.

Die Biodiversität wird von einer Vielzahl von Faktoren beeinflusst. Dazu gehören Prozesse natürlichen Ursprungs, wie Klimaveränderungen, geologische Ereignisse oder Naturkatastrophen. Verändern diese Faktoren einen Lebensraum, so wird die Artendiversität direkt beeinflusst.

Der Einfluss von menschlichen Aktivitäten spielt jedoch eine immer größere Rolle. Die Zerstörung von Lebensräumen durch Änderungen der Landnutzung, die Abholzung von Wäldern und die Urbanisierung reduziert die verfügbaren Lebensräume für viele Arten. Die Einführung von nicht heimischen invasiven Arten kann die einheimische Fauna und Flora verdrängen und die natürlichen Gleichgewichte stören.

Die Anpassungsfähigkeit von Arten ist von entscheidender Bedeutung im Bezug auf die Veränderung von Einflussfaktoren. Sie bezieht sich auf die Fähigkeit einer Art, sich an neue Umweltbedingungen anzupassen und ist der entscheidende Faktor für deren Überleben und den langfristigen Fortbestand.

Durch die vielfältigen Umweltbedingungen, die sich durch menschliche Aktionen regelmässig und in kurzer Zeit ändern sind urbane Gebiete besonders interessant für die Untersuchung der Anpassungsfähigkeit von Arten.

## 2.2 UX / UI Design

In diesem Unterkapitel werden relevante UX / UI Grundlagen und Richtlinien beschrieben. Die beschriebenen Grundlagen stammen neben den direkt referenzierten Quellen aus einem Coursera Kurs von Meta [1] und aus weiterer Fachliteratur [2].

### 2.2.1 Seitenlayout

Mehrere Studien [3] [4] [5] ergaben, dass Menschen Web Content in Mustern absuchen und konsumieren. Bei Seiten mit wenig Inhalt wurde ein Z-Pattern, welches in Abbildung 1 links visualisiert ist, beobachtet. Der Blick der Testpersonen bewegte sich zeilenweise von der oberen linken Ecke bis zu der unteren rechten Ecke über den gesamten Content. Bei informationsreichen Webseiten wurde ein F-Pattern beobachtet. Die Spur des Blickfelds startet, wie beim Z-Pattern, oben links. Der Content wird anschliessend in einem F-förmigen Muster nach relevanten Mustern abgesucht [6]. Abbildung 1 zeigt auf der rechten Seite eine Visualisierung des F-Pattern.

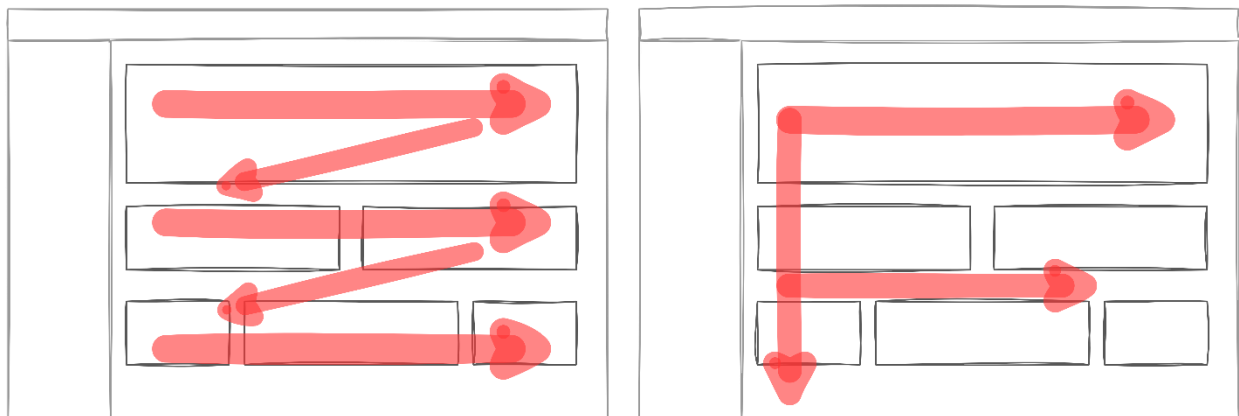


Abbildung 1 Z- und F-Pattern

Bei der Entwicklung einer Webanwendung ist es sinnvoll, die Strukturierung des Inhalts so zu gestalten, dass relevante Informationen intuitiv konsumiert werden. Mit Typografie, Farben und Formen kann der Blick geleitet werden.

### 2.2.2 Rückmeldungen und Systemstatus

Personen, die ein System nutzen, sollen zu jedem Zeitpunkt über den aktuellen Status informiert sein. Mit Rückmeldungen können Aktionen bestätigt werden. Länger andauernde Prozesse wie Datenabfragen im Hintergrund sollen klar kommuniziert werden. Durch eine klare Kommunikation der Abläufe werden die Benutzer und Benutzerinnen mit einem System vertraut und verstehen Abläufe schneller [7].

Neben dem Status soll kommuniziert werden, wo sich die betrachtende Person im Informationsraum befinden. In vielen Fällen wird dafür eine Navigationsleiste mit einem hervorgehobenen Label eingesetzt. Nicht alle Nutzenden gelangen über die Startseite auf die Applikation. Eine klare Kommunikation des aktuellen Standortes ist relevant für die Orientierung und die Navigation [8].

### 2.2.3 Konsistenz und Standards

Anwendungen sollen konsistent sein und einen standardisierten Aufbau aufweisen. Es wird in interne und externe Konsistenz unterschieden. Die interne Konsistenz beschreibt die Wiederverwendung von definierten Strukturen innerhalb einer Anwendung. Solche Strukturen beinhalten die Anordnung von Schaltflächen oder die Schritte zur Ausführung einer Aktion.

Die externe Konsistenz referenziert auf etablierte Konventionen im Web, die über die Anwendung hinausgehen. Jakob's Law of Internet User Experience [9] besagt, dass Nutzende den grössten

Teil ihrer Zeit auf anderen Webseiten verbringen und es bevorzugen, wenn eine Applikation gemäss bekannten Mustern und Abläufen funktioniert. Ein Beispiel von Standard Design ist ein Logo in der Navigationsliste, das auf die Startseite einer Webanwendung verlinkt ist [10].

#### 2.2.4 Notausgang

Das versehentliche Ausführen von Aktionen passiert vielen Nutzern und Nutzerinnen. In jeder Situation soll eine gut signalisierte Möglichkeit (*Notausgang*) bestehen, um Aktionen abubrechen oder rückgängig zu machen. Bei kritischen Aktionen, die irreversibel sind, soll eine zusätzliche Bestätigung der Aktion implementiert werden [11].

#### 2.2.5 Hilfe und Dokumentation

Ein gutes Tool soll intuitiv bedienbar sein. Trotzdem ist es wichtig, die Funktionen zu dokumentieren, damit die nutzenden Personen eine Applikation kennenlernen und verstehen. Anleitungen und Hilfe in Applikationen können proaktiv und reaktiv gestaltet werden. Proaktive Hilfe sind Meldungen oder Tipps, die angezeigt werden, um zu verhindern, dass ein Problem auftritt. Beispiele sind Onboarding Tutorials oder Tipps zum aktuellen Kontext.

Statische Dokumentationen oder Tutorials sind reaktiv. Sie werden aktiv aufgerufen, wenn Komplikationen auftreten oder wenn die Funktionalität erlernt wird. Dokumentationen und Anleitungen sollen umfassend und detailliert sein und unterstützen, Probleme zu identifizieren und zu beheben. Unterteilungen in Kategorien und klare Strukturen sind nötig für eine gute Orientierung [12].

#### 2.2.6 Ästhetik und Minimalismus im Design

User Interfaces sollen ästhetisch ansprechend und minimalistisch gestaltet werden. Die Bedienoberfläche soll einfach und leicht zu verstehen sein, ohne unnötige Elemente oder Ablenkungen. Die Ästhetik soll den Nutzenden eine positive Erfahrung bieten, woraus deren Zufriedenheit positiv beeinflusst wird. Sie wird mit konsistenten Designs, ansprechenden Grafiken und klaren Schriftarten erhöht.

Minimalismus wird durch die Geringhaltung der Anzahl Elemente ohne relevanten Informationsgehalt (Noise) angestrebt. Elemente mit hohem Informationsgehalt kommen somit besser zur Geltung [13].

### 2.3 Datenvisualisierung

Als Datenvisualisierung wird die visuelle Präsentation von Informationen und Datensätzen bezeichnet, die zum Ziel hat, Eigenschaften wie Trends oder Muster oder Ausreisser der darunterliegenden Daten zu kommunizieren. Mit unterschiedlichen Methoden werden verschiedene Achsen von Datensätzen hervorgehoben. Dieses Unterkapitel behandelt die Grundsätze der Datenvisualisierung und beschreibt relevante Methoden zur Darstellung von unterschiedlichen Datenstrukturen. Die folgend dokumentierten Grundlagen der stammen unter Anderem aus dem Buch *Better Data Visualizations* [14].

#### 2.3.1 Grundsätze der Datenvisualisierung

Datenvisualisierungen dienen unterschiedlichen Zwecken. In den meisten Fällen werden sie eingesetzt, um Daten zu erklären. In der Literatur wird oft die Bezeichnung 'Explanatory Data Visualization' für diesen Einsatz verwendet. Erklärende Visualisierungen dienen zur Kommunikation von Erkenntnissen wie Muster, Trends oder Ausreisser, die durch Daten belegt werden.

Die explorative Datenvisualisierung wird eingesetzt, um Erkenntnisse aus Daten zu gewinnen. Unterschiedliche Darstellungsmethoden zeigen verschiedene Sichtweisen auf die Daten um Trends, Muster und Ausreisser zu erkennen. Die Visualisierungen in dieser Projektarbeit sind explorative Darstellungen. Die folgenden Punkte sollten bei allen Visualisierungen beachtet werden [14].

## Data-Ink Ratio

Die Data-Ink Ratio ist ein von Edward Tufte entwickeltes Mass für die Effizienz einer Darstellung. Data Ink bezeichnet die Tinte, die zum Visualisieren der Daten erforderlich ist. Das Verhältnis der Data Ink zur Gesamtmenge der Tinte in der Visualisierung wird als Data-Ink Ratio bezeichnet. Für effiziente Visualisierungen soll die Ratio maximiert werden.

## Chart Junk

Chart Junk ist eine Bezeichnung für verzichtbare Elemente in einer Visualisierung. Mit dekorativen Elementen, die nicht notwendig sind, wird von den Daten abgelenkt. Chart Junk ist das Gegenteil von Data Ink und soll so gering wie möglich gehalten werden.

## Redundanz

In einem Diagramm kann es verschiedene Elemente geben, die dieselbe Information vermitteln. Elemente, die der Visualisierung keine einzigartigen Informationen hinzufügen, sollten entfernt werden. Redundante Informationen werden oft mit Legenden oder Achsenbeschriftungen verursacht.

### 2.3.2 Visualisierung von Zeitreihendaten

Zeitreihendaten sind zweidimensionale Datensätze, wobei die erste Dimension die Zeit darstellt. Visualisierung von Zeitreihendaten sollen dazu dienen, Veränderungen eines Wertes über die Zeit zu kommunizieren. Die meistverwendeten Visualisierungsarten sind folgend beschrieben.

#### Scatter Plots

Die simpelste Visualisierungsform von Zeitreihendaten sind Scatter Plots. Jeder Datenpunkt wird von einem Marker im Chart repräsentiert.

#### Line Plots

Ein Line Plot ist eine erweiterte Form des Scatter Plots. Die Marker der einzelnen Datenpunkte werden mit Linien verbunden. Line Plots sollten nur dann eingesetzt werden, wenn davon auszugehen ist, dass die Werte interpolierbar sind. Interpolierbare Datensätze sind zum Beispiel die Temperatur oder die Luftfeuchtigkeit. Detektionen von Spezies pro Tag sind nicht interpolierbar, da es sich um absolute Werte handelt. Als Minimum der Y-Achse wird in den meisten Fällen der niedrigste Y Wert definiert, um relative Änderungen klar darzustellen.

#### Area Plots

Die Grundlage von einem Area Plot ist ein Line Plot. Der Bereich unter der Linie wird bis zum Nullpunkt der Y-Achse eingefärbt. Die Y-Achse startet somit beim Nullpunkt, womit die Absoluten Werte zur Geltung kommen. Wie bei den Line Plots sollten nur interpolierbare Datensätze mit Area Plots dargestellt werden.

#### Bar Plots

Bar Plots eignen sich für die Darstellung von nicht interpolierbaren Datensätzen, deren Absolutwerte relevant sind. Ein Beispiel dafür ist die Niederschlagssumme pro Tag. Jeder Datenpunkt wird durch einen Balken repräsentiert. Die Balken beginnen jeweils beim Nullpunkt der Y-Achse. Abbildung 2 zeigt die beschriebenen Visualisierungsformen an Beispielen.

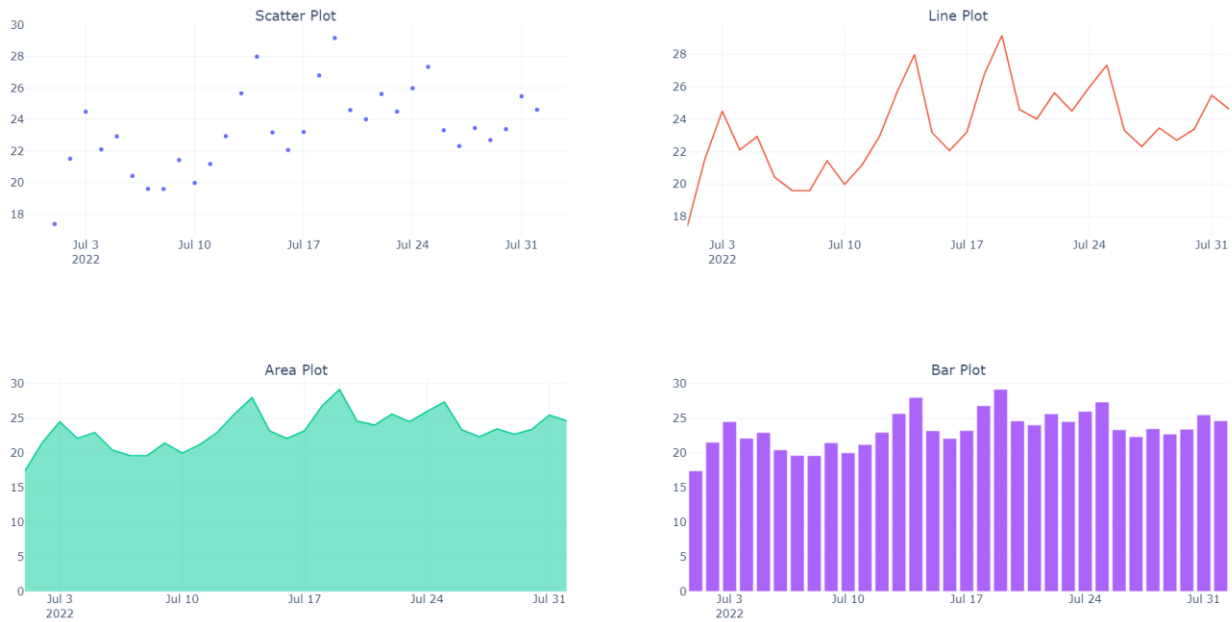


Abbildung 2 Visualisierung von Zeitreihendaten

### 2.3.3 Visualisierung von räumlichen Daten

Für die Visualisierung von räumlichen Datensets, zum Beispiel eine Liste von Orten, an welchen Beobachtungen von Spezies gemacht wurden, werden oft Karten eingesetzt. Aus einer Karte können neben den visualisierten Datenpunkten weitere Informationen über die Umgebung gelesen werden.

#### Scatter Maps

Die einfachste Visualisierungsform für räumliche Daten sind Scatter Maps, wobei für jeden Punkt im Datensatz ein Marker auf eine Karte gezeichnet wird. Abbildung 3 zeigt einen Scatter Map Plot mit 120 visualisierten Datenpunkten. Die grobe Verteilung der Punkte lässt sich gut aus der Visualisierung lesen. Aussagen über die Konzentration von Punkten an bestimmten Orten hingegen können nur nach dem Hineinzoomen in die Regionen gemacht werden.



Abbildung 3 Scatter Map Plot

## Density Heatmaps

Oft werden Density Heatmaps für die Visualisierung der Dichte von Datenpunkten eingesetzt. Abbildung 4 zeigt die selben Datenpunkte wie in Abbildung 3 in Form einer Heatmap. Auf der Visualisierung kann man erkennen, dass die Region in der Mitte der Karte über eine sehr hohe Punktedichte verfügt.

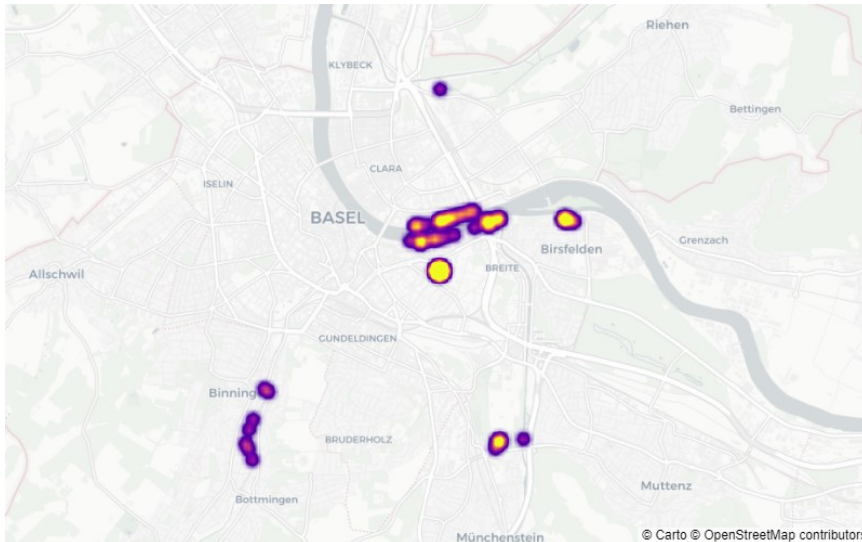


Abbildung 4 Visualisierung der Punktedichte mittels Heatmap

Der Einsatz von Heatmaps zur Visualisierung von räumlichen Daten bringt mehrere Nachteile mit sich:

- Einzelne Punkte werden nicht an ihrer genauen Position eingezeichnet, was zu einem Informationsverlust führt.
- Ausreisser werden durch die Aggregation nicht deutlich dargestellt.
- Die räumliche Wahrnehmung wird durch die Farbverläufe verzerrt. Bestimmte Regionen können die Aufmerksamkeit stark auf sich ziehen, wobei andere Regionen weniger betont werden.

## Hexbinmaps

Als Alternative zu Density Heatmaps eignen sich Hexbinmaps zur Visualisierung der Punktdichte in bestimmten Bereichen. Für die Visualisierung wird die Oberfläche der Karte in Hexagon förmige Zellen, auch Bins genannt, aufgeteilt. Pro Zelle werden die sich darin befindenden Punkte aggregiert.

Der Einsatz von Hexagons als Form hat folgende Vorteile gegenüber anderen Formen:

- Sie ermöglichen eine gleichmässige Abdeckung.
- Die kompakte Anordnung ohne Überlappungen von Clustern oder Lücken zwischen Zellen nutzt den Platz effizient.
- Robust gegen Verzerrungen, die durch Kartenprojektionen entstehen.
- Gute Erkennbarkeit von Mustern und Trends durch die Kontinuität.

Die Aggregation kann der Anzahl Datenpunkte in einer Zelle entsprechen oder aus darunterliegenden Messwerten berechnet werden. Die Zellen können entsprechend des berechneten Wertes eingefärbt werden, um die Werte zu kommunizieren. Abbildung 5 zeigt die selben Datenpunkte wie in Abbildung 3 und Abbildung 4. Die originalen Datenpunkte sind in blauer Farbe visualisiert. Durch die Einfärbung ist erkennbar, dass die Anzahl Punkte in einer Zelle signifikant höher ist als die anderen.

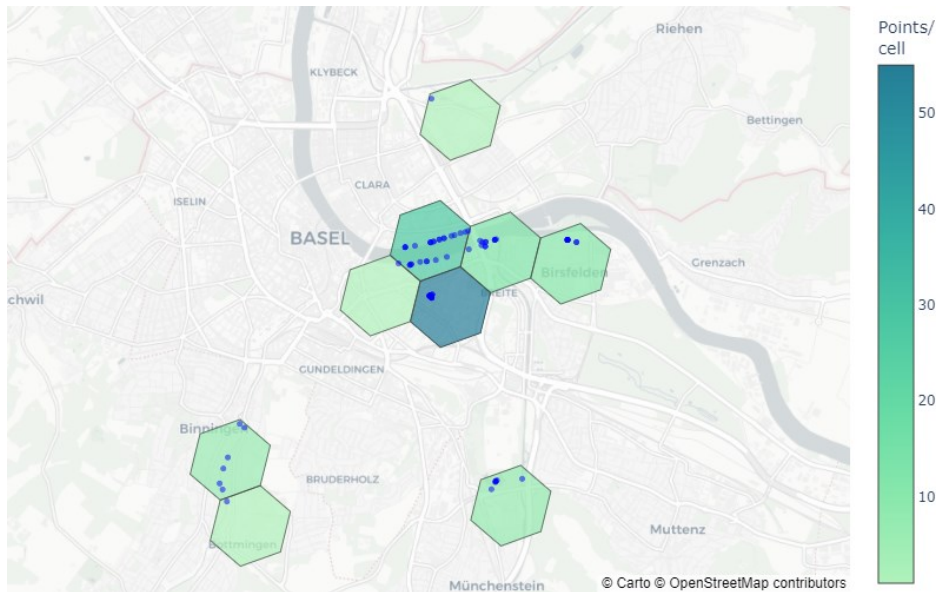


Abbildung 5 Hexbinmap: Punkte pro Zelle

Im Gegensatz zu Scatter Maps muss bei Hexbinmaps jede Ansicht berechnet werden. Dynamische Hexbinmaps berechnen den Layer für die Zoom Stufen der Karte jeweils neu. Abbildung 6 zeigt einen Ausschnitt aus Abbildung 5, wobei in das Hexagon mit den meisten Datenpunkten gezoomt wurde. Durch das Zoomen wurden neue Zellen und eine neue Farbskala berechnet.

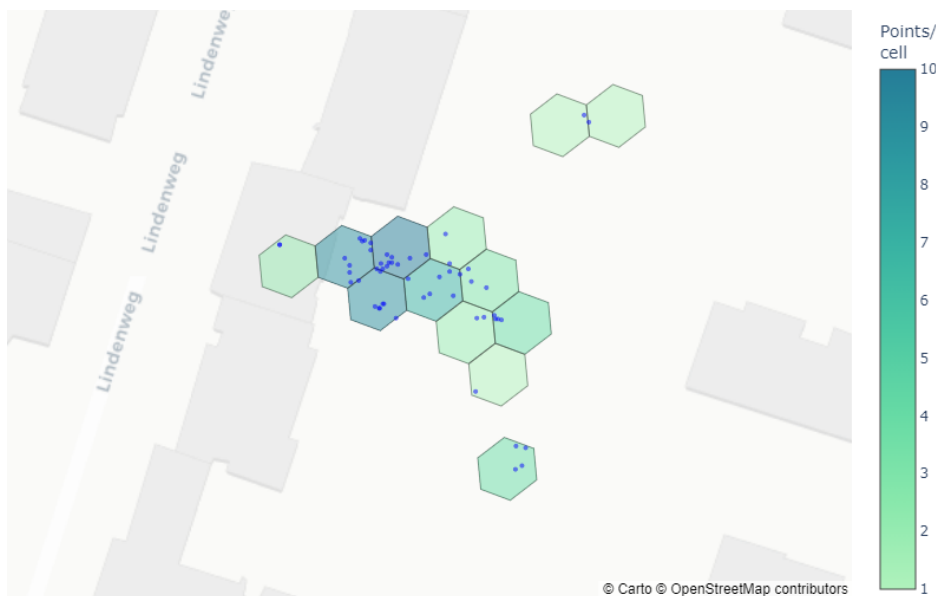


Abbildung 6 Hexbinmap: Punkte pro Zelle, Zoom

## Bubble Map

Bei multidimensionalen Datensets, die neben dem Standort einen weiteren Wert, wie ein Temperaturwert oder eine Anzahl detektierter Spezies beinhalten, können mit Bubble Maps dargestellt werden. Pro Datenpunkt wird ein Marker auf den Kartenlayer gezeichnet. Die Grösse der Datenpunkte sind proportional zu den darunterliegenden Werten. In der Regel repräsentiert die Oberfläche der Marker den entsprechenden Messwert. Abbildung 7 zeigt eine Bubble Map.



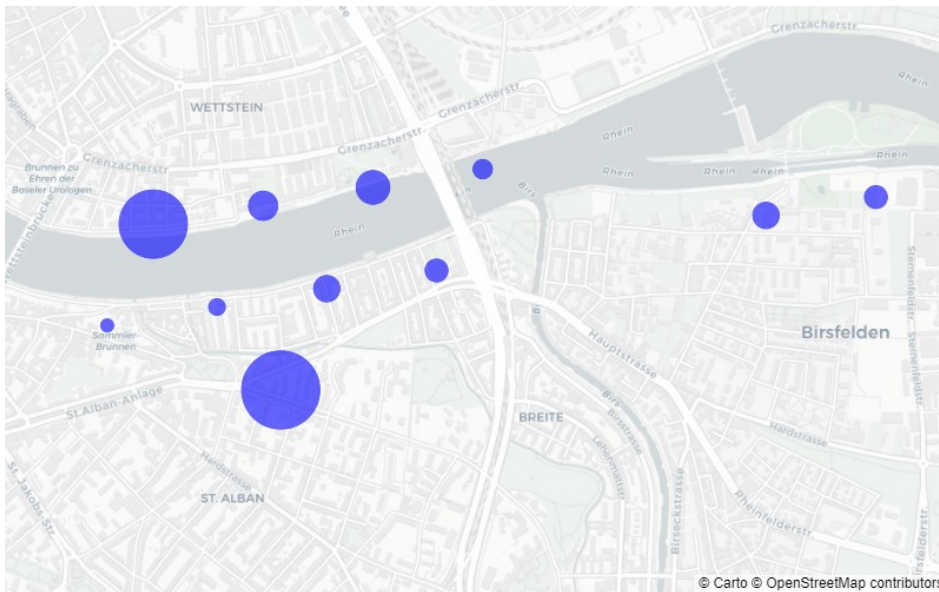


Abbildung 7 Bubble Map

Ein Nachteil von Bubble Maps sind Überlappungen von Markern. Durch Anpassen der Opazität werden verdeckte Punkte sichtbar. Damit die Marker weder zu gross noch zu klein werden, müssen die Werte normalisiert und mit einem Skalierungsfaktor errechnet werden.

### 2.3.4 Visualisierung von Proportionen

Es gibt verschiedene Methoden zur Visualisierung der Zusammensetzung eines kategorischen Datensets. Ein verbreitetes Beispiel ist die Darstellung von Männer- und Frauenanteil in einer Personengruppe oder Wähleranteile von politischen Parteien.

Abbildung 8 zeigt drei häufig eingesetzte Darstellungsformen zur Kommunikation der Zusammensetzung eines Datensatzes. Die Daten stammen aus dem Bird-Datenset und zeigen die Anzahl erkannter Vogelstimmen pro Art an einem spezifischen Ort. Weniger oft detektierte Arten sind in der Kategorie 'others' zusammengefasst.

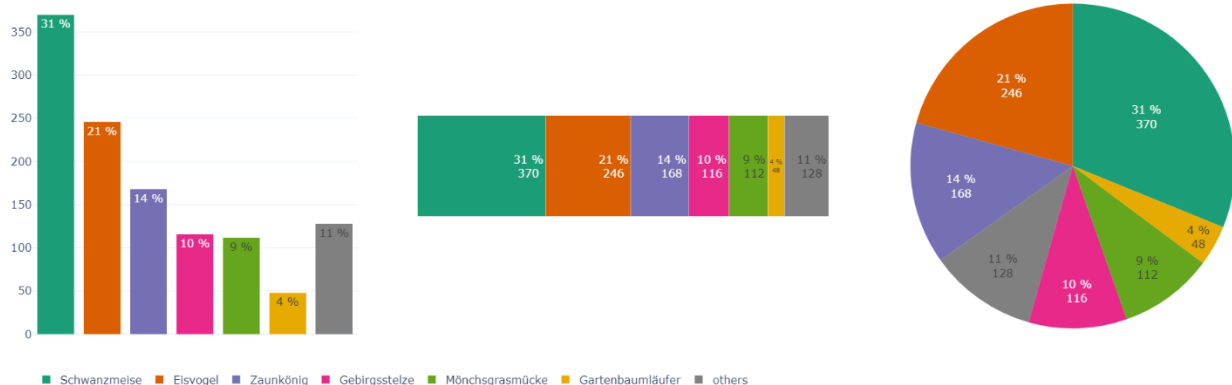


Abbildung 8 Visualisierung von Proportionen - Aktivste Vogelarten Deployment 679

In dem Bar Chart sind die Proportionen der Kategorien gut erkennbar. Auch kleinere Unterschiede, wie der Unterschied von 11 % und 10 % in Abbildung 8 sind klar ersichtlich. Der Nachteil dieser Methode liegt darin, dass es nicht klar ist, dass die Daten Anteile einer Gesamtmenge darstellen. Der Einsatz eines Stacked-Bar Chart kommuniziert diese Eigenschaft besser. Die Proportionen der Kategorien sind jedoch weniger gut ersichtlich. Sowohl horizontale als auch vertikale Ausrichtungen sind üblich. Bei sehr vielen Kategorien mit kleinen Werten kann diese Darstellungsmethode unübersichtlich wirken. Der Pie Chart ist die am häufigsten eingesetzte

Visualisierungsmethode. Durch die radiale Anordnung sind Proportionen in gewissen Fällen sehr schwer erkennbar, wodurch die Darstellung in der Wissenschaft häufiger kritisiert wird. Dass der Chart die Zusammensetzung einer Gesamtmenge darstellt, ist eindeutig. Ein Vorteil gegenüber den anderen Visualisierungsformen liegt darin, dass einfache Bruchteile wie  $\frac{1}{2}$ ,  $\frac{1}{3}$  oder  $\frac{1}{4}$  visuell betont werden [15].

Welche Visualisierungsform eingesetzt werden sollt, hängt in jedem Fall von den Daten und der damit verbundenen Aussage ab. Während Bar Charts die einzelnen Werte klar darstellen, kommuniziert das Kuchendiagramm klar, was dargestellt wird. Der Einsatz von Beschriftungen ermöglicht die Kommunikation von exakten Werten in jeder Darstellung.

### 2.3.5 Interaktive Visualisierungen

Interaktive Visualisierungen ermöglichen dynamische Manipulationen des angezeigten Inhalts. Typische Interaktionen sind:

- Hinein- oder Hinauszoomen
- Skalieren der Achsen
- Ein- oder Ausblenden von Elementen
- Anzeigen von weiteren Informationen bei Hover über einem Element
- Selektieren von Datenpunkten

Durch den Einsatz von interaktiven Elementen kann sich die bedienende Person aktiv mit den darunterliegenden Werten beschäftigen und erhält mehr Informationen als mit einer statischen Visualisierung.

### 2.3.6 Libraries zur Datenvisualisierung

Es gibt viele verschiedene Libraries für die Visualisierung von Daten. High-Level- Libraries bieten eine höhere Abstraktionsebene und sind einfacher zu bedienen, während Low-Level-Libraries mehr Kontrolle über die Ausgabe und die Konfiguration bieten. JavaScript, Python und R sind die am besten unterstützten Sprachen, da die Datenvisualisierung grösstenteils im Web oder im Bereich der Data Science Anwendung findet. Die populärsten Libraries für Python sind Matplotlib [16], Plotly [17] und Seaborn [18], wobei Plotly die grösste Flexibilität bietet [19].

Plotly.py ist aufgebaut auf Plotly.js [20], der Implementierung für JavaScript. Plotly.js basiert auf d3.js und stack.gl. Eine Plotly Figure besteht aus der data Komponente und dem layout. Der modulare Aufbau der Grafik erlaubt es, mehrere Datensets mit unterschiedlichen Visualisierungsformen zu plotten. Mit der Klasse `plotly.graph_objects` werden Grafiken konstruiert, manipuliert und serialisiert für das Rendering mit Plotly.js. Sie stellt das Low-Level Interface zu Plotly dar. Die High-Level Klasse `plotly.express` beinhaltet Konstruktoren für die gängigsten Visualisierungstypen und ermöglicht die Generation von Grafiken in einer Linie Code. Im Gegensatz zu Matplotlib und Seaborn sind Plotly Figures standardmässig interaktiv. Zur Steigerung der Effizienz bei der Visualisierung von grossen Datenmengen wird WebGL für das Rendering eingesetzt, das die Berechnungen auf der GPU ausführt [21].

## 2.4 Datenmanagement

Die Speicherung, Organisation, Verwaltung und die Bereitstellung von Daten wird unter dem Begriff Datenmanagement zusammengefasst. Für die Speicherung von grösseren Datenmengen werden Datenbanksysteme eingesetzt.

### 2.4.1 Speicherung und Organisation von Daten

Daten lassen sich in verschiedene Kategorien einteilen, die im Datenmanagement unterschiedlich zu behandeln sind. Oft werden strukturierte, semi-strukturierte und unstrukturierte Daten unterschieden [22].

**Strukturierte** Daten sind organisiert in einem klar definierten Format. Sie können einfach in Tabellen oder andere Datenstrukturen gespeichert und verarbeitet werden. Beispiele aus dieser Kategorie sind Listen von beobachteten Spezies oder Messwerte von IoT Systemen. Aufgrund der festgelegten Strukturen sind die Daten am einfachsten zu verarbeiten. In den meisten Fällen kommen relationale Datenbankmanagementsysteme (DBMS) für die Speicherung zum Einsatz. In relationalen Datenbanken werden die Daten in einer oder in mehreren Tabellen gespeichert, wobei alle Zellen in einer Spalte den selben Datentyp haben.

Zeitreihendaten sind zeitlich geordnete Datenpunkte, zum Beispiel periodisch gemessene Temperaturwerte. Sie gehören zu den strukturierten Daten und haben die Zeit als zentrales Merkmal. Für die Speicherung kommen neben relationalen DBMS vermehrt Zeitreihendatenbanken zum Einsatz. Zeitreihendatenbanken sind optimiert für zeitbasierte Operationen, wie die Berechnung von täglichen Maximalwerten oder von stündlichen Summen. Es gibt Erweiterungen für relationale DBMS, die gewisse Tabellen umstrukturieren, um den Inhalt als Zeitreihendaten bereitzustellen und zeitbasierte Operationen zu ermöglichen.

**Semistrukturierte** Daten haben eine gewisse Teilstruktur, unterliegen insgesamt aber keiner streng definierten Strukturierung. Beispiele für semistrukturierte Daten sind Tweets, die anhand der Hashtags organisiert werden, oder HTML Files von Webseiten. Zur Speicherung von dieser Datenkategorie werden oft NoSQL Datenbanken eingesetzt. Es existieren relationale Datenbanken, die Datentypen zur Speicherung von unstrukturierten Komponenten der Daten bereitstellen. In PostgreSQL ist es möglich, JSON Objekte ohne festgelegte Struktur zu speichern [23].

**Unstrukturierte** Daten unterliegen keiner definierten Struktur. Bilder, Ton- und Videoaufnahmen und PDF Dokumente sind Beispiele aus dieser Kategorie. Die Verarbeitung von Daten aus dieser Kategorie ist am schwierigsten, da für die Extraktion von Informationen zusätzliche Technologien und Verfahren gefordert sind. In vielen Anwendungsfällen werden unstrukturierte Daten im binären Format in Objektspeicher gespeichert. Ein Objektspeicher ist ein Speicherdienst, der Daten zusammen mit Metadaten als Objekte speichert. Jedes Objekt wird durch den Objektnamen identifiziert und kann über die entsprechende Schnittstelle heruntergeladen werden. Funktionen für die Zugriffsregelung, das Erstellen von Backups und eine automatische Skalierung des Speichers werden oft vom Provider des Objektspeichersystems bereitgestellt.

Für das Management der Daten im Objektspeicher wird in vielen Fällen eine relationale Datenbank angelegt, worin alle Objekte indexiert werden. Somit lassen sich zusätzliche Informationen oder extrahierte Informationen zu den Dateien speichern und verwalten. Insbesondere die Leistung von Suchvorgängen kann durch eine Indexierung massiv effizienter gestaltet werden.

### 2.4.2 Bereitstellung von Daten

Der Einsatz von REST (Representational State Transfer) API's ist eine weit verbreitete Methode für den Zugriff auf Daten aus Datenbanken und Objektspeichern. Sie bietet eine standardisierte Möglichkeit der Abfrage, Manipulation, Löschung und Erstellung von Daten über das HTTP Protokoll. Durch die Abstraktion der Datenstrukturen werden Abhängigkeiten zu darunterliegenden Systemen aufgehoben. Die flexible Zugriffsmethode erhöht die Skalierbarkeit, die

Wiederverwendbarkeit und die Wartbarkeit der Anwendung. Abbildung 9 zeigt die schematische Anordnung der API zwischen dem Client und den Datenspeichersystemen.

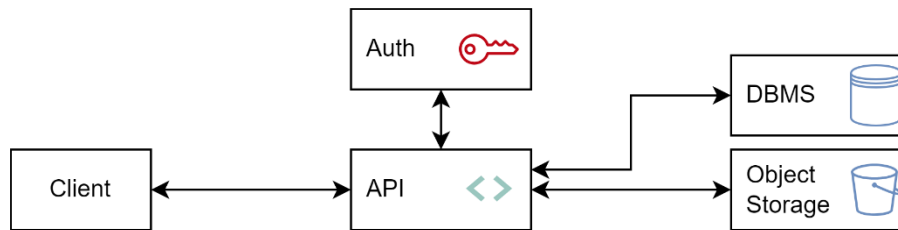


Abbildung 9 Bereitstellung von Daten: REST API

Eine REST API besteht aus mehreren Endpunkten, die auf bestimmte Ressourcen verweisen, die in unterschiedlichen Systemen gespeichert werden. Requests an die Endpunkte werden entsprechend transformiert. Für eine Abfrage von Einträgen aus einer SQL Tabelle wird ein SQL Query gebildet und an das DBMS gesendet. Die Antwort wird wiederum in das standardisierte Format konvertiert und dem Client als Response übermittelt. SQL Queries können Parameter erhalten, die in einem Request mitgeschickt wurden und sind potentiell anfällig auf SQL Injection Attacks. Bei der Konstruktion von SQL Queries müssen Parameter in jedem Fall validiert werden, um das System sicher zu halten.

Es gibt Ressourcen, deren Abfrage oder Modifikation spezielle Berechtigungen erfordert. In der API können Routen geschützt werden. Bei Anfragen überprüft sie, ob ein Client über die entsprechenden Berechtigungen verfügt. Diese Funktionalität kann direkt implementiert werden oder auf ein IAM System ausgelagert werden. Das User Management wird im nächsten Kapitel dokumentiert.

## 2.5 Access Management

Dieses Kapitel dokumentiert die wichtigsten Grundlagen des Zugriffsmanagements im Bezug auf Webanwendungen.

### 2.5.1 Autorisierung und Authentifizierung

In Webanwendungen gibt es öffentlich zugängliche Ressourcen und Informationen, die nur für eine definierte Nutzergruppe zugänglich sein dürfen. Als Autorisierung wird das Erteilen von Zugriffsrechten auf gewisse Ressourcen an Nutzende bezeichnet. Der Prozess der Identitätsüberprüfung als Authentifizierung bezeichnet. In der Regel wird die Autorisierung nach der Authentifizierung geprüft.

Das User Management umschließt die Autorisierung, die Authentifizierung und die Verwaltung von Nutzern. Folgend werden die wichtigsten Protokolle und Formate beschrieben, die in Access Management Systemen zum Einsatz kommen.

### 2.5.2 OAuth

OAuth ist ein Protokoll für den Austausch der Autorisierung zwischen verschiedenen Diensten, ohne die Anmeldedaten des Users zu teilen. Für die Anmeldung in einer Applikation wird auf einen Autorisierungsserver weitergeleitet, wo die Eingabe der Anmeldedaten erfolgt. Im Autorisierungsserver wird bei erfolgreicher Anmeldung ein Token generiert, der an die Applikation übermittelt wird. Wird ein Autorisierungsserver für mehrere Applikationen eingesetzt, wird das auch als Single-Sign-On (SSO) bezeichnet.

### 2.5.3 OpenID Connect (OIDC)

OIDC ist ein auf OAuth 2.0 basierendes Identitätsprotokoll. Während OAuth ein Protokoll zur Autorisierung ist, liegt der Schwerpunkt bei OIDC bei der Authentifizierung, also der Verifikation der Identität des Users gegenüber einem Client Dienst. Wie bei OAuth erfolgt eine Weiterleitung auf einen Autorisierungsserver für die Anmeldung. Bei erfolgreicher Anmeldung wird neben dem ID Token ein Access Token und ein Refresh Token an die Client Applikation weitergeleitet. Der Access Token hat eine relativ kurze Gültigkeitsdauer und ermöglicht Zugriff auf geschützte Ressourcen. Der Refresh Token, der in der Regel länger gültig ist, wird verwendet, um neue Access Token zu beschaffen.

### 2.5.4 JSON Web Token (JWT)

JWT ist ein offener Standard nach RFC 7519 [24] , der einen kompakten Weg für den sicheren Austausch von Informationen zwischen zwei Parteien ermöglicht. JWT enthalten sämtliche authentifizierungsrelevanten Informationen, womit sie sich zur Implementierung von Stateless Sessions eignen.

JWT bestehen aus drei JSON Objekten, die jeweils Base64-encodiert werden und mit einem Punkt getrennt aneinandergereiht sind:

- Der Header enthält Metadaten, die den Typ und den Algorithmus der Verschlüsselung des Tokens beschreiben.
- Die Payload enthält neben der Gültigkeitsdauer des Tokens Informationen zum User, wie Benutzername, Nutzer-ID und Berechtigungen.
- Die Signatur ist im dritten Teil enthalten. Sie wird eingesetzt, um den Token zu validieren, um sicherzugehen, dass er nicht modifiziert wurde.

### 2.5.5 IAM Systeme

Identity and Access Management (IAM) Systeme werden für die Autorisierung, die Authentifizierung und die Verwaltung von nutzenden Personen eingesetzt. Sie erfordern eine sichere Speichermöglichkeit für die Nutzerdaten. Für die Berechtigungen werden oft rollenbasierte Konzepte eingesetzt. Von den meisten IAM Systemen wird die Möglichkeit einer Anmeldung mit einem Konto aus einem anderen System, zum Beispiel bei Google, angeboten. IAM Systeme können lokal betrieben werden oder von Drittanbietern als Service genutzt werden. Oft besteht eine systemweite Abhängigkeit zu der Erreichbarkeit des IAM Systems, wodurch ein zuverlässiger Betrieb unerlässlich ist.

## **3 Analyse**

### **3.1 Anforderungen an das Tool**

Die Anforderungen an das Tool wurde in Zusammenarbeit mit dem Team des Forschungsprojekts Mitwelten festgelegt. Sie sind folgend beschrieben.

#### **3.1.1 Darstellung von Daten**

Die Hauptaufgabe des Tools liegt bei der Visualisierung der im Projekt Mitwelten erhobenen Daten. Es sollen verschiedene Achsen von Teilmengen der Datensätze dargestellt werden, um Trends, Muster oder Anomalien darin zu erkennen. Zusätzlich zu den vorhandenen Daten sollen weitere Daten aus externen Quellen in das Tool einbezogen werden, die einen möglichen Einfluss auf die Biodiversität haben. Es sollen verschiedene Arten der Datenvisualisierung eingesetzt werden, um verschiedene Aspekte der Daten hervorzuheben.

#### **3.1.2 Filtern von Daten**

Die im Rahmen des Projekts Mitwelten erhobenen Daten liegen in komplexer Struktur vor. Das Tool soll es ermöglichen, Teilmengen aus den Datensätzen untereinander oder mit anderen Datensätzen zu vergleichen. Zudem soll es möglich sein, den Zeitraum der angezeigten Daten einzugrenzen.

#### **3.1.3 Reproduzierbarkeit der Ansichten**

Um Dashboards zu speichern oder mit anderen Nutzenden zu Teilen sollen alle Ansichten, die im Tool generiert werden, reproduzierbar sein.

#### **3.1.4 Verfassen von Annotationen**

Die Applikation soll eine Funktion bereitstellen, um Erkenntnisse oder Kommentare im Bezug auf Ansichten zu speichern. Die Annotationen sollen für alle registrierten Personen sichtbar sein, um den Informationsaustausch zu gewährleisten.

#### **3.1.5 Daten Schnittstelle für externe Tools**

Die im Projekt Mitwelten erhobenen Daten sollen für weiterführende Analysen mit unabhängigen Tools bereitgestellt werden.

#### **3.1.6 Authentifizierung**

Die Applikation soll über eine Authentifizierung verfügen, um den Datenzugriff gegebenenfalls einzuschränken. Zur Vergabe von Rechten sollen persönliche Konten verwendet werden.

## 3.2 Daten

In der zu entwickelnden Applikation sollen verschiedene Datensätze visualisiert und analysiert werden. Am relevantesten sind die im Projekt Mitwelten erhobenen Daten. In einer Recherche wurden Datensets aus externen Quellen ermittelt. Folgend werden die Datensätze beschrieben.

### 3.2.1 BirdNET Auswertungen

Der Datensatz BirdNET Resultate wurde im Rahmen des Projekts Mitwelten erhoben. Er repräsentiert Vogelstimmen auf Audioaufnahmen, die mit dem BirdNET Algorithmus [25] ausgewertet wurden.

#### 3.2.1.1 Messprinzip

Die Audioaufnahmen wurden mit AudioMoths, Vollspektrum Mikrofonen von Open Acoustic Devices [26], an verschiedenen Orten während unterschiedlichen Zeiten erhoben. Die Aufnahmen wurden auf einer S3 Instanz gespeichert und auf einer VM mit dem BirdNET Algorithmus ausgewertet. Für die Auswertung wird jede Aufnahme in drei Sekunden lange Stücke geteilt. Das Model analysiert jedes Segment und analysiert, ob sich Vogelstimmen darauf befinden. Falls das Segment Vogelstimmen enthält, werden die Wahrscheinlichkeiten von bis zu drei Spezies, die in dem Segment zu hören sind, als Resultat gespeichert. Durch die Aufteilung der Dateien besteht die Gefahr, dass Rufe von Vögeln zerteilt werden und somit nicht erkannt werden. Um dieses Situation vorzubeugen, wird jede Aufnahme mit einem Offset ein zweites Mal analysiert.

#### 3.2.1.2 Datenstruktur

Die Resultate des BirdNET Algorithmus haben die Form einer Liste mit folgenden Parametern:

- Der wissenschaftliche Name des erkannten Vogels.
- Die Wahrscheinlichkeit (Confidence), dass es sich um den Vogel handelt.
- Der Zeitpunkt, relativ zum Start der Aufnahme, an welchem er erkannt wurde.

Durch den Dateinamen sind Ort und Zeit der Aufnahme bekannt.

#### 3.2.1.3 Interpretation und Qualität der Daten

Bei den Einträgen im Datenset handelt es sich um erkannte Vogelstimmen, nicht um Individuen. Die akustische Aktivität von Vögeln kann nach Art und Zeit stark variieren, womit direkte Vergleiche nur mit entsprechenden Hintergrundinformationen aussagekräftig sind. Es ist möglich, dass Rufe nicht oder falsch erkannt wurden. Es existieren Arten, die Rufe von anderen Arten nachahmen. Es ist möglich, dass nachgeahmte Rufe zu falschen Resultaten führen.

Die Aufnahmezeiträumen wurden nicht bei allen Mikrofonen einheitlich festgelegt. Bei einem kleineren Aufnahmezeitraum werden potentiell weniger Rufe erkannt. Die Qualität der Audioaufnahmen kann die Resultate beeinflussen. Bei Standorten mit lauten Umgebungsgeräuschen sind Qualitätseinbussen möglich.

### 3.2.2 Bestäuber Datenset

Im Rahmen des Mitwelten Projekts wurde eine Pollinator Studie durchgeführt. An verschiedenen Orten wurden Blumentöpfe platziert. Mit Kameras wurden in regelmässigen Zeitabständen Bilder von den Blumen aufgenommen. Mit Hilfe einer ML Pipeline wurden Bestäuber auf den Bildern erkannt.

### 3.2.2.1 Auswertung

Die aufgenommenen Bilder wurden von einer Machine Learning Pipeline ausgewertet. In einem ersten Schritt wurde von einem Objekterkennungsmodell Blumen auf den Bildern ausgeschnitten. Es wurden nur Margariten, Flockenblumen und Wilde Möhren berücksichtigt. Ein weiteres Objekterkennungsmodell wurde eingesetzt, um die Bildausschnitte nach Bestäubern abzusuchen. Das Bestäuber Modell kann die Insekten in die Klassen Honigbienen, Wildbienen, Hummeln, Fliegen und Schwebfliegen kategorisieren. Zum Verfassungszeitpunkt dieser Dokumentation werden nur Margariten und Flockenblumen nach Bestäubern abgesucht.

### 3.2.2.2 Interpretation der Daten

Die Werte sind eine Momentaufnahme, die eine Aussage darüber machen, welche Bestäuber mit welcher Confidence auf dem Bild vorhanden sind. Die Taxonomie der Bestäuber Kategorien ist in Abbildung 10 ersichtlich. Alle Bestäuber der Familie Apidae, die nicht zu den Gattungen *Bombus* oder *Apis* zählen, werden als Wildbienen erkannt. Durch die hohe Anzahl gemeinsamer optischer Merkmale der Kategorie Wildbienen mit den Kategorien Honigbienen und Hummeln können gelegentlich Fehlklassifikationen vorkommen.

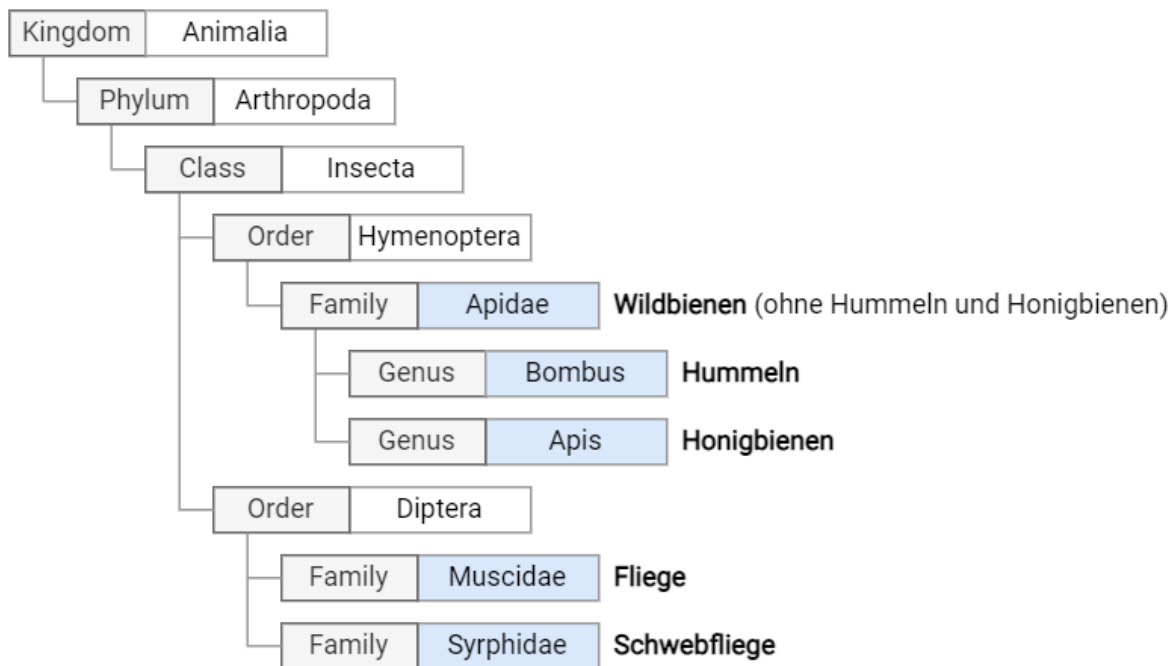


Abbildung 10 Taxonomie Kategorien Bestäuber Datenset

### 3.2.3 PAX Counter Messungen

Als PAX Counter wird ein System bezeichnet, das Personen zählt. Im Rahmen des Projekts Mitwelten wurden PAX Counter entwickelt, die periodisch ausgesendete Signale von Smartphones empfangen und die Anzahl eindeutiger Smartphones im Umkreis von einigen Metern als Messwert übertragen.

#### 3.2.3.1 Messprinzip

Smartphones senden in regelmässigen Abständen BLE Advertising Pakete. Die Pakete bestehen aus einer MAC Adresse und zusätzlicher Payload. Bei neueren Geräten wird aus Datenschutzgründen eine randomisierte Adresse gesendet, die täglich oder in kürzeren Abständen neu generiert wird. Der PAX Counter Empfängt Advertising Pakete und speichert einen Hash Wert der



MAC Adresse. In einem festgelegten Intervall wird die Anzahl eindeutiger Adressen via LoRa oder MQTT an das Backend übertragen.

### 3.2.3.2 Interpretation der Daten

Die Summe der erkannten Geräte pro Zeitintervall wird übertragen. Falls sich ein Gerät über mehrere Messintervalle hinweg in Reichweite befindet, so wird es in jeder Messung mitgezählt. Bei den PAX Countern, die über WiFi angebunden sind, ist der Messintervall viel kleiner als bei den LoRa Geräten, was ein direkter Vergleich verunmöglicht.

### 3.2.4 Environment Sensor Messungen

Während den Sommermonaten im Jahr 2021 wurde an verschiedenen Orten das lokale Klima mit LoRa Sensor Nodes gemessen. Folgende Parameter wurden alle 15 Minuten gemessen:

- Die Lufttemperatur, ca. 30 cm über Boden
- Die relative Luftfeuchtigkeit, ca. 30 cm über Boden
- Die Bodenfeuchtigkeit ca. 10cm unter der Oberfläche

#### 3.2.4.1 Qualität und Interpretation der Daten

Die Messwerte im Datenset wurden, bedingt durch den Messaufbau, teilweise verfälscht. Durch Komplikationen der Datenübertragung und durch ausgefallene Sensoren sind die Messreihen teilweise unvollständig. Für Vergleiche von relativen Änderungen der Werte an verschiedenen Standorten eignet sich das Datenset.

### 3.2.5 Informationen Taxa

Die BirdNET Resultate und die Einträge in Bestäuber Datenset repräsentieren Detektionen von gewissen Arten. BirdNET klassifiziert die Rufe auf der Taxonomie Ebene Art. Die erkannten Arten werden in Form des wissenschaftlichen Namens angegeben. Im Bestäuber Datenset befinden sich die Kategorien auf den Taxonomie Ebenen Gattung und Familie, als Ausgabeformat wurde der deutsche Namen festgelegt.

#### 3.2.5.1 Taxonomie

Für die meisten Personen sind Arten unter dem Namen in lokaler Sprache und nicht unter der wissenschaftlichen Bezeichnung bekannt. Das GBIF Backbone Taxonomy [27] Datenset wird von GBIF zur Verwaltung der Taxa verwendet. Es enthält gut vier Millionen Einträge davon knapp 1.5 Millionen aus dem Königreich Animalia (Tiere). Für jedes Taxon wurde eine eindeutige ID definiert. Neben der ID enthalten die Einträge unter anderem folgende Informationen:

- Der wissenschaftliche Namen.
- Die Taxonomie Stufe.
- Gebräuchliche Namen in unterschiedlichen Sprachen.
- ID und Namen der übergeordneten Taxa.

Über die API von GBIF können Taxa anhand der ID oder des Namens abgefragt werden. Um die Abhängigkeiten des zu entwickelnden Tools möglichst gering zu halten, wurden Informationen zu der Taxonomie und Namensübersetzungen der Taxa in den BirdNET Resultaten und im Bestäuber Datenset in einer Datenbank im Mitwelten Backend gespeichert. Die Miteinbeziehung der Taxonomie Informationen ermöglichen Abfragen von Erkennungen einer bestimmter Gattung, Familie oder Ordnung.

### 3.2.5.2 Kurzbeschreibung

Mit einem kurzen textuellen Beschrieb eines Taxons können relevante Hintergrundinformationen direkt in das Tool integriert werden. Wikipedia hat für die meisten Taxa einen Eintrag in mehreren Sprachen. Der erste Abschnitt eines Eintrags enthält die wichtigsten Punkte des Artikels. Über die MediaWiki Rest API [28] kann das Summary in abgefragt werden.

### 3.2.5.3 Referenzbilder

Neben dem Namen und einer Beschreibung eines Taxons kann ein Bild viel Informationen enthalten. Wikispecies [29] hat für eine Vielzahl der im Projekt detektierten Taxa ein Referenzbild, wovon die URL über die MediaWiki Rest API [28] abgefragt werden kann. Für alle von BirdNET erkannten Arten wurde die URL des Referenzbildes in der Datenbank im Mitwelten Backend gespeichert. Für jedes Bild kann die Quellenangabe über den Dateinamen in der URL von der MediaWiki API [28] abgefragt werden.

### 3.2.6 Meteodaten

Meteodaten haben einen möglichen Einfluss auf die Aktivität von Spezies. Das Bundesamt für Meteorologie und Klimatologie MeteoSchweiz [30] betreibt ein engmaschiges Messsystem, welches die wichtigsten Meteodaten periodisch misst. Mehrere Stationen liegen in der direkten Umgebung der Mitwelten Feldstudien. Sie sind in Abbildung 11 ersichtlich.

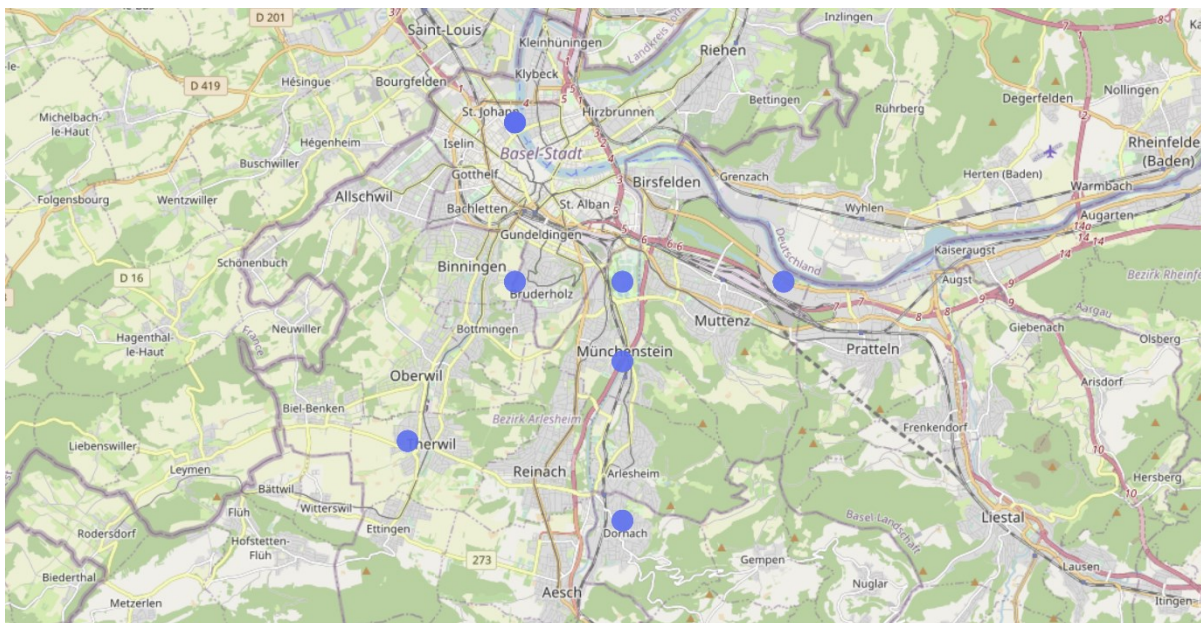


Abbildung 11 MeteoSchweiz Stationen in den Gebieten der Mitwelten Feldstudien [30]

Die Parameter sind in roher Form und als aggregierte Werte vorhanden, wobei die Aggregation im jeweiligen Parameternamen beschrieben ist.

MeteoSchweiz stellt die erhobenen Daten über das Datenportal IDAWEB [31] für Lehre und Forschung zur Verfügung. Das Projekt Mitwelten fällt in die Kategorie Forschung, was die Nutzung erlaubt.

### 3.2.7 Beobachtungen aus externen Quellen

Es gibt eine Vielzahl an Institutionen und Communities, die Beobachtungen von Spezies erfassen. GBIF (Global Biodiversity Information Facility) ist eine Plattform, die Datensets von verschiedenen Stellen sammelt und Beobachtungen in standardisiertem Format bereitstellt.

Die Beobachtungen aus externen Quellen können eingesetzt werden, um teile der automatisierten Beobachtungen aus dem Projekt Mitwelten zu validieren. Bei seltenen Arten kann überprüft werden, ob und wann die Art in der Umgebung bereits erkannt wurde.

Eine Suche nach Beobachtungen der Klasse Aves (Vögel) in den Kantonen Basel Stadt und Basel Landschaft zwischen 2020 und 2022 ergibt gut 23'000 Resultate. Die Quellen der Beobachtungen sind in Abbildung 12 aufgeführt.

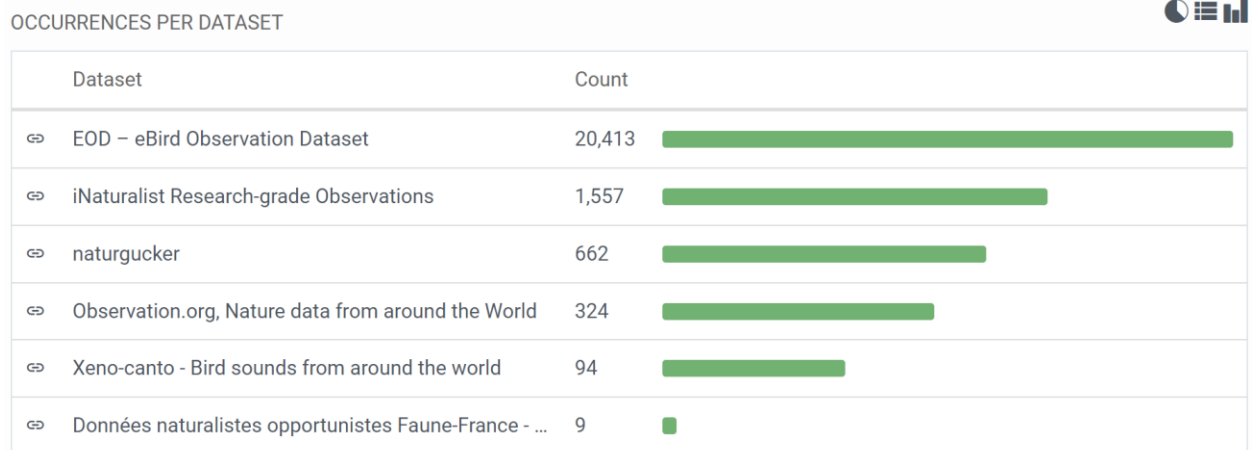


Abbildung 12 GBIF Resultate Klasse Aves Kt. Basel Stadt & Kt. Basel Land, 2020-2022 [32]

Über 20'000 Beobachtungen entstammen dem eBird Observation Dataset [33], welches Auswertungen der BirdNET Mobile App [34] aufführt. An zweiter Stelle sind Beobachtungen von iNaturalist [35], einer Community aus Biologen, Biologinnen und interessierten Personen, aufgeführt.

### 3.2.7.1 Qualität der Datensets

Bei einigen Beobachtungen wurden fehlerhafte Erfassungszeitpunkte festgestellt. Die Tageszeiten der Beobachtungen sind teilweise nicht vorhanden. Beobachtungen, die aus dem eBird Observation Dataset stammen, beinhalten, bedingt durch die Auswertung mit BirdNET, möglicherweise Fehldetektionen.

### 3.2.7.2 Datenabfrage

Über die GBIF Occurrence API [36] werden sämtliche Beobachtungen in standardisiertem Format bereitgestellt. Pro Request können maximal 300 Einträge abgefragt werden. Jedes in GBIF vorhandene Taxon wird durch einen eindeutigen `taxon_key` identifiziert. Abfragen von Beobachtungen bestimmter Taxa werden anhand des `taxon_keys` gemacht. Der GBIF Taxonomy Tree kann genutzt werden, um die Parent-Taxa anhand des Keys zu ermitteln.

## 3.2.8 Charakterisierung der Erfassungsstandorte

Als Zusatzinformationen zu den im Projekt Mitwelten erhobenen Daten wurde beschlossen, die Eigenschaften der Messstandorte mit festgelegten Parametern zu beschreiben. Um Orte zu vergleichen, sollen die Parameter normiert werden. Die Werte sollen jeweils zwischen 0 und 10 liegen. Ein erster Vorschlag von möglichen Attributen ist in Tabelle 1 ersichtlich.

Bezeichnung	Beschreibung Minimalwert	Beschreibung Maximalwert
Siedlungsfaktor	Offene Landschaft	Stark besiedeltes Gebiet
Bodenversiegelung	Keine Versiegelung	Vollständig versiegelter Boden
Sonneneinstrahlung	Keine direkte Sonneneinstrahlung	Kontinuierliche Sonneneinstrahlung
Gewässernähe	Keine Gewässer in der Nähe	Direkt am Gewässer
Blühangebot	Geringes Blühangebot	Sehr hohes Blühangebot
Substratvorkommen	Kein Vorkommen von organischem Substrat	Sehr hohes Vorkommen von organischem Substrat
Nistmöglichkeit	Sehr ungeeignet für Bestäubernester	Sehr geeignet für Bestäubernester
Fragmentierung	Geringe Fragmentierung der Landschaft	Hohe Fragmentierung der Landschaft
Habitatvielfalt	Geringe Vielfalt natürlicher Lebensräume	Hohe Vielfalt natürlicher Lebensräume
Pflanzenvielfalt	Geringe Pflanzenvielfalt	Hohe Pflanzenvielfalt

Tabelle 1 Vorschlag Attribute zur Beschreibung von Umgebungsfaktoren

### 3.2.9 Map Layer

Karten eignen sich, um räumliche Informationen zu vermitteln. Das Bundesamt für Landestopografie [37] stellt digitale Karten und Bilder der Landschaft in hoher Qualität zur Verfügung. Neben Basiskarten können Informationen wie Lebensräume, Oberflächenstruktur oder Bevölkerungsdichte auf einem Map Layer visualisiert werden. Im Auftrag des Bundesamt für Umwelt [38] werden verschiedenste Datensätze als Map Layer bereitgestellt. Interessant im Bezug auf die Biodiversität ist die Lebensraumkarte, in der die Schweiz flächendeckend gemäss der TypoCH-Klassifikation «Lebensräume der Schweiz» [39] kategorisiert ist. Ein Ausschnitt der Lebensraumkarte mit Legende ist in Abbildung 13 ersichtlich.



Abbildung 13 Kartenausschnitt Lebensraumkarte [38]

Die Basiskarten von Swisstopo und die Layer vom Bundesamt für Umwelt werden unter anderem als WMS Layer über die GeoAdmin API [40] bereitgestellt.

### 3.3 Datensets Explore Applikation

In der zu entwickelnden Applikation sollen Teilmengen der vorhandenen Daten analysierbar und untereinander vergleichbar sein. Eine Referenz auf ein Subset der Daten wird als Datenset bezeichnet. Folgend werden Datensets, die sich für den Einsatz im Tool eignen, dokumentiert.

#### 3.3.1 Taxa

Ein Taxa Datenset beinhaltet Beobachtungen von Tieren aus einer bestimmten Taxonomiegruppe, zum Beispiel der Art Mauersegler, der Familie Apoidea oder der Klasse Insecta. Damit Beobachtungen vom Projekt Mitwelten mit Beobachtungen aus externen Quellen vergleichbar sind, existieren pro Taxon jeweils zwei Datensets. Das Datenset **Taxon** beinhaltet nur Datenpunkte, die im Rahmen des Projekts Mitwelten erhoben wurden, das Datenset **Taxon (GBIF)** beinhaltet Datenpunkte aus externen Quellen.

Die darstellbaren Dimensionen von Taxon Datensets umfassen die zeitliche Aktivität der Taxa und die räumliche Aktivität der Taxa. Die zeitliche Aktivität kann saisonal dargestellt werden, um zu analysieren, wie aktiv die Taxa zu den jeweiligen Jahreszeiten sind. Für die Analyse von Tageszeiten mit erhöhter Aktivität kann die stündliche Aktivität berechnet und dargestellt werden.

#### 3.3.2 Pollinators

Ein Pollinator Datenset referenziert Beobachtungen von Bestäubern, die aus den Daten des Projekts Mitwelten stammen. Die Erfassungsorte und Klassen der beobachteten Bestäuber können eingegrenzt werden. Durch eine Eingrenzung des Erfassungsorts ist es möglich, verschiedene Ökosysteme zu vergleichen. Neben der zeitlichen und räumlichen Aktivität der Bestäuber kann die Diversität im Bezug auf einen Ort dargestellt werden.

#### 3.3.3 Bird Diversity

Ein Bird Diversity Datenset referenziert Beobachtungen, die von bestimmten Mitwelten Audio Deployments erhoben wurden. Im Gegensatz zu den Taxa Datensets, welche die Summe von Beobachtungen eines bestimmten Taxons referenzieren, wird die Anzahl verschiedener Spezies und somit die Artenvielfalt innerhalb der Klasse Aves repräsentiert.

Die darstellbaren Dimensionen sind identisch zu denen des Pollinator Datensets. Anstelle der Aktivität wird die Diversität dargestellt. Zusätzlich kann die Zusammensetzung der erhobenen Daten an einem bestimmten Ort untersucht werden.

Für die Bestäuber ist diese Ansicht weniger relevant, da praktisch kein Fall existiert, in dem nicht alle Kategorien auftreten.

#### 3.3.4 Meteodata

Ein Meteo Datenset referenziert eine Messreihe, die von einer definierten Meteo Station erhoben wurde. Die Stationen haben unterschiedliche Parameter, die gemessen werden und sind somit nicht direkt vergleichbar. Für die Darstellung ist die Zeitachse am relevantesten. Meteodaten sind besonders interessant, um deren Einflüsse und Abhängigkeiten auf die Aktivität von Taxa oder die Artendiversität zu untersuchen.

#### 3.3.5 PAX

Ein PAX Datenset referenziert PAX Werte, die von einem definierten Deployment erhoben wurden. Wie bei den Meteodaten ist die Zeitachse am Aussagekräftigsten. Mit einer Darstellung der PAX Werte im Bezug auf die Tageszeit kann analysiert werden, an welchen Stunden die



Anwesenheit von Menschen einen potentiellen Einfluss auf andere Daten, wie zum Beispiel die Aktivität von Bestäubern, haben können.

### **3.3.6 Env Measurements**

Ein ENV Measurement Datenset referenziert Messwerte eines Sensors, die von einem bestimmten Deployment erhoben wurden. Aus funktionaler Sicht und auch im Bezug auf darstellbare Dimensionen ist das Datenset identisch zu einem Meteodataset.

## **3.4 Bedienstruktur und Workflow der Applikation**

Als Workflow wird eine Abfolge von Schritten in einer bestimmten Reihenfolge beschrieben, die nötig sind, um ein Ziel zu erreichen. In der Explore Applikation ist das Ziel die Analyse von Datensets oder das Verfassen einer Annotation. Dafür muss ein Datenset selektiert und in dem gewünschten Dashboard geöffnet werden. Damit das Datenset nicht jedes Mal neu ausgewählt werden muss, kann es zwischengespeichert werden. Der Workflow ist ähnlich zu der Bedienung eines Onlineshops, wobei die Datensets in einem Shop Produkte wären. Die einzelnen Schritte des Workflows werden folgend beschrieben.

### **3.4.1 Suchen**

Als erstes muss ein Datenset gesucht werden. Die Datensets sind nach Kategorie unterteilt, Filter helfen bei der Suche nach einem bestimmten Datenset. In einem Onlineshop entspricht dieser Schritt der Suche nach einem Produkt. Für eine Suche nach einer SSD auf Digitec können passende Produkte in der Kategorie 'Festplatten' mit der Festlegung von bestimmten Filtern gefunden werden.

### **3.4.2 Sammeln**

Wurde ein Datenset gefunden, wird es in der Collection gespeichert. Die Collection ist eine Liste von Referenzen auf Datensets, die ein Nutzer gesammelt hat. In einem Onlineshop entspricht die Collection dem Warenkorb oder in gewissen Fällen einer Vergleichsliste.

### **3.4.3 Visualisieren**

In verschiedenen Dashboards können einzelne Datensets sowie Kombinationen von Datensets visualisiert und analysiert werden. Beim Öffnen eines Dashboards werden alle kompatiblen Datensets, die in der Collection vorhanden sind, zur Auswahl bereitgestellt. In der Collection-Ansicht sind für alle Datensets Direktlinks zu kompatiblen Dashboards aufgeführt.

### **3.4.4 Annotieren**

Bemerkungen, Beobachtungen und Erkenntnisse, die in Dashboards gemacht werden, können als Annotationen verfasst werden. Annotationen werden mit den registrierten Personen der Applikation geteilt und dienen zum Informationsaustausch. Diese Funktion hat Ähnlichkeiten zu Produktbewertungen in Webstores.

### 3.5 Dashboard Layout

Ein Dashboard besteht aus Modulen zur Darstellung von Informationen, aus Elementen zur Bedienung und einem Titel. Bedienelemente ermöglichen es den nutzenden Personen, mit den angezeigten Informationen zu interagieren. Über die Steuerelemente können Daten gefiltert, Ansichten geändert und Aktionen ausgeführt werden. Häufig verwendete Eingabeelemente sind Schaltflächen, Dropdowns, Schieberegler oder Kontrollkästchen. Elemente zur Modifikation von dargestellten Daten werden auch als Filter bezeichnet.

Für jedes Bedienelement muss für die Anwender und Anwenderinnen verständlich sein, was es bewirkt und auf welche Daten es sich bezieht. Durch Beschriftungen mit Labels oder Tooltips kann die Funktionalität kommuniziert werden. Die Positionierung von Eingabeelementen kann indizieren, welche Daten davon abhängen.

Eingabeelemente in der Navigationsleiste, wie ein Schalter zur Auswahl zwischen Hell- und Dunkelmodus, haben in der Regel eine Auswirkung auf alle Bereiche einer Applikation. In einem Dashboard sind Filter oft über den abhängigen Elementen positioniert. Inline Filter beziehen sich auf einzelne Elemente und werden in unmittelbarer Nähe angeordnet [41].

Gemäss mehreren Studien betrachten Personen beim Öffnen einer Webseite oder eines Dashboards als erstes den oberen linken Bereich. Um zu vermitteln, was im Dashboard angezeigt wird, ist es sinnvoll, den Titel an dieser Stelle zu platzieren. Filter, die alle Daten betreffen, können auf gleicher Höhe angeordnet werden. Abbildung 14 zeigt einen Entwurf mit der entsprechenden Anordnung. Module zur Präsentation von Informationen und Daten werden in der blau eingefärbten Content Area dargestellt.

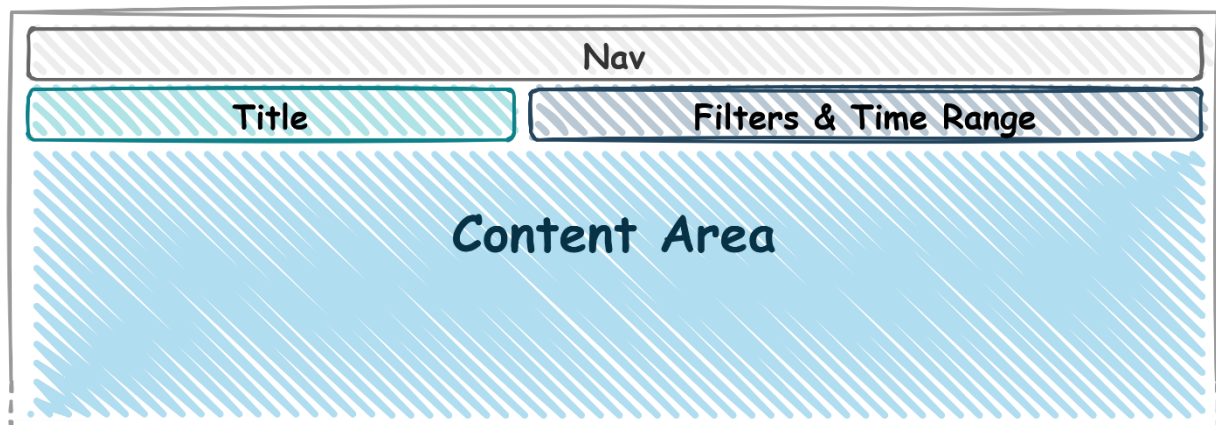


Abbildung 14 Layout Design Entwurf - Titel und Filter

Die Content Area ist modular aufgebaut. Abhängig von der Art eines Dashboards werden unterschiedliche Datenvisualisierungsmodule eingesetzt. Die Module haben im jeweiligen Kontext unterschiedliche Relevanzen. Die Anordnung der Module muss so gewählt werden, dass die wichtigsten Informationen als erstes wahrgenommen werden. F- und Z-Pattern besagen, dass sich der Blick grundsätzlich vom oberen linken Bereich zum unteren rechten Bereich bewegt.

Im Rahmen einer Studie zum Detektieren der Nutzeraktivität anhand des Blicks wurden Eye-Tracking Daten von mehreren Testpersonen erhoben [42]. Die Daten sind nach Aktivität kategorisiert. Um die Prominenz der Bildschirmregionen zu erforschen wurden die Datensätze bei den Aktivitäten Browsing, Interpreting, Debugging und Searching analysiert. Die Datenexploration in einem Dashboard besteht aus diesen Aktivitäten. Aus den knapp 700'000 Datenpunkten wurde eine Heatmap generiert, die den Bildschirm auf 384x240 Pixel darstellt. Abbildung 15 zeigt eine Visualisierung der aggregierten Datenpunkte. Rot eingefärbte Bereiche wurden häufig erblickt, die blauen Bereiche sehr wenig.

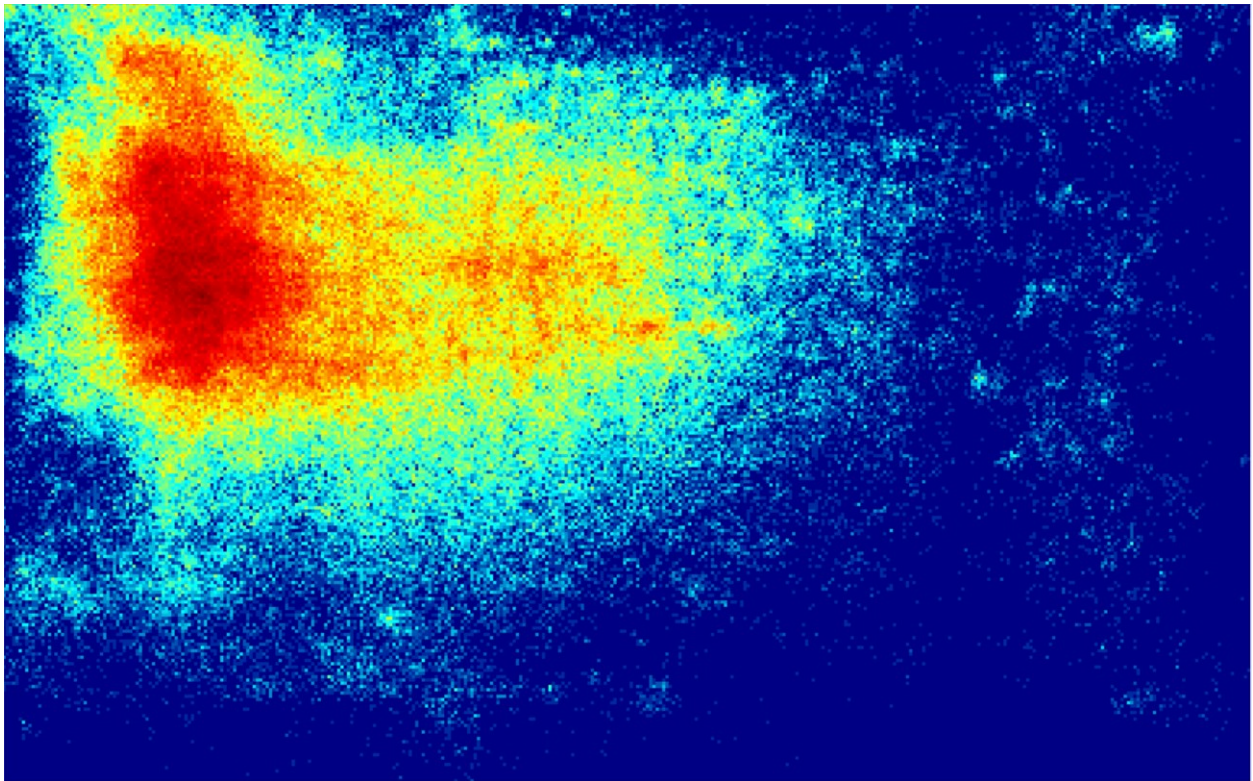


Abbildung 15 Prominente Bereiche auf dem Bildschirm [42]

Die Eye-Tracking Daten bestätigen die hohe Priorität des oberen linken Bereichs. Gelegentlich wurde die rechte Seite betrachtet, der untere Rand wird, vor Allem auf der rechten Seite, praktisch nie erblickt.

Auf Basis der Literaturrecherche mit Einbezug der Eye Tracking Daten wurden mehrere Layouts mit verschiedenen Anordnungen von unterschiedlich priorisierten Modulen entworfen. Abbildung 16 die entworfenen Grundstrukturen für drei Arten von Dashboards. Die Priorität der Elemente ist farblich gekennzeichnet. Durch den Einsatz von Tabs kann ein Modul mehrere Elemente beinhalten. Die Priorität von nicht direkt angezeigten Elementen liegt auf der niedrigsten Stufe.

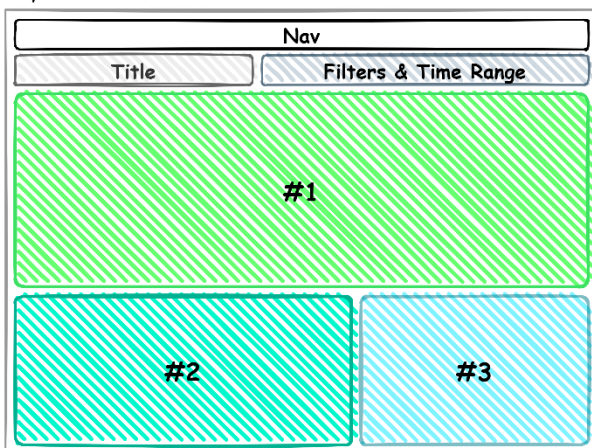
Das erste Layout beinhaltet drei Elemente. Rund die Hälfte der Content Area wird vom Hauptelement eingenommen. Diese Struktur eignet sich für die Analyse von Zeitreihendaten, da der Time Series Chart für diese Ansicht sehr relevant ist. Das sekundäre Element kann für die Darstellung der Daten in einer anderen Dimension eingesetzt werden. Hintergrundinformationen, zum Beispiel der Erfassungsort, sind im Modul mit Priorität 3 anzuordnen.

Das zweite Layout wurde für die Analyse von räumlichen Daten konzipiert. Das Hauptelement ist für eine Kartendarstellung vorgesehen und füllt die gesamte Höhe der Content Area. Indikatoren, Legenden und weitere Bedienelemente sind in den Modulen rechts davon anzuordnen.

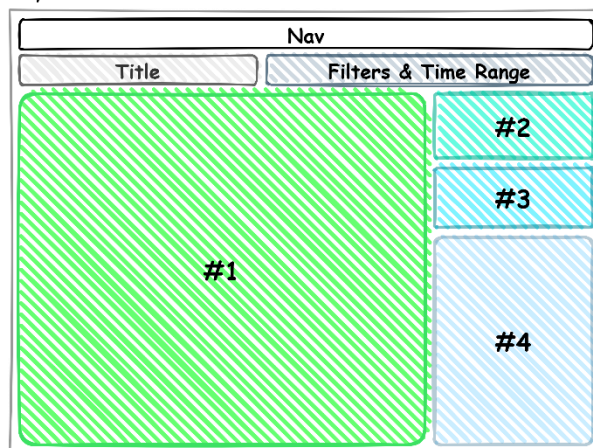
Das dritte Layout beinhaltet fünf Module, die in zwei Spalten angeordnet sind. Unterhalb des Hauptelements befinden sich zwei Module mit identischer Grösse, wobei das linke Modul höher priorisiert ist. In der rechten Spalte befindet sich ein kleines Modul an oberster Stelle. Die restliche Fläche wird durch ein Modul mit niedriger Priorität ausgefüllt. Dieses Layout wurde zur Visualisierung von Tierbeobachtungen entwickelt. In der linken Spalte können zeitliche und räumliche Aktivitäten visualisiert werden. Hintergrundinformationen können in dem Modul auf der rechten Seite dargestellt werden.



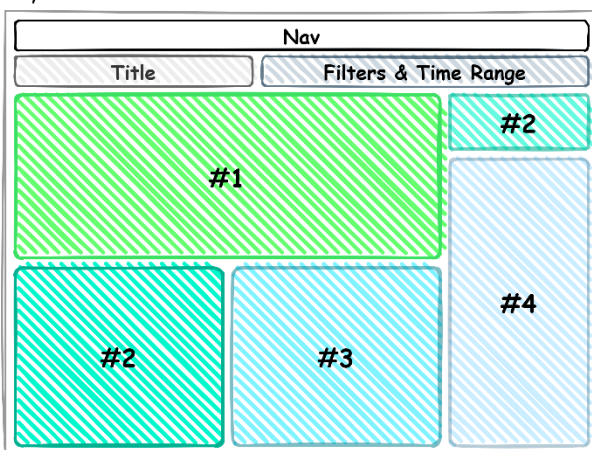
Layout 1



Layout 2



Layout 3



### Priority

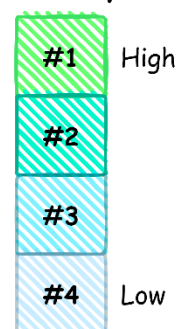


Abbildung 16 Dashboard Layout Entwurf

Neben den Bedienelementen zur Filterung und zur Darstellung von Daten beinhalten Dashboards Schaltflächen zum Ausführen von Aktionen. Die Aktionen umschliessen das Teilen einer Ansicht und das Hinzufügen einer Annotation. Damit die Übersichtlichkeit der Dashboards durch diese Eingabeelemente nicht beeinträchtigt wird, kann ein Floating Action Button (FAB) die Aktionen zusammenfassen. FAB werden oft in Mobilapplikationen eingesetzt, um wichtige Aktionen schnell zu erreichen [43]. Abbildung 17 zeigt den Entwurf des ausgeklappten FAB.



Abbildung 17 Dashboard Layout - FAB User Actions

Der FAB schwebt über dem Inhalt des Dashboards, verdeckt aber keine Informationen. Bei einem Mausklick auf den Button erscheinen die enthaltenen Aktionen. Ein Vorteil dieses Konzepts ist, dass sich die Aktionen immer am selben Ort befinden und das Funktionsprinzip den meisten Personen bekannt ist.

### 3.6 Datenmanagement

#### 3.6.1 Mitwelten Datenbank

Die im Forschungsprojekt Mitwelten erfassten Daten sind in einer PostgreSQL Datenbank im Mitwelten Backend gespeichert. Abbildung 18 zeigt ein Diagramm des Datenbankschemas. Es ist auf die wichtigsten Tabellen reduziert.

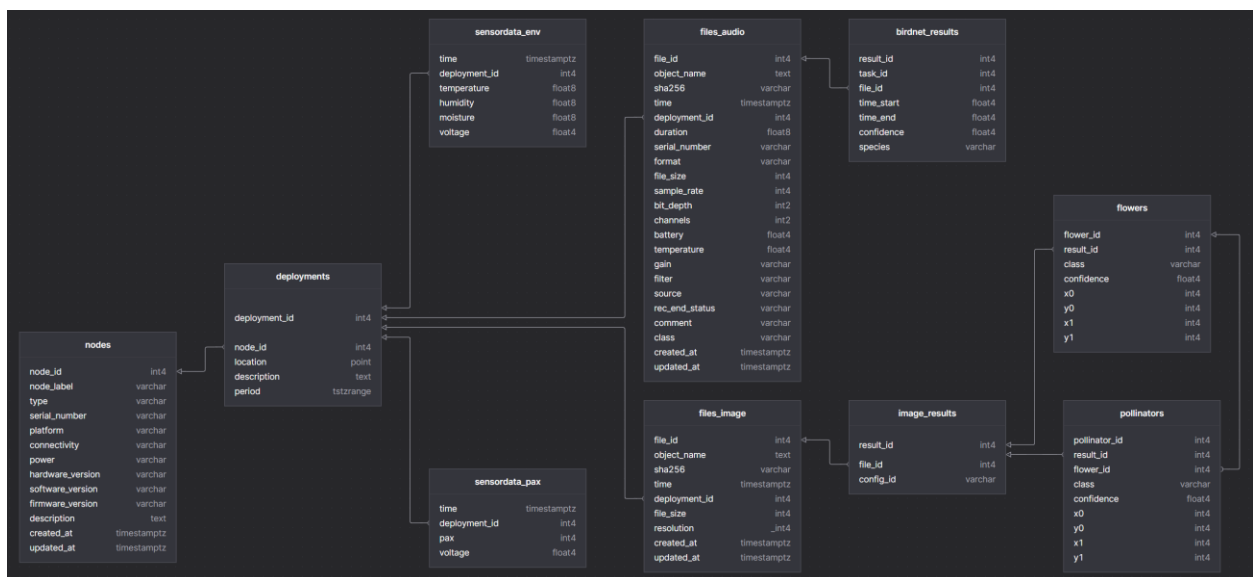


Abbildung 18 Datenbankschema Mitwelten

Im Zentrum des Schemas befindet sich die Deployment Table. Jeder Datenpunkt abhängig von einem Deployment, womit er erfasst wurde. Ein Deployment beinhaltet Informationen über den Ort und die Zeitperiode der Datenerfassung. Das Gerät, womit die Daten erfasst wurden, ist über die Table nodes referenziert.

#### 3.6.2 Daten API

Für die Bereitstellung der Daten aus der Mitwelten Datenbank wurde eine REST API mit FastAPI entwickelt. Die API dient als Schnittstelle für alle Applikationen des Projekts Mitwelten und soll auch für das zu entwickelnde Explore Tool erweitert werden. Folgend werden erforderliche funktionale Erweiterungen dokumentiert.

##### 3.6.2.1 Zeitliche Aggregation

Die zu visualisierenden Daten liegen in unterschiedlichen zeitlichen Granularitäten vor. PAX Counter mit LoRa Connectivity produzieren alle 15 Minuten einen Datenpunkt während das BirdNET Datenset mehrere Datenpunkte pro Sekunde enthalten kann. Aggregationen werden eingesetzt, um effiziente Datenbankabfragen zu ermöglichen und die Datenmenge zu reduzieren. Durch das Definieren eines Zeitfensters, zum Beispiel 60 Minuten, wird die Granularität festgelegt. Mittels einer Funktion zur statistischen Aggregation wird pro Zeitfenster einen Wert berechnet und als Resultat der Abfrage zurückgegeben.

Abbildung 19 zeigt die zeitliche Aggregation anhand eines Beispiels mit PAX Counter Werten. Die blauen Punkte zeigen die Rohdaten in der originalen Auflösung. Die rote Linie zeigt die

Summe der Werte pro Stunde. Durch die Aggregation wurden die knapp 12'000 originalen Datenpunkte auf 48 Datenpunkte zusammengefasst. Neben der Effizienzsteigerung lassen sich Muster und Trends besser erkennen.

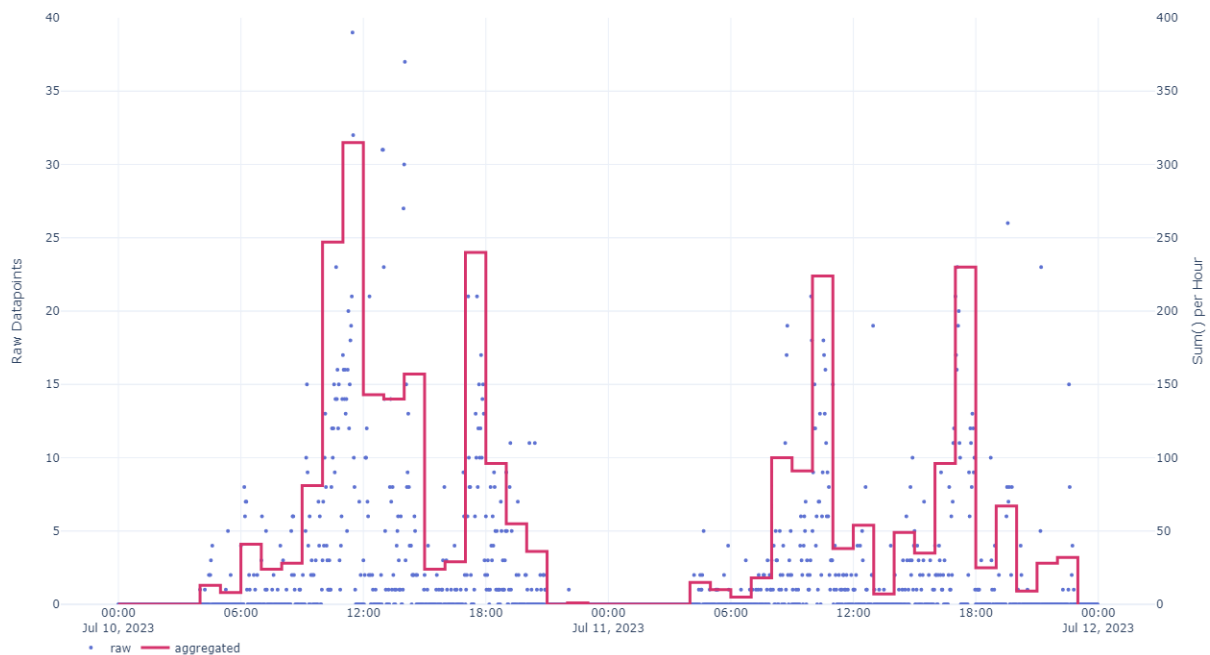


Abbildung 19 Zeitliche Aggregation von PAX Messwerten

Zur Ermöglichung von Abfragen der Daten in standardisiertem Format müssen sie zeitlich aggregiert werden. TimeScaleDB ist eine Erweiterung für PostgreSQL, die effiziente Abfragen und Aggregationen von Zeitreihendaten ermöglicht. Sie erstellt Hypertables mit zeitlich indexierten Inhalten.

### 3.6.2.2 Räumliche Aggregation

Neben der zeitlichen Aggregation können Daten anhand der räumlichen Achse aggregiert werden. Jeder im Projekt Mitweltten erfasste Datenpunkt referenziert auf ein Deployment, von welchem er erfasst wurde. Der Deployment Eintrag beinhaltet Informationen über den Ort und die Dauer der Datenerfassung. Für eine Aggregation nach Aufnahmeort werden Daten bei der Abfrage nach Deployment gruppiert. Pro Gruppe wird aus den Datenpunkten ein zusammengefasster Wert berechnet.

### 3.6.2.3 Aggregation nach Taxonomie

Datenpunkte von detektierten Vögeln und Bestäubern haben mit der Taxonomie ein kategorisches Attribut. Die Taxonomie ermöglicht es, Aussagen über die Ähnlichkeit von Arten zu machen. Eine Zusammenfassung von Beobachtungen nach Gattung, Familie oder Ordnung ermöglicht das Erkennen von Mustern auf einer höheren taxonomischen Ebene.

## 3.6.3 Caching

Eine Datenabfrage kann, abhängig von der Datenmenge und der Komplexität von Berechnungen in der Datenbank, eine gewisse Zeit in Anspruch nehmen. Wenn Abfragen wiederholt durchgeführt werden und sich die Daten nicht oft ändern, kann eine Zwischenspeicherung der Resultate zur Steigerung der Effizienz beitragen. Die Zwischenspeicherung wird als Caching bezeichnet und kann im Server oder beim Client erfolgen. Ein verbreiteter Ansatz ist das Speichern von Resultaten in einer zusätzlichen Caching-Datenbank. Die Resultate werden zusammen mit einem Timestamp zwischengespeichert. Bei einer erneuten Anfrage wird geprüft, ob die Antwort auf die

Anfrage bereits bekannt ist. Falls die Resultate in der Caching Datenbank verfügbar sind, werden sie direkt zurückgeschickt. Die zwischengespeicherten Antworten werden nur für eine bestimmte Zeit zwischengespeichert.

### 3.7 Datenanalyse

Im folgenden Abschnitt werden statistische Instrumente mit Relevanz in der zu entwickelnden Applikation beschrieben.

#### 3.7.1 Pearson Korrelationskoeffizient

Der Pearson Korrelationskoeffizient  $r$  ist ein Wert, der die Stärke eines linearen Zusammenhangs zwischen zwei Variablen repräsentiert. Im wesentlichen wird für die Berechnung eine Linie bestmöglich durch die Datenpunkte zweier Variablen zu ziehen. Der Korrelationskoeffizient liegt im Wertebereich  $[-1, 1]$  und beschreibt, wie weit die Punkte von der Linie entfernt sind. Bei einem starken linearen Zusammenhang zweier Variablen liegen die Punkte sehr nahe an der Linie. Ist der  $r$ -Wert grösser als 0, besteht ein positiver Zusammenhang zwischen den Variablen. Ist der  $r$ -Wert kleiner als 0, besteht ein negativer Zusammenhang; falls eine Variable grösser wird, wird die andere kleiner.

Für die Interpretation des  $r$ -Wertes wurden folgende Grundsätze formuliert [44]:

- $\{-0.3 \dots -0.1 \mid 0.1 \dots 0.3\}$  schwache Abhängigkeit
- $\{-0.5 \dots -0.3 \mid 0.3 \dots 0.5\}$  mittlere Abhängigkeit
- $\{-1 \dots -0.5 \mid 0.5 \dots 1\}$  starke Abhängigkeit

Um eine Aussage über die effektive Relevanz und Stärke der Korrelation zu machen, müssen Art und Qualität der Variablen berücksichtigt werden. Abbildung 20 zeigt Beispiele für starke positive Korrelation, starke negative Korrelation und Variablen ohne Zusammenhang.

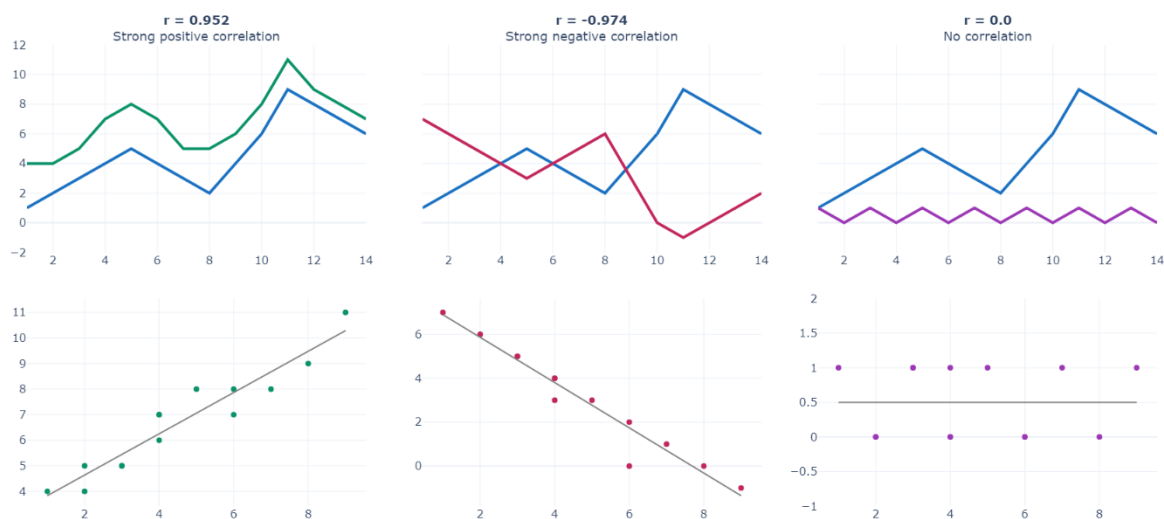


Abbildung 20 Pearson Korrelation

Bei der Berechnung des Korrelationskoeffizienten von Zeitreihendaten ist zu beachten, dass die zeitliche Auflösung übereinstimmt. Ausreisser in den Daten haben einen direkten Einfluss auf den  $r$ -Wert womit starke Verfälschungen auftreten können. Korrelationen von mehr als zwei Variablen werden oft in Form einer Korrelationsmatrix, welche die Koeffizienten zwischen allen Variablen beinhaltet, dargestellt.

### 3.7.2 Fourier Transformation

Ein Zeitreihendatenset besteht aus Messwerten, die in der Regel in gleichmässigen zeitlichen Abständen erhoben wurden. Viele Variablen, wie Lufttemperatur, Windgeschwindigkeit oder Aktivitäten von Tieren haben eine saisonale Komponente. Durch den Tag ist es wärmer als in der Nacht, im Sommer sind Bestäuber aktiver als im Winter. Saisonalitäten sind sich wiederholende Muster in Zeitreihendaten. Sie können täglich, jährlich oder in anderen Intervallen auftreten. Neben den regelmässigen Schwingungen beinhalten reale Messreihen Noise und Anomalien. Abbildung 21 visualisiert ein Signal im Zeitbereich (rot) und im Frequenzbereich (blau). Die Peaks in der Darstellung des Frequenzbereichs sind Saisonalitäten.

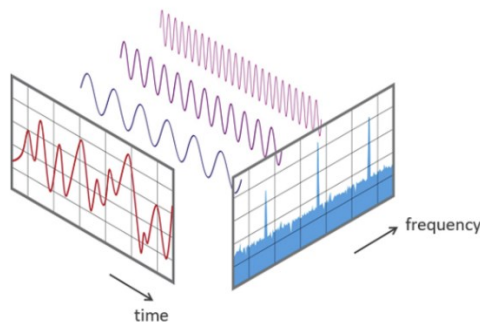


Abbildung 21 Darstellung eines Signals im Zeit- und Frequenzbereich [45]

Mit der diskreten Fourier Transformation (DFT) lassen sich saisonale Muster aus einer Zeitreihe (Signal) extrahieren. In einem ersten Schritt wird das Signal in ein Set von harmonischen Schwingungen aufgespalten. Für jede Schwingung werden anschliessend eine Amplitude und einen Phasenverschiebungswinkel berechnet, um damit das originale Signal zu rekonstruieren. Aus der Darstellung im Frequenzbereich, welche die Frequenzen und die Amplituden der harmonischen Schwingungen enthält, können Periodizitäten anhand von hohen Amplituden identifiziert werden.

Die Berechnung der DFT besteht aus vielen komplexen Operationen. Eine sehr verbreitete, effizientere Methode zur Abbildung von Zeitreihendaten im Frequenzbereich ist die Fast-Fourier-Transformation. Die FFT tastet das Signal immer in einer ganzzahligen Potenz zur Basis 2 ab. Als maximal bestimmbare Frequenz ist doppelt so gross wie die Sampling Rate der Zeitreihe.

Abbildung 22 zeigt Messwerte der Lufttemperatur, die über zwei Jahre im 10-Minuten Intervall erhoben wurden. Die jährliche Schwingung des Signals ist deutlich erkennbar, die tägliche Schwingung ist nicht direkt ersichtlich.

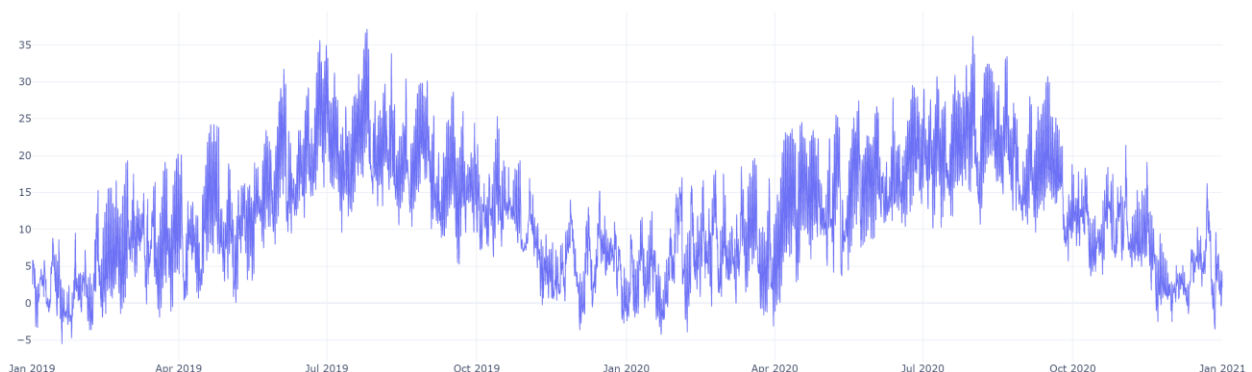


Abbildung 22 Zeitreihe: Temperaturwerte über zwei Jahre

Für die Darstellung der Temperaturwerte im Frequenzbereich wurde eine FFT berechnet, die in Abbildung 23 visualisiert ist. Der höchste bestimmbare Intervall entspricht 20 Minuten. Der in rot eingezeichnete Peak bei einem Intervall von einem Jahr repräsentiert die jährliche Saisonalität.

Der mit einem grünen Marker hervorgehobene Peak zeigt die tägliche Schwingung. Die Peaks bei 12h, 8h und 6h können als Aliasing Artefakte interpretiert werden.

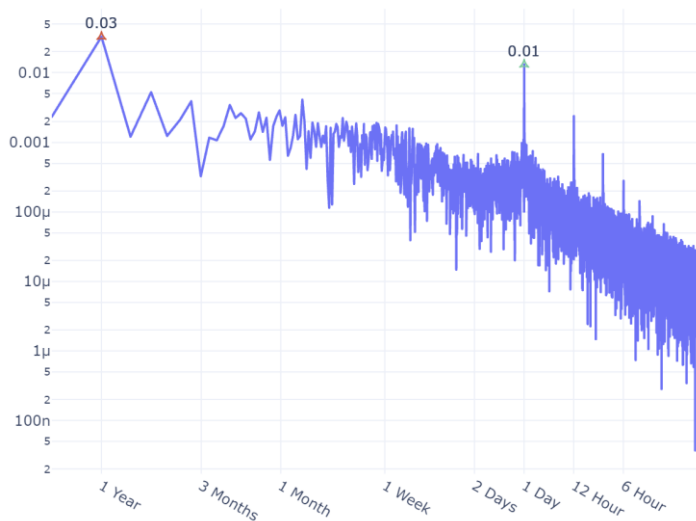


Abbildung 23 FFT von Temperaturwerten

Falls eine dominante Frequenz nicht genau einem Frequenz-Bin der FFT entspricht, wird die Stärke der Frequenzkomponente auf benachbarte Bins verteilt. Diese Verteilung erschwert das Erkennen von Peaks.

### 3.8 Plotly Dash

Dash ist ein Open Source Framework [46] von Plotly zur Entwicklung von interaktiven Webanwendungen zur Datenvisualisierung in Python. Im Vergleich zu den anderen analysierten Frameworks (Streamlit [47], Tableau [48]) bietet Dash die grösste Flexibilität. Folgend werden die wichtigsten Eigenschaften und Funktionen von Dash beschrieben.

#### 3.8.1 Architektur

Eine Dash Applikation ist ein Flask Webserver [49], der mit JSON Paketen über HTTP kommuniziert. Das Frontend der Applikation besteht aus gerenderten React.js Komponenten [50]. Die Interaktivität wird mittels Callback Funktionen implementiert.

#### 3.8.2 Callbacks

Ein Callback muss mindestens ein Input und ein Output beinhalten. Als Input, Output oder State werden Eigenschaften von Komponenten definiert. Die Callback Funktion wird bei der Änderungen der Input Komponente ausgelöst und aktualisiert die Output Komponente. State Komponente triggern das Callback nicht. Codeausschnitt 1 zeigt ein simples Beispiel der Implementierung einer Callback Funktion. In einem HTML Text wird angezeigt, wie oft auf den Button geklickt wurde.



```

app.layout = html.Div(
    [
        html.Button("Button", id="in_id", n_clicks=0),
        html.P(id="out_id"),
    ]
)

@app.callback(
    Output("out_id", "children"),
    Input("in_id", "n_clicks"),
)
def show_n_clicks(button_clicks):
    return f"the button was clicked {button_clicks} times"

```

Codeausschnitt 1 Callback Beispiel

Sobald auf den Button geklickt wird ändert sich die Eigenschaft im Callback Input und es wird ein POST Request an den Flask Server gesendet. Die Payload des Requests ist in Tabelle 2 auf der linken Seite ersichtlich. Sie beinhaltet Informationen über die getriggerte Input Komponente und über die Output Komponente, die zu aktualisieren ist. Die Liste 'changedPropIds' wird in Callback Funktionen mit mehreren Inputs eingesetzt, um zu kommunizieren, welcher davon getriggert wurde. Wenn im Callback neben Input und Output State Komponente definiert sind, werden deren Eigenschaften im Request mitgeschickt.

Die Response auf den Request ist in Tabelle 2 auf der rechten Seite abgebildet. Sie enthält die ID und die Attribute der zu aktualisierenden Komponente und den zu setzenden Wert. Der JavaScript Code beim Client setzt die Werte entsprechend der Response.

Request	Response
<pre> {   "output": "out_id.children",   "outputs": {     "id": "out_id",     "property": "children"   },   "inputs": [     {       "id": "in_id",       "property": "n_clicks",       "value": 1     }   ],   "changedPropIds": [     "in_id.n_clicks"   ] } </pre>	<pre> {   "multi": true,   "response": {     "out_id": {       "children":         "the button was clicked 1 times"     }   } } </pre>

Tabelle 2 Callback Requests

Callbacks, die schnell sein müssen und keine Interaktion mit dem Backend erfordern, können als Clientside Callbacks implementiert werden. Identisch zu den normalen Callbacks werden Inputs, Outputs und optional States definiert. Die Callback Funktion selbst wird anstelle von Python Code in JavaScript geschrieben. Codeausschnitt 2 zeigt die als Clientside Callback implementierte Callback Funktion aus Codeausschnitt 1. Längere JavaScript Funktionen können in separaten Files

definiert werden, um die Leserlichkeit des Codes zu erhöhen. Bei der Clientside Implementierung werden keine Requests an das Backend gesendet, was schnellere Updates ermöglicht und die Performance der Applikation steigert.

```
clientside_callback(
    """
    function show_n_clicks(n_clicks) {
        return `the button was clicked ${n_clicks} times`;
    }
    """, Output("out_id", "children"), Input("in_id", "n_clicks"),
)
```

Codeausschnitt 2 Clientside Callback Beispiel

### 3.8.3 Components

Das Frontend einer Dash Applikation besteht aus mehreren Komponenten, die gerendert und dargestellt werden. Dash stellt verschiedene Open Source Component Libraries bereit, darunter HTML Components und Core Components. Dash HTML Components beinhaltet Wrapper für die meisten HTML Elemente, um auf HTML Templating zu verzichten. Dash Core Components ist eine Sammlung von interaktiven Komponenten wie Dropdowns, Sliders, Date Pickers und der Graph Komponente. Neben sichtbaren Komponenten existieren Wrapper für den Zugriff auf den Browser Speicher oder für die Aktualisierung der URL.

Eine Vielzahl von Dash Komponenten wurden von der Community in Open Source Projekten entwickelt. Zu den populärsten Community Component Libraries zählen Dash-Bootstrap-Components [51], welche die wichtigsten Bootstrap Komponente bereitstellt und Dash-Mantine-Components [52], eine Library mit über 70 Mantine React Komponenten [53]. Plotly stellt eine Anleitung und Templates für die Entwicklung von eigenen Komponenten bereit.

All-In-One (AIO) Komponente können zum Einsatz kommen, falls eine Zusammensetzung von abhängigen Komponenten mehrfach in einer Applikation vorkommen. Ein Beispiel für eine AIO-Komponente ist eine Listenansicht mit Seitenanzeige, die geblättert oder mit einem Suchfeld gefiltert werden kann. Eine AIO Komponente für die Listenansicht besteht aus einem Layout, das die Grundkomponente und ihre Positionen definiert, und den entsprechenden Callbacks für das Blättern und die Suchfunktion. Insbesondere für Komponenten mit komplizierten Callbacks kann die Entwicklung einer AIO Komponente die Übersichtlichkeit der Applikation erhöhen.

Für alle Dash Komponenten, die als Callback Argumente eingesetzt werden, muss eine Component-ID gesetzt werden. Bei grösseren Applikationen ist es sinnvoll, die IDs zu generieren, um zufällige Duplikate zu vermeiden.

### 3.8.4 Datenvisualisierung

Die Graph Komponente von Dash-Core-Components dient als Wrapper für Datenvisualisierungen. Sie unterstützt alle Plotly-gestützten Charts und stellt User Interaktionen als mögliche Callback Argumente bereit. Interaktionen werden über vier Attribute bereitgestellt:

- Hover Data beinhaltet Datenpunkte, welche sich unter dem Mauszeiger befinden.
- Click Data beinhaltet Datenpunkte, auf die geklickt wurde.
- Selection Data beinhaltet alle Datenpunkte, die mit der Select Option ausgewählt wurden.
- Zoom / Relayout Data beinhaltet Informationen über Layoutänderungen.

Das Erscheinungsbild der Charts, die Sichtbarkeit der Toolbar und weitere Eigenschaften können über die 'config' Attribute modifiziert werden.



### 3.8.5 Layout und Style

Für das Layout der Applikation können externe Stylesheets wie Bootstrap Themes eingesetzt werden. Jede Komponente hat ein Attribut 'style', worin der Style mit CSS Parametern modifizierbar ist. Styles von mehrfach verwendeten Elementen können in einem CSS File definiert werden. Im Rahmen des Open Source Projekts Dash-Bootstrap-Templates [54] wurde eine AIO Komponente für das Umschalten zwischen Hell- und Dunkelmodus entwickelt. Komponente wie Charts, die unabhängig vom Stylesheet gerendert werden, müssen durch entsprechende Callbacks mit dem ThemeSwitch als Input neu gerendert werden.

### 3.8.6 Pages

Dash unterstützt Multi Page Applikationen, wobei jede Page in einem eigenen File definiert wird. Für jede Page ist die Definition des Layouts erforderlich, welches alle Komponente der Page beinhaltet. Callbacks für die Interaktionen in einer Page befinden sich ebenfalls im Page File, was die Strukturierung der Applikation vereinfacht. Über die Funktion 'register\_page' wird eine Page mit dem entsprechenden Path Prefix in der Page Registry erfasst. Für jede Anfrage wird anhand der URL den jeweiligen Content geladen.

## 4 Implementierung

Dieses Kapitel beschreibt die Implementierung der Applikation.

### 4.1 Architektur

Die Mitwelten Explore App läuft in einem Docker Container im Mitwelten Backend. Im Container läuft eine FastAPI [55] Applikation, die Uvicorn [56] als Server nutzt. Die Dash Applikation wurde als WSGI Middleware an die Route '/app' gemountet und ist nur für eingeloggte Benutzer und Benutzerinnen aufrufbar. Die Landing Page wird von FastAPI auf der Route '/' bereitgestellt und ist öffentlich. Alle relevanten Bestandteile des Systems sind in Abbildung 24 ersichtlich und werden folgend beschrieben.

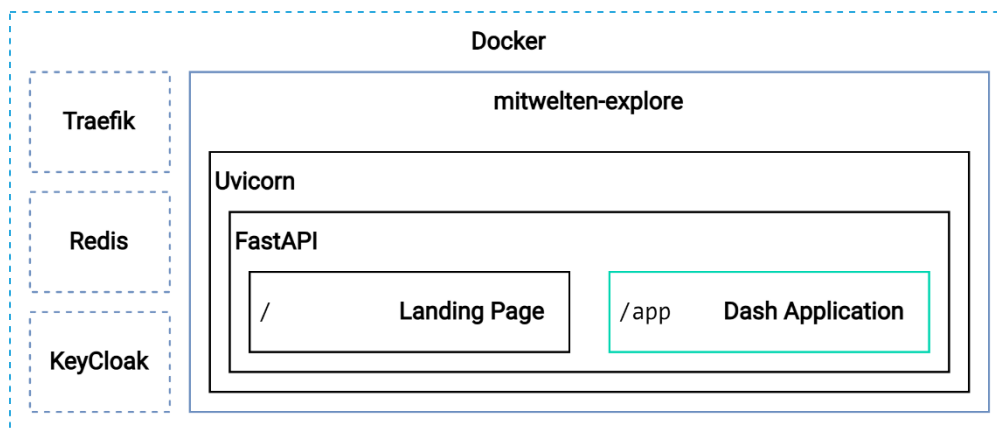


Abbildung 24 Architektur Mitwelten Explore

#### 4.1.1 FastAPI

Die API der Applikation wurde mit FastAPI entwickelt. Zu den Hauptaufgaben der API gehört die Abwicklung von Authentifizierungsprozessen und die Bereitstellung der statischen Landing Page. Als nicht eingeloggte Person ist nur die Landing Page zugänglich.

Eine Middleware validiert die Autorisierung aller Anfragen an die geschützten Routen. Ist eine Person nicht autorisiert, wird sie auf die Login Route weitergeleitet. Anfragen an die Login Route werden an KeyCloak weitergeleitet, damit eine Anmeldung oder eine Registrierung erfolgen kann.

Nach einer erfolgreichen Anmeldung leitet KeyCloak den Client auf die Explore Applikation zurück, wo ein Auth-Cookie gesetzt wird. Das Cookie wird in jedem Request automatisch mitgeschickt und von der Middleware validiert. Das Cookie wird gelöscht, sobald die Logout Route aufgerufen wird.

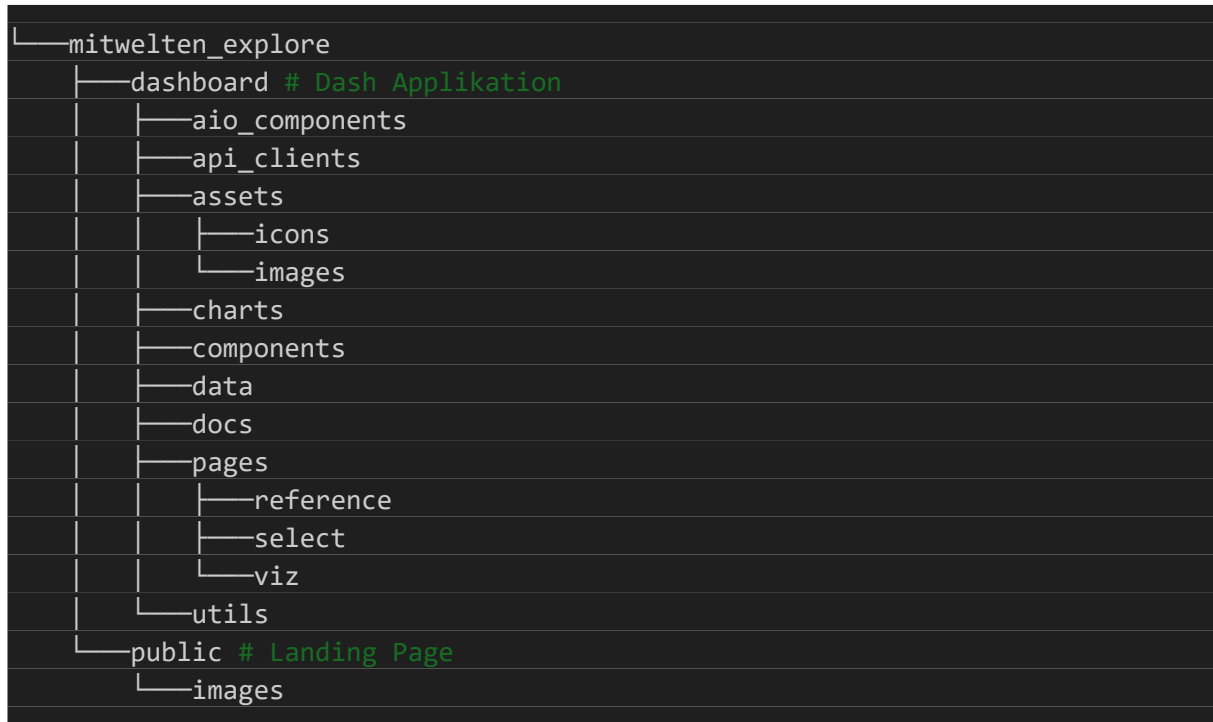
Ein ausgestellter Access Token hat eine Gültigkeitsdauer von 15 Minuten. Um eine erneute Anmeldung alle 15 Minuten zu verhindern wurde eine automatisierte Erneuerung des Tokens implementiert. Bei jedem Request an eine geschützte Route wird berechnet, wie lange der Token noch gültig ist. Wenn ein konfigurierbares Limit erreicht ist, wird der Token automatisch erneuert.

Die Files der statischen Landing Page befinden sich im Verzeichnis 'public'. Durch den Einsatz einer 'StaticFiles' Middleware werden Nutzende, die nicht eingeloggt sind, auf die Landing Page weitergeleitet.

#### 4.1.2 Dash Applikation

Die Dash Applikation wurde als WSGI Middleware unter der Route '/app' gemountet. Die Ordnerstruktur der gesamten Applikation ist in Codeausschnitt 3 ersichtlich. Die Dash Applikation

befindet sich im Verzeichnis 'dashboard'. Die Komponente der Applikation wurden in verschiedene Module gegliedert. Die wichtigsten Module werden nachfolgend dokumentiert.



Codeausschnitt 3 Ordnerstruktur Mitwelten Explore

Für jede Seite der Applikation existiert ein Python File im Ordner 'pages'. Die Namen der Dateien werden für das Routing verwendet. Ein Aufruf der Route '/app/annotations' zeigt das Layout der Seite in 'annotations.py' an. Dateien in Unterordnern sind unter dem entsprechenden Präfix erreichbar.

Für die Applikation wurden mehrere All-In-One Komponente entwickelt. Sie bestehen aus mehreren Dash Komponenten mit internen Callbacks und funktionieren wie eine einzelne Komponente. Die entwickelten Module befinden sich im Ordner 'aio\_components'.

Im Ordner 'components' befinden sich weitere Elemente, die mehrfach eingesetzt werden. Die Elemente sind kategorisch in mehrere Files unterteilt. Der Ordner 'charts' enthält Elemente zur Datenvisualisierung.

Die Tool-Interne Dokumentation besteht aus mehreren Markdown Files. Sie sind im Verzeichnis 'docs' gespeichert und werden für die Anzeige gerendert. Alle in der Dokumentationsseite angezeigten Bilder befinden sich im Verzeichnis 'assets'. Funktionen für Datenabfragen wurden auf mehrere Files im Ordner 'api\_clients' aufgeteilt.

Der Source Code und eine Anleitung zum Erstellen eines Docker Images sind auf GitHub verfügbar [57].

## 4.2 Navigation

Eine Navigationsleiste im oberen Bereich der Applikation dient zum Navigieren zwischen den Seiten. Der Name des Tools und das Mitwelten Logo auf der linken Seite ist auf die Startseite des Tools verlinkt. In der Mitte der Leiste sind die Abschnitte des Tools verlinkt. Die Anordnung entspricht dem Workflow des Tools. Damit es für die Benutzer und Benutzerinnen zu jedem Zeitpunkt klar ist, in welchem Bereich der Applikation sie sich befinden, wird die aktive Seite mit einem blauen Balken unterstrichen.

Die Seiten *Collection*, *Annotations* und *Docs* sind direkt verlinkt. Neben dem Link zu der Collection wird die Anzahl gespeicherter Datensets mit einem Badge angezeigt. Manipulationen der Collection werden durch Änderungen dieser Zahl direkt bestätigt.

In den Abschnitten *Datensets* und *Dashboards* gibt es mehrere Unterbereiche. Eine möglichst schnelle Navigation wird durch den Einsatz von Dropdown Elementen ermöglicht, die auf die Unterbereiche weiterleiten. Mit Pfeilen wird kommuniziert, dass sich hinter den Schaltflächen mehrere Optionen befinden. Ein Dropdown wird geöffnet, sobald sich der Mauszeiger über dem Element befindet.

Auf der rechten Seite der Leiste wird die eingeloggte Person mit einem Avatar und dem Namen angezeigt. Der Nutzerbereich beinhaltet ein Link zu den erstellten Annotationen, einen Knopf zum löschen der Collection Einträge und die Logout Funktion.

Abbildung 25 zeigt die Navigationsleiste mit ausgeklappten Dropdowns. Der Abschnitt Datasets ist hervorgehoben und weist auf den momentan angezeigten Abschnitt hin.

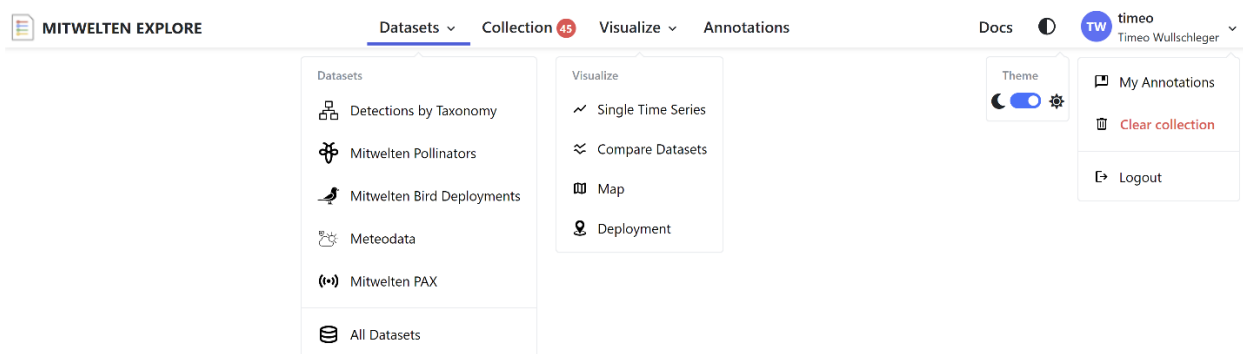


Abbildung 25 Navigationselemente Explore Tool

## 4.3 Datenauswahl

Die Datenauswahl zum Filtern und Zusammenstellen von Datensets wurde, strukturiert nach Kategorie, in mehrere Pages gegliedert.

Datensets sind Referenzen auf Mengen von Datenpunkten, die im Projekt Mitwelten oder von externen Quellen erhoben wurden. Diese Referenzen werden durch Datenset Objekte im JSON Format repräsentiert. Jedes Datenset Objekt enthält den Datenset-Typ und alle Parameter für die Datenabfrage. Zusätzlich sind Informationen wie Namen, Einheit und Standort enthalten, um ein Datenset in der Collection aufzulisten.

### 4.3.1 Datenkategorien

Eine Referenz auf eine Teilmenge eines Datensatzes wird als Datenset bezeichnet. Im Tool werden folgende Kategorien von Datensets unterschieden:

- **Taxon:** Die Anzahl Detektionen einer *Art, Gattung, Ordnung, Familie, Klasse, Stamm* oder *Königreich* in den im Forschungsprojekt Mitwelten erhobenen Daten.
- **Pollinator:** Die Anzahl detektierter Bestäuber aus der Mitwelten Bestäuber Studie.
- **Bird Diversity:** Die Anzahl verschiedener Spezies aus den Mitwelten BirdNET Auswertungen.
- **Meteodata:** Meteodaten von MeteoSchweiz.
- **PAX:** Messwerte der PAX Sensoren.
- **ENV Measurements:** Temperatur, Luftfeuchtigkeit und Bodenfeuchtigkeit, die von den Environment Sensor Nodes aufgenommen wurden.
- **Taxon (GBIF):** Beobachtungen einer *Art, Gattung, Ordnung, Familie, Klasse, Stamm* oder *Königreich* aus externen Quellen, die von GBIF bereitgestellt werden.
- **Location Characteristics:** Merkmale von Ortschaften.

In folgenden Unterkapiteln wird die Auswahl eines Datensets für jede Kategorie beschrieben.

### 4.3.2 Taxa

Die Auswahl eines Taxa Datensets ist über die Schaltfläche Datasets → Detections by Taxonomy erreichbar. Die Seite ist in Abbildung 26 ersichtlich. Eine durchsuchbare Liste enthält alle bekannten Spezies. Über die RadioGroup oberhalb der Liste kann die Taxonomie Ebene geändert werden, um nach Gattungen, Ordnungen und weiteren Ebenen zu suchen.

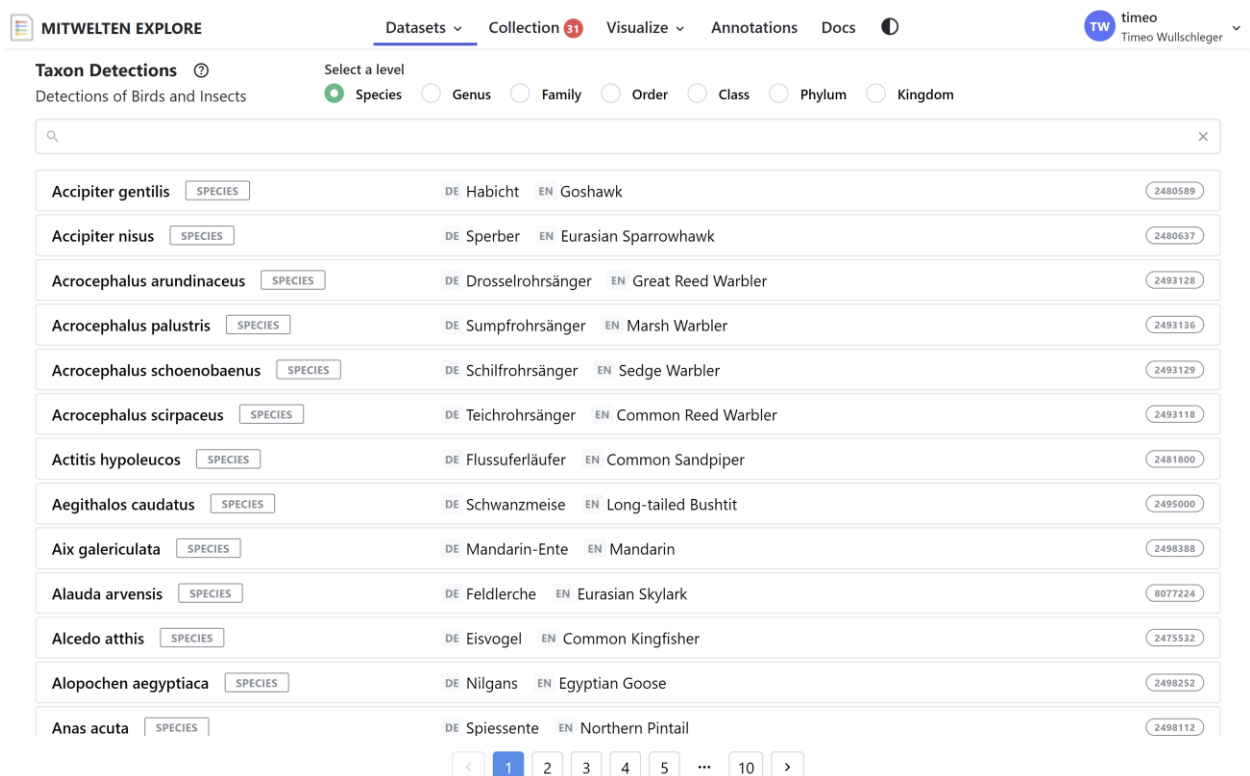


Abbildung 26 Datenauswahl - Taxa

Wird auf ein Listenelement geklickt, erscheint ein Popup, welches die wichtigsten Informationen über diese Art beinhaltet. Abbildung 27 zeigt das Element nach Anklicken der Art Mauersegler. Im Header befinden sich die Namen des Taxon und Informationen über die Taxonomie. Rechts neben den Namen befinden sich Indikatoren, welche die Anzahl Detektionen im Mitwelten Projekt

und die Anzahl Beobachtungen aus externen Quellen darstellen. Unterhalb der Headers wird ein Referenzbild und eine kurze Beschreibung des Taxons von Wikipedia angezeigt. Im unteren Bereich des Popups sind Schaltflächen angeordnet. Für jedes Taxon existieren zwei mögliche Datensets. Das Mitwelten Detections Datenset beinhaltet nur die Detektionen aus dem Projekt Mitwelten. Das GBIF Observation Datenset beinhaltet alle Beobachtungen des Taxons, die von externen Quellen gemacht wurden. Die dritte Schaltfläche dient als Direktlink auf das Taxon Dashboard.

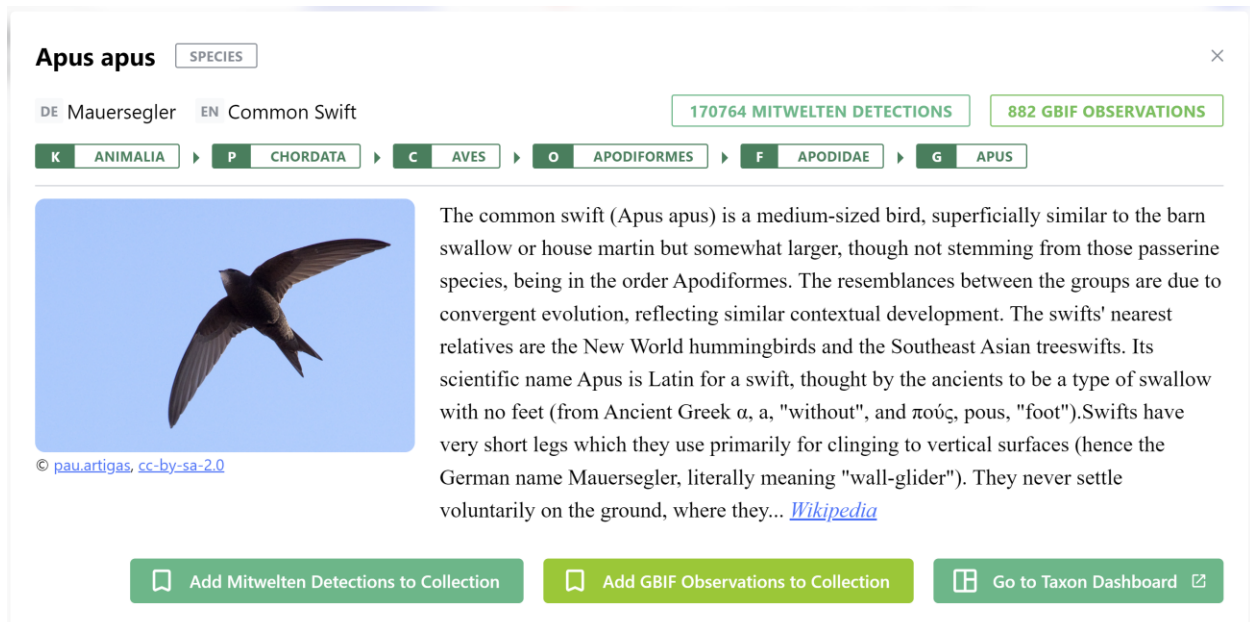


Abbildung 27 Datenauswahl - Taxa: Info Modal

### 4.3.3 Pollinators

Bestäuber Datensets können in der Kategorie Mitwelten Pollinators zusammengestellt werden. Die Seite ist in Abbildung 28 ersichtlich. Im Gegensatz zu der Auswahl eines Taxon Datensets wird das Datenset konstruiert und nicht ausgewählt. Im linken Bereich befinden sich Filter, um die Daten einzugrenzen. Im rechten Bereich werden Informationen über das Datenset dargestellt. Die Standardeinstellung beinhaltet alle Detektionen von allen Deployments.

Über das Eingabefeld 'Deployment' können einzelne oder mehrere Deployments ausgewählt werden. Durch das Selektieren von Deployments wird das Datenset gemäss der Auswahl eingegrenzt. Unterhalb dieses Eingabeelements befindet sich eine interaktive Karte, die alle Deployments enthält. Durch Klicken auf einen Marker kann ein Deployment an- oder abgewählt werden. Durch die Option 'Box Select', ist es möglich, mehrere Punkte auf einmal anzuwählen. Die Option ist über das Quadratische Symbol in der Toolbar der Karte erreichbar.

Die Kategorien von Bestäubern, die im Datenset enthalten sind, lassen sich durch das Eingabelement 'Pollinator Class' eingrenzen. In Abbildung 28 besteht das Datenset aus Beobachtungen von Schwebfliegen die von zwei Deployments erfasst wurden. Diese Informationen sind im rechten Bereich ersichtlich. Für jedes Deployment wird zusätzlich die Zeitperiode der Datenerfassung angezeigt.

Über das Bookmark Symbol kann ein konstruiertes Datenset in der Collection gespeichert werden. Für eine schnelle Überprüfung der Daten kann das Datenset über die Schaltfläche 'Viz' im Timeseries Dashboard geöffnet werden.

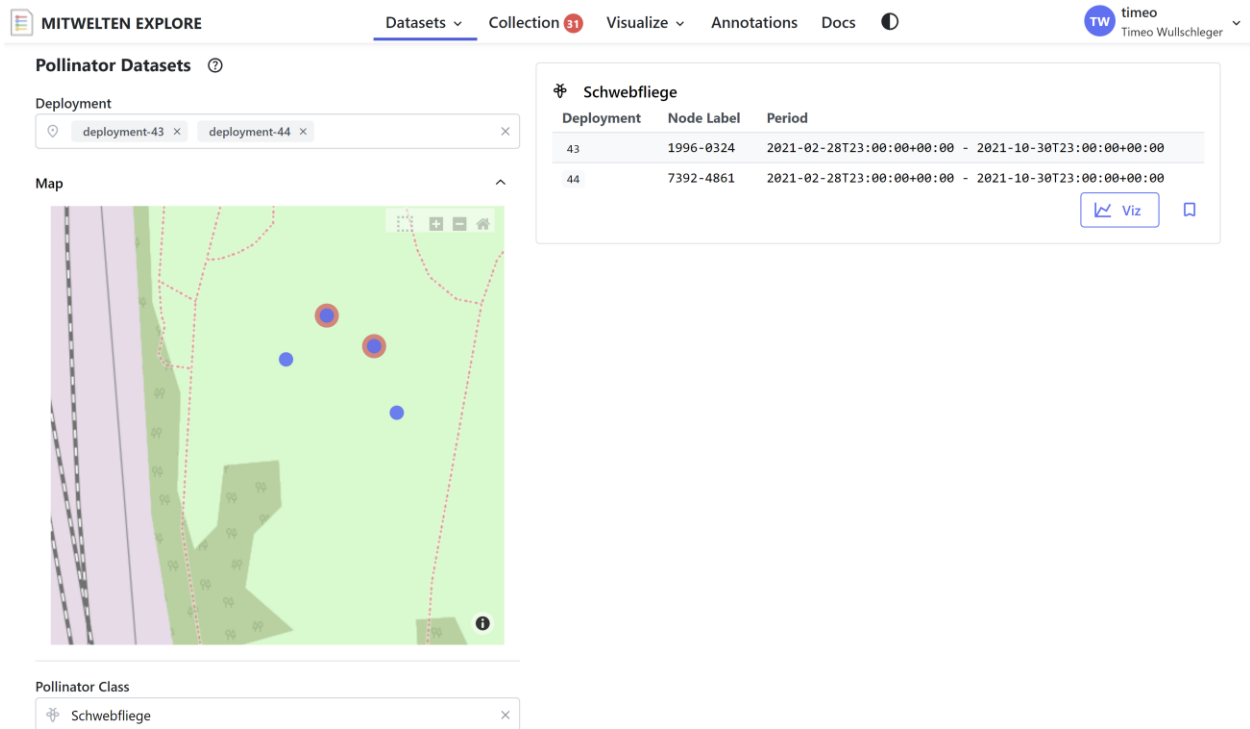


Abbildung 28 Datenauswahl - Pollinators

#### 4.3.4 Bird Diversity

Datensets der Kategorie 'Bird Diversity' repräsentieren die Anzahl verschiedener Arten, die an definierten Orten erkannt wurden. Abbildung 29 zeigt die Seite zur Datenauswahl. Das Funktionsprinzip ist bis auf die Eingrenzung der Kategorie identisch zu der Auswahl eines Bestäuber Datensets. Zur schnellen Überprüfung der Daten ist das Deployment Dashboard verlinkt.

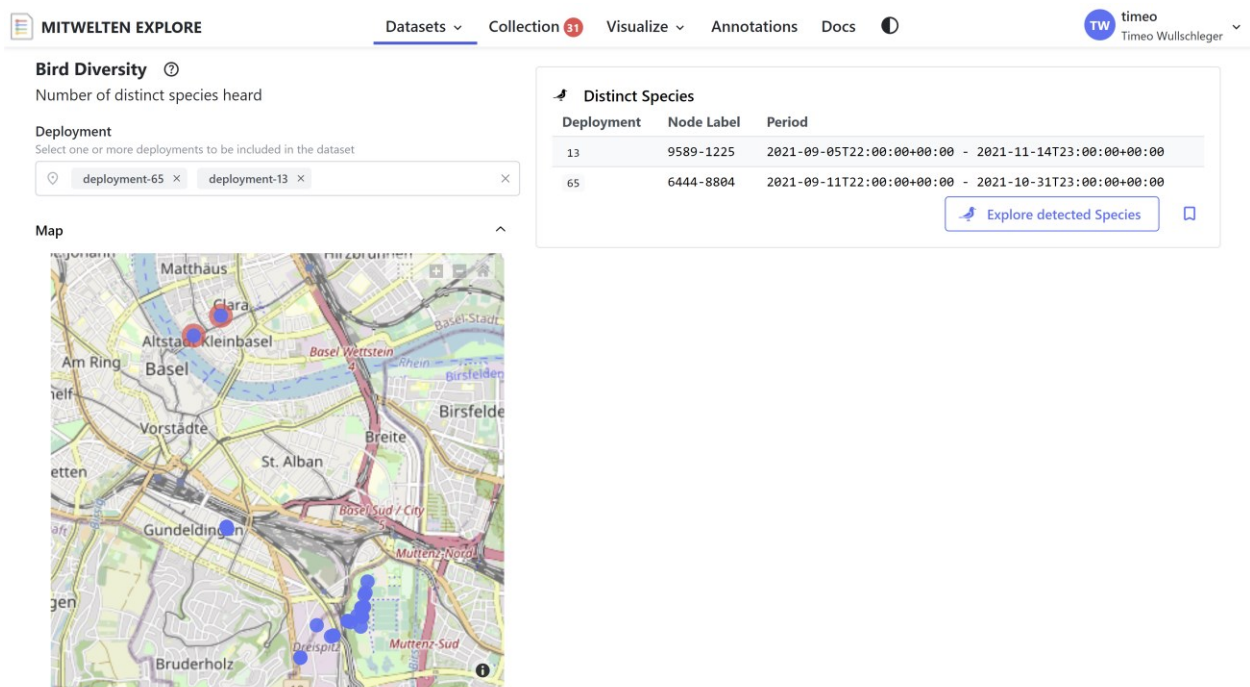


Abbildung 29 Datenauswahl - Bird Diversity



### 4.3.5 Meteodata

Die von MeteoSchweiz bezogenen Meteodaten können in der Applikation unter Datasets → Meteodata ausgewählt werden. Die Auswahlseite ist in Abbildung 30 ersichtlich. Auf der rechten Seite befindet sich eine Liste mit allen vorhandenen Datensets. Im linken Bereich befinden sich Filter zum einschränken der Datensets. Durch anklicken eines Markers in der Karte oder durch die Auswahl einer Station im Eingabefeld 'Meteo Station' kann eine Station ausgewählt werden. Die Einheit der Messwerte kann über das Eingabefeld 'Unit' festgelegt werden.

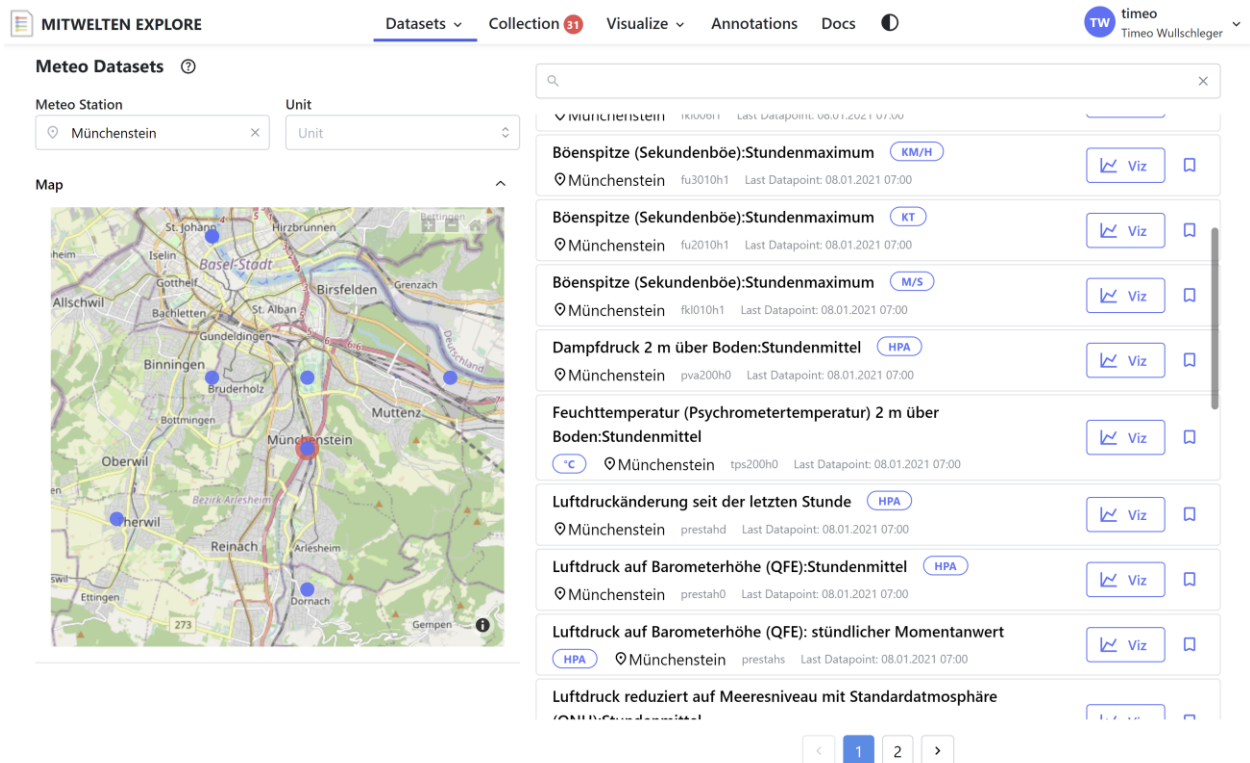


Abbildung 30 Datenauswahl – Meteodaten

Gewisse Datensets wurden von MeteoSchweiz aggregiert. Die Methode ist im Dataset Namen enthalten. Gewisse Messreihen sind auf MeteoSchweiz nur von gewissen Zeitperioden verfügbar, die nicht unbedingt aktuell sind. In der Auflistung der Datensets wird jeweils der Zeitpunkt der neusten Datenpunktes angezeigt.

### 4.3.6 PAX

Ein PAX Datenset repräsentiert die Messwerte eines Deployments. Abbildung 31 zeigt die Auswahlseite für PAX Datensets. Ein Datenset repräsentiert ein Deployment. Wie bei den anderen Auswahlseiten befindet sich auf der rechten Seite eine durchsuchbare Liste, die alle vorhandenen Datensets anzeigt. Über die Filter im linken Bereich kann ein Deployment ausgewählt werden.

Jedes aufgelistete Datenset enthält neben der Deployment ID weitere Informationen über das Deployment. Informationen zur Connectivity und Plattform des Nodes sind als Badges abgebildet. Die Beschreibung des Deployments enthält in vielen Fällen weitere Informationen zum Standort. Weiter werden Start- und Endzeitpunkt der Einsatzperiode dargestellt.

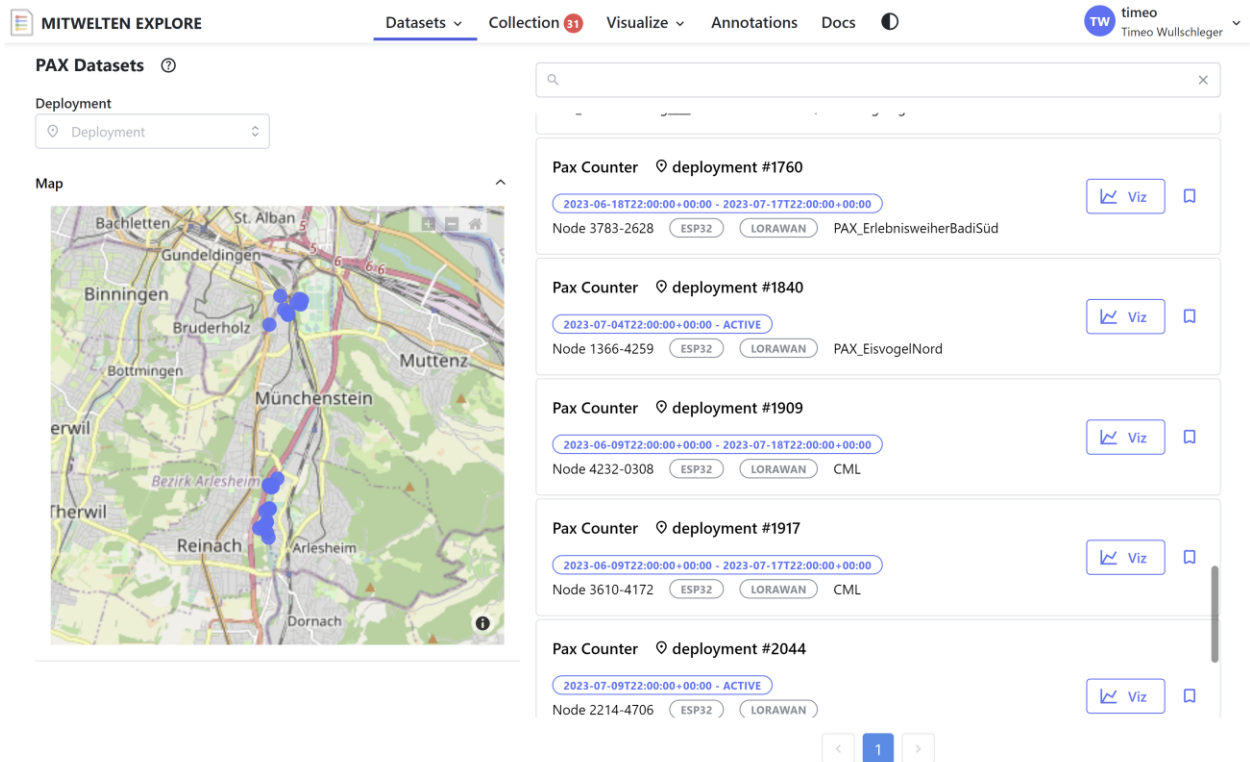


Abbildung 31 Datenauswahl - PAX

### 4.3.7 ENV Measurements

Die Environment Sensor Nodes wurden nur im Jahr 2021 eingesetzt. Die Datenqualität ist in vielen Fällen sehr niedrig, womit die Bedeutung der Datensets in der Applikation nicht sehr hoch ist. Abbildung 32 zeigt die Seite zur Auswahl von Environment Sensor Messwerten. Der Aufbau ist identisch zu der Auswahlseite für PAX Datensets. Für jedes Deployment existieren drei Datensets.

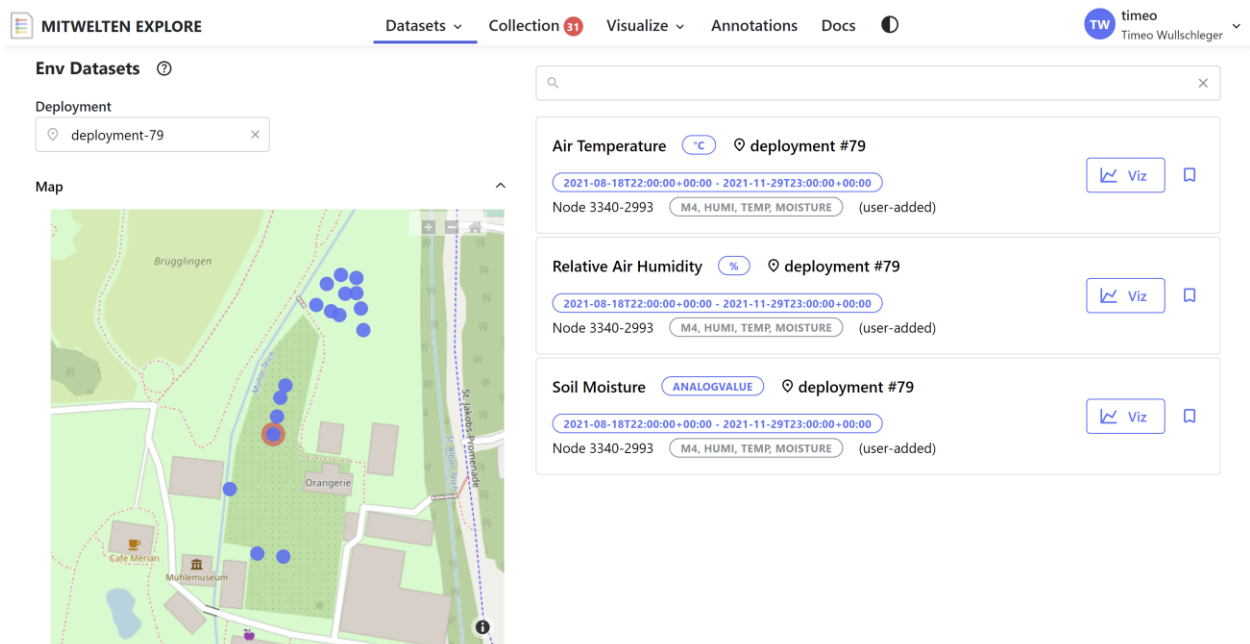


Abbildung 32 Datenauswahl - ENV Measurements

### 4.3.8 Feedback

User Actions sollten bestätigt werden, um mitzuteilen, ob die Aktion erfolgreich war oder um zu informieren, wenn etwas nicht funktioniert hat. Für diese Rückmeldungen wurden Notifications entwickelt, die im unteren rechten Bildschirmrand erscheinen und nach fünf Sekunden automatisch verschwinden. Es ist möglich, sie manuell zu schliessen. Notifications, die über eine Modifikation der Collection informieren, sind in Abbildung 33 ersichtlich.

Eine erfolgreiche Speicherung eines Datensets in der Collection wird mit einem grünen Haken symbolisiert. Das Label enthält den Namen des Datensets. In Abbildung 33 wurden zwei PAX Datensets erfolgreich gespeichert. Wird versucht, ein Datenset, welches bereits in der Collection enthalten ist, erneut hinzuzufügen, erscheint eine Informationsmeldung. Der Hinweis ist in Abbildung 33 an dritter Stelle ersichtlich. Eine Bestätigung für die Entfernung eines Datensets aus der Collection ist an letzter Stelle ersichtlich.

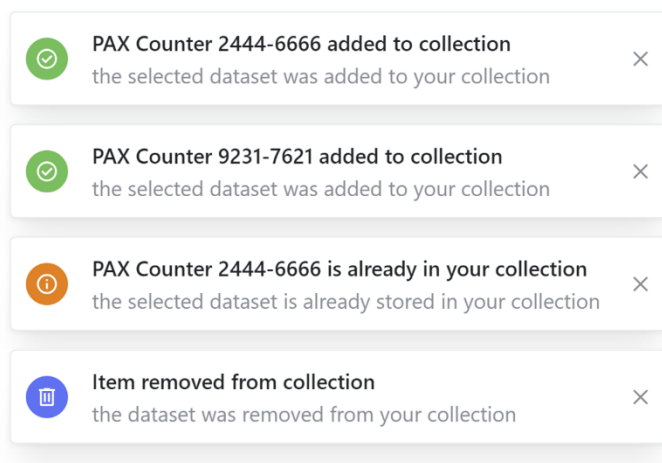


Abbildung 33 Notifications Collection Manipulation

Es können mehrere Meldungen gleichzeitig angezeigt werden. Die aktuellsten Meldungen werden zu unterst eingefügt.

## 4.4 Collection

Die Collection ist eine mit dem Backend synchronisierte Liste von Datenreferenzen, die von der angemeldeten Person gespeichert wurden. Die gespeicherten Datensets werden nicht mit den anderen Nutzenden geteilt. Die gespeicherten Datensets sind zeitunabhängig. In der Navigationsleiste des Tools wird angezeigt, wie viele Elemente sich in der Collection befinden. Die Synchronisierung der Collection mit dem Backend erfolgt bei jeder Modifikation des Inhaltes und immer, wenn die Seite neu geladen wird. Somit ist es möglich, mit mehreren Geräten gleichzeitig zu arbeiten.

Über die Route '/collection' können die gespeicherten Datensets verwaltet werden. Die Einträge werden in Form einer durchsuchbaren Liste aufgelistet. Für jedes Datenset sind die kompatiblen Dashboards verlinkt. Über die Schaltfläche mit dem Papierkorb Symbol können einzelne Einträge aus der Collection entfernt werden. Die Entfernung des gesamten Collection Inhalts ist über die Nutzer-Aktionen in der Navigationsleiste möglich.

Ein Screenshot der Collection Page ist in Abbildung 34 ersichtlich. Für jedes Datenset sind Namen, Einheit und Erfassungsort dargestellt. Bei Datensets mit mehreren Standorten werden Listen der entsprechenden Deployments angezeigt. Damit die Übersichtlichkeit der Ansicht nicht beeinträchtigt wird, sind nur die ersten zwei Standorte sichtbar. Die Schaltfläche 'Show All' expandiert die Liste. Die Kategorie der jeweiligen Datensets wird mit Icons kommuniziert.

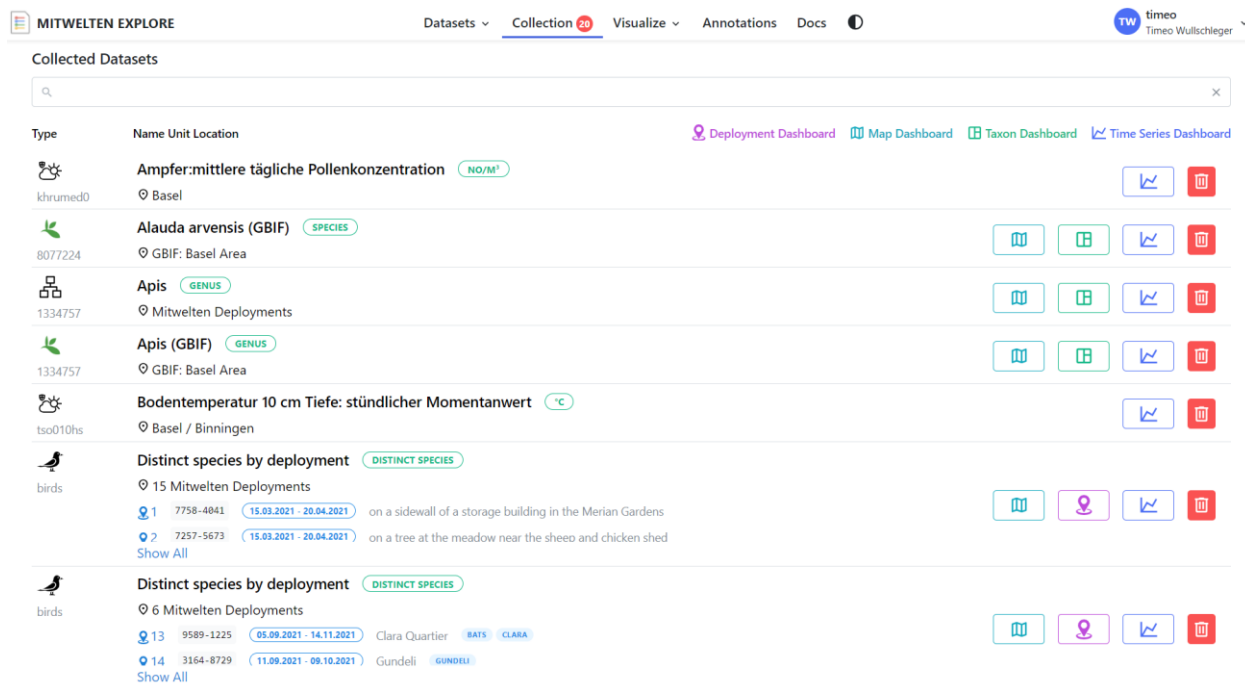


Abbildung 34 Screenshot Collection

## 4.5 Dashboards

Für die explorative Visualisierung und Analyse der Daten wurden mehrere Dashboards entwickelt. In folgenden Unterkapiteln werden sie dokumentiert.

### 4.5.1 Bedienung der Dashboards

Bei der Entwicklung der Bedienkonzepte für die Applikation wurde eine interne Konsistenz angestrebt. Die Anordnung von Bedienelementen und das Funktionsprinzip ist für alle Dashboards identisch.

#### 4.5.1.1 Zeitspanne

In jedem Dashboard werden Daten über einen bestimmten Zeitbereich visualisiert. Das Element zur Auswahl der Zeitspanne befindet sich in allen Dashboards im oberen rechten Bereich. Abbildung 35 zeigt das Bedienelement mit ausgeklapptem Date Picker.

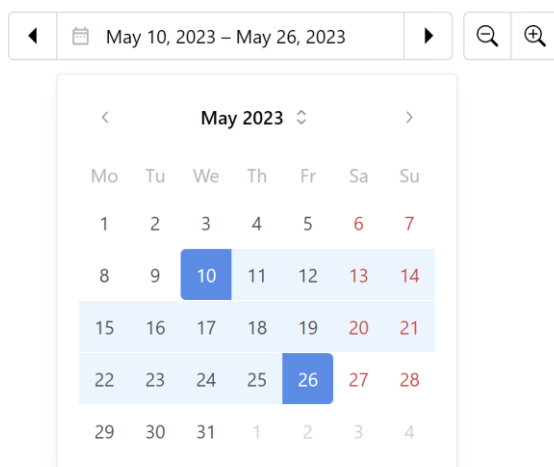


Abbildung 35 Eingabe des anzuzeigenden Zeitraums

Es besteht aus einem Date Range Picker und zusätzlichen Schaltflächen für die Manipulation der Werte. Mit den Pfeilen wird das Zeitfenster nach Links oder Rechts verschoben. Die Schrittweite beträgt 15 % der gesamten Zeitspanne. Mit den Zoom Icons wird das Fenster symmetrisch verkleinert oder vergrößert. Die zusätzlichen Schaltflächen werden nur angezeigt, falls genügend Platz auf dem Bildschirm vorhanden ist.

Anwendungen zur explorativen Datenanalyse, wie Grafana [58], verwenden oft die Zoom Funktion der Charts für die Manipulation der globalen Zeitspanne. Die Aktualisierung des globalen Zeitfensters hat den Vorteil, dass alle anderen Charts auch aktualisiert werden und mit dem Haupt Chart übereinstimmen. Diese Funktionalität ist in den Dashboards mit der Zeit als Hauptachse standardmässig aktiviert. Oberhalb des Zeitreihen Charts befindet sich jeweils ein Schalter, über welchen die Funktion deaktiviert werden kann. Abbildung 36 zeigt den Schalter im aktivierten Zustand, welcher bei jeder Interaktion mit dem Chart die Time Range anpasst.



Abbildung 36 Aktualisierung des Zeitbereichs bei Zoom

#### 4.5.1.2 Konfigurationsparameter

In Dashboards mit der Zeit als Hauptachse werden alle Daten zeitlich aggregiert. Die Auflösung der Daten wird durch die Time Bucket Width definiert. Über einen Input kann die Auflösung aus einer Liste von vorgegebenen Werten selektiert werden. Datensets, die Beobachtungen von Spezies oder PAX Werte beinhalten, werden mit der Summenfunktion aggregiert. Für andere Datensets besteht eine Auswahl an Methoden in einem Dropdown. Alle mit Machine Learning ausgewerteten Datenpunkte beinhalten einen Konfidenzwert, welcher die Wahrscheinlichkeit einer korrekten Klassifizierung beschreibt. Wird ein entsprechendes Datenset visualisiert, kann die Confidence über einen Filter nach unten begrenzt werden. Abbildung 37 zeigt die Bedienelemente mit ausgeklappten Dropdowns. Sie befinden sich im oberen Bereich der Dashboards neben dem Date Picker.

Abbildung 37 Auswahl Aggregation, Confidence und Zeitfenster

### 4.5.1.3 Dashboard Aktionen

Zu den Dashboard Aktionen gehört das Teilen einer Ansicht und die Erstellung von Annotationen. Sie werden über die schwebende Schaltfläche im unteren rechten Bereich der Dashboards, ersichtlich in Abbildung 38, erreicht. Jede Ansicht ist direkt reproduzierbar. Die Time Range und alle anderen Einstellungen sind in der URL als Query-Parameter aufgeführt. Die URL wird bei Änderungen der Konfigurationsparameter aktualisiert und ist immer aktuell. Bei Betätigung der Schaltfläche *'Share this View'* erscheint ein Modal, worüber die URL in die Zwischenablage kopiert werden kann.

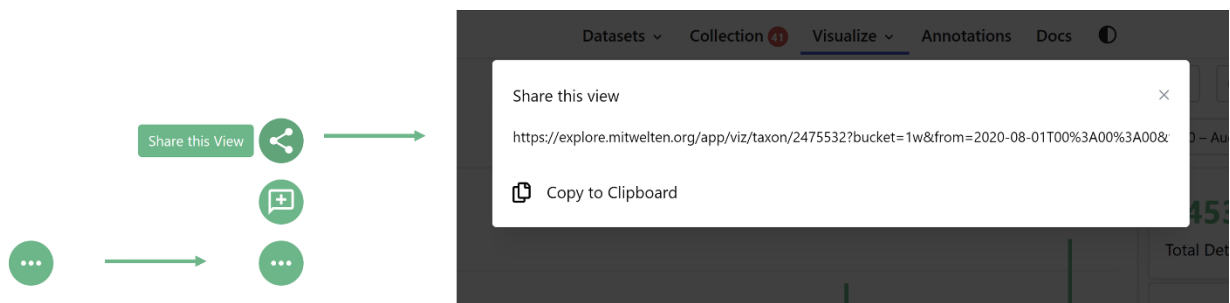


Abbildung 38 Dashboard Aktionen -Teilen einer Ansicht

## 4.5.2 Time Series Dashboard

Das Time Series Dashboard dient zur Analyse von Datenreihen mit der Zeitachse als Primärdimension. Abbildung 39 zeigt einen Screenshot des Dashboards mit PAX Counter Daten.

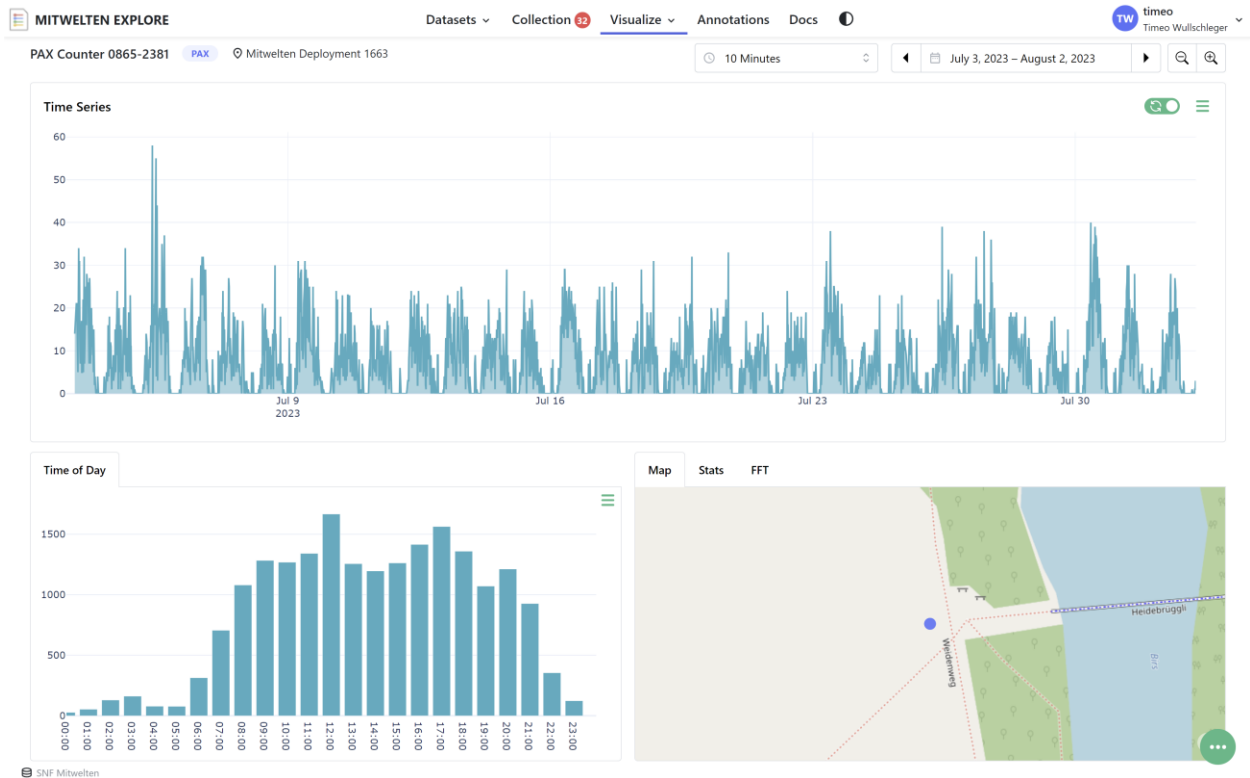


Abbildung 39 Times Series Dashboard

Das zentrale Element ist ein modifizierbares Zeitreihendiagramm. Muster, Trends oder Ausreisser sollen damit gut erkennbar sein. Unterhalb des Hauptdiagramms befindet sich eine Visualisierung der Daten in Abhängigkeit von der Tageszeit. Insbesondere für die Analyse der Aktivitäten von Spezies oder von Menschen können diese Informationen von hoher Bedeutung sein. Die Auflösung beträgt 60 Minuten. Die Aggregation erfolgt mit der selben Funktion wie beim Hauptchart.

Die Orte der Datenerfassung haben in diesem Dashboard eine geringere Relevanz. Im unteren rechten Bereich werden sie in einer Karte dargestellt. Als Erweiterte Informationen enthält das Dashboard statistische Aggregate und eine FFT der Messreihe. Durch Bedienung der Tabs oberhalb der Karte kann der Bereich durch erweiterten Content ersetzt werden. Abbildung 40 zeigt auf der linken Seite eine Tabelle mit berechneten Werten einer Messreihe und auf der rechten Seite die Visualisierung einer FFT. Die Werte wurden aus einem Datensatz der Bodentemperatur berechnet. Der in Abbildung 40 dargestellte Content wird erst nach der Auswahl des entsprechenden Tabs gerendert, um die Performance hoch zu halten.



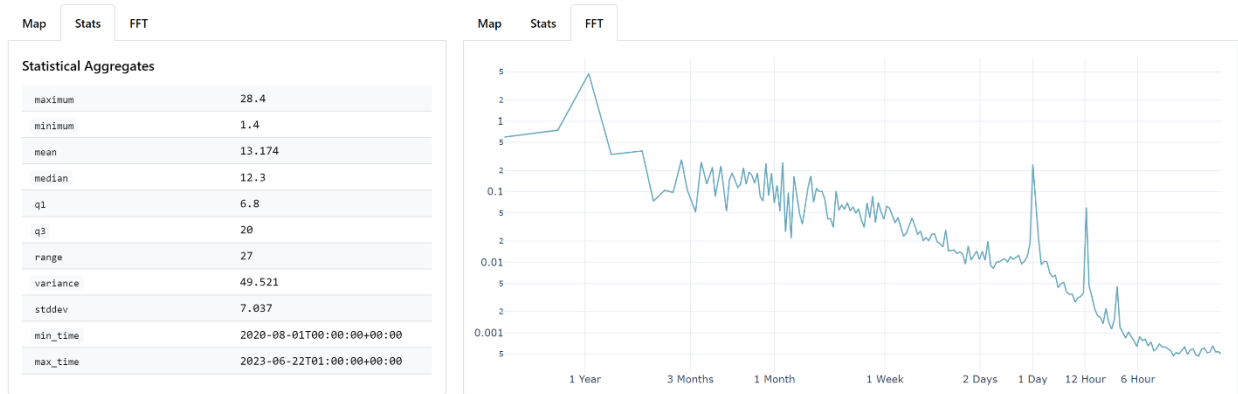


Abbildung 40 Statistische Aggregate und FFT im Time Series Dashboard

### 4.5.3 Compare Dashboard

Das Compare Dashboard ist für die Analyse von Zusammenhängen zwischen verschiedenen Datensätzen in einem definierten Zeitfenster konzipiert. Das in Abbildung 41 ersichtliche Dashboard vergleicht Meteodaten mit PAX Werten und der Bird Diversity. Für jeden Datensatz wird eine eindeutige Farbe eingesetzt. Die Legende, die immer sichtbar ist, zeigt die dargestellten Datensätze und die zugehörige Farbe.

Das Herzstück des Dashboards ist ein modifizierbarer Zeitreihenchart. Im Standard Layout besteht er aus mehreren Subplots mit einer gemeinsamen y-Achse. Muster, Abhängigkeiten, Ausreisser oder fehlende Daten sind leicht erkennbar.

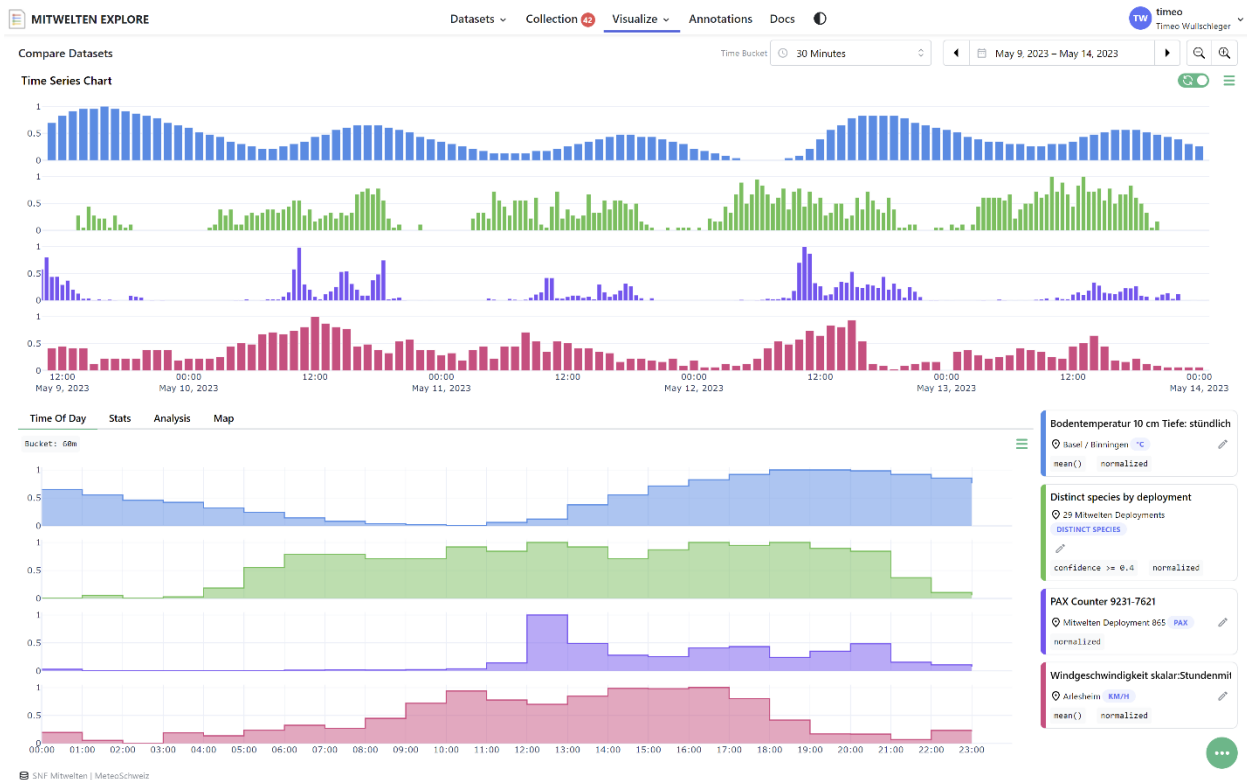


Abbildung 41 Compare Dashboard

Unterhalb des Haupt Charts befinden sich Tabs, die Informationen mit Sekundär- und Tertiärpriorität beinhalten. Beim ersten Aufruf des Dashboards wird ein Chart mit nach Tageszeit aggregierten Daten dargestellt, der in Abbildung 41 ersichtlich ist. Wie beim Haupt Chart ist das Layout und die Art der Visualisierung modifizierbar.

Durch Bedienung der Tabs oberhalb des Time-of-Day Charts kann dieser durch andere Komponenten ersetzt werden. Die Information, welche Registerkarte aktiv ist, wird im Browser gespeichert und beim nächsten Aufruf entsprechend gesetzt.

Der Tab *Stats* beinhaltet pro Datenset eine Tabelle mit statistischen Aggregaten. Die Kennwerte sind insbesondere bei der Untersuchung von Datensätzen desselben Typs von möglicher Relevanz. Eine Darstellung der Tabellen sind in Abbildung 42 ersichtlich.

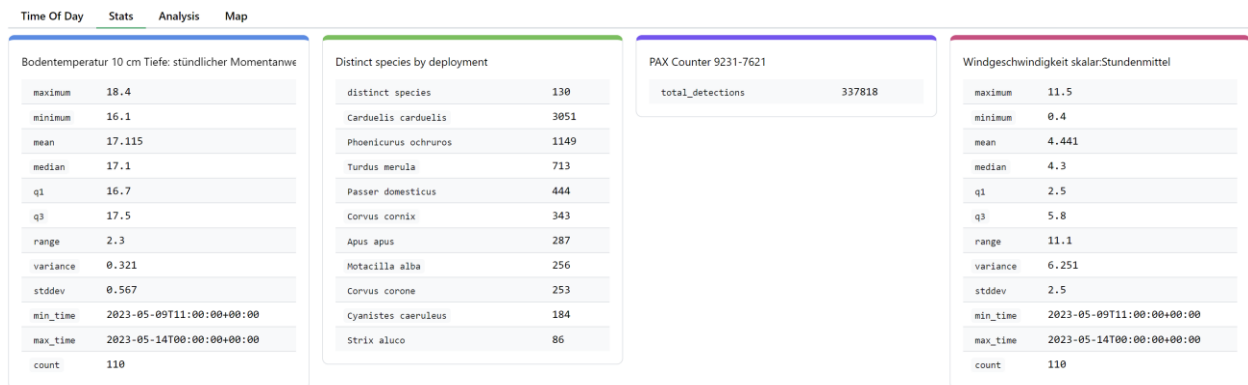


Abbildung 42 Compare Dashboard - Statistische Aggregate

Die Darstellungen der Kategorie Zeitreihenanalyse befinden sich im Tab *Analysis*, der in Abbildung 43 gezeigt wird. Der Pearson-Korrelationskoeffizient wird für jedes mögliche Messreihenpaar berechnet. Für die Berechnung werden die selben Datenpunkte verwendet, die im Haupt Chart dargestellt werden. Die Aggregierung und die Breite des Time Buckets hat einen direkten Einfluss auf die Korrelationskoeffizienten. Für die Darstellung der Werte wurde eine Korrelationsmatrix entwickelt, die im linken Bereich ersichtlich ist. Mit einer Farbskala werden starke positive und negative Korrelationen hervorgehoben.

Periodische Komponenten in einer Datenreihe sind hauptsächlich im Frequenzbereich identifizierbar. Für jedes Datenset wird eine FFT berechnet. Wie bei den Korrelationskoeffizienten werden die bereits geladenen Daten verwendet. Die eingestellte Time Range und die Bucket Size haben einen direkten Einfluss auf die berechneten Werte. Die Darstellung wurde mit einem Line Plot realisiert. Logarithmische Achsen dienen der besseren Identifizierbarkeit von Peaks. Zur Erhöhung der Performance wird ein Binning mit logarithmischen Bin Sizes angewendet.

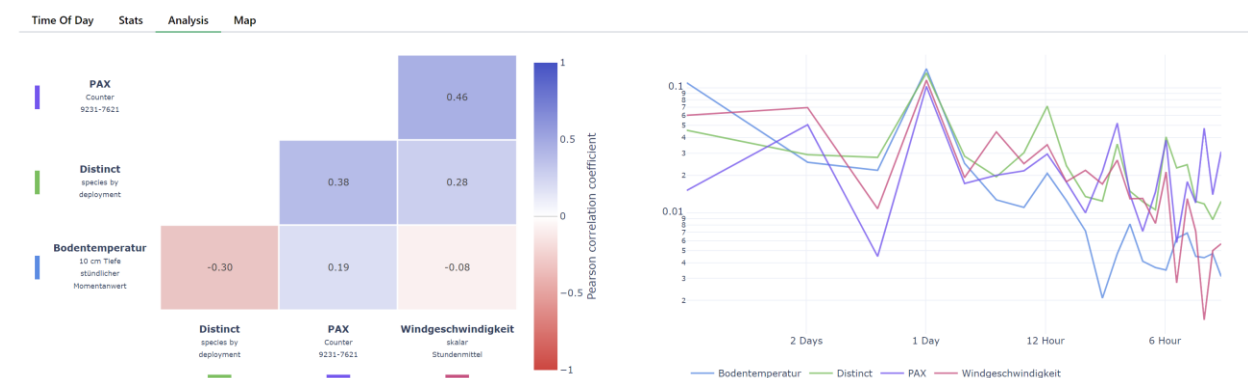


Abbildung 43 Compare Dashboard - Korrelationsmatrix und FFT

Die Erfassungsorte der Datenpunkte werden in einem weiteren Tab als Scatter Map Plot visualisiert. Falls ein Datenset mehrere Erfassungsorte referenziert, werden Punkte regional als Cluster zusammengefasst. Wenn in die Karte gezoomt wird, teilen sich die Cluster, bis die originalen Datenpunkte angezeigt werden. Das Clustering dient der Übersichtlichkeit der Karte und ermöglicht eine schnelle Kommunikation der Erfassungsorte.

Die Konfigurationsparameter Time Range und Time Bucket werden global festgelegt und auf alle Datensets angewendet. Damit die Nutzenden bei der Analyse möglichst viel Freiheiten haben, werden Methode zur Aggregation oder die minimale Confidence für jedes Datenset einzeln festgelegt. Jedes Datenset in der Legende verfügt über eine Schaltfläche, über welche ein Modal zur Konfiguration geöffnet wird. Die Schaltfläche ist in Abbildung 44 hervorgehoben. Das Modal enthält Eingabeelemente für Konfigurationen, die auf das entsprechende Datenset anwendbar sind. Zudem kann die Messreihe normalisiert werden, indem die Checkbox 'Normalize' angewählt wird.

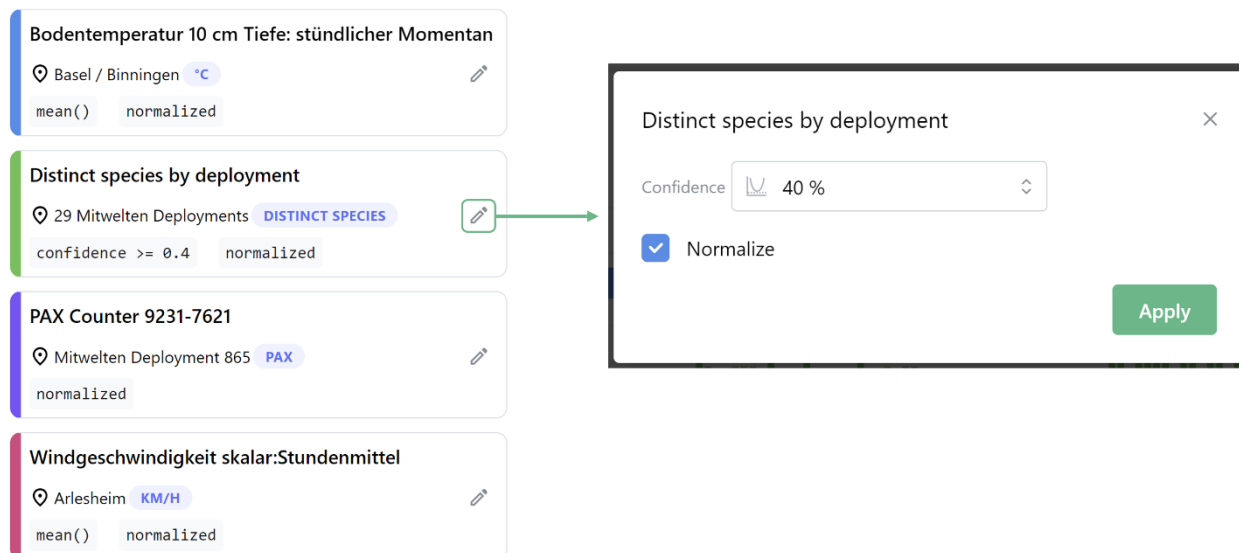


Abbildung 44 Compare Dashboard - Konfiguration der Datensets

#### 4.5.4 Taxon Dashboard

Das Taxon Dashboard ist für die Erforschung von Arten oder Taxonomie Gruppen konzipiert. Abbildung 45 zeigt Daten des Eisvogels im Taxon Dashboard. Die Saisonale Aktivität wird mit einem Bar Chart visualisiert. Die Gesamtzahl der Entdeckungen und die Anzahl verschiedener Deployments, welche das Taxon detektierten, werden als Kennzahlen rechts neben dem Bar Chart dargestellt.

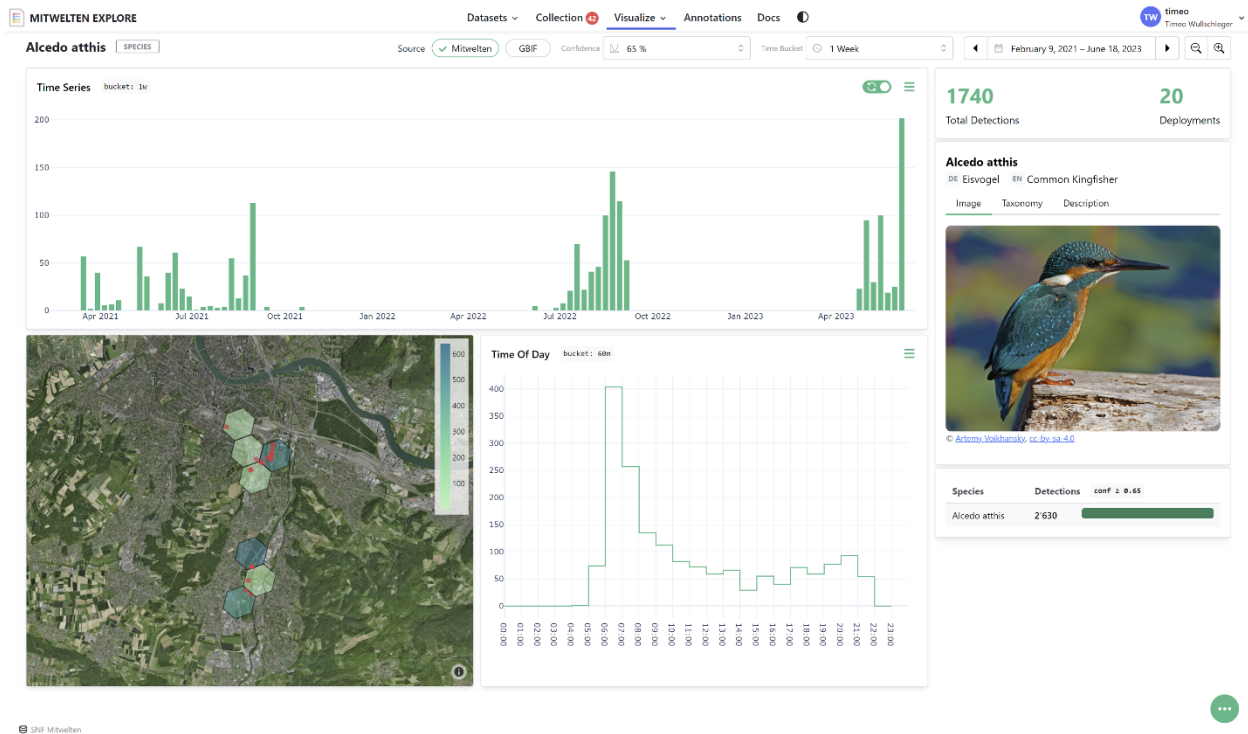


Abbildung 45 Taxon Dashboard

Unterhalb des Bar Charts wird die räumliche Aktivität in Form einer Hexbinmap visualisiert. Die Dimensionen der Sechsecke werden automatisch dem Zoom Level angepasst. Rechts davon befindet sich ein Line Chart, der die Aktivitäten im Bezug auf die Tageszeit kommuniziert.

Im rechten Bereich des Dashboards befinden sich zusätzliche Informationen über das ausgewählte Taxon. Ein Referenzbild soll helfen, visuelle Merkmale einer Art zu kommunizieren. Die Bilder stammen von WikiSpecies. Die Quellenangaben der Bilder werden von der MediaWiki API angefordert und unterhalb des Bildes angezeigt. Das Bild kann durch Bedienung der Tabs durch Informationen zur Taxonomie oder durch eine textuelle Beschreibung des Taxons ersetzt werden.

Eine Visualisierung der Taxonomie wird in Abbildung 46 auf der linken Seite gezeigt. In einer hierarchischen Darstellungsform sind alle höheren Ebenen aufgeführt und auf die jeweiligen Dashboards verlinkt. Im mittleren Bereich in Abbildung 46 ist eine Kurzbeschreibung des ausgewählten Taxons ersichtlich. Die Informationen stammen von Wikipedia. Über die Wikipedia API wird der Summary-Bereich der Seite über das ausgewählte Taxon abgefragt. Das Summary einer Wikipedia Page enthält gemäss den Richtlinien [59] eine kurze Zusammenfassung der wichtigsten Punkte des Eintrags. Aus dem Dashboard wird auf den Wikipedia Originaleintrag verlinkt, um effizient an detailliertere Informationen zu gelangen.



Abbildung 46 Taxon Dashboard - Erweiterte Informationen

Bei der Analyse einer Gattung, Familie, Ordnung oder Klasse werden die meisterkannten Arten der Taxonomie Gruppe unterhalb des Referenzbildes angezeigt. Die Visualisierung ist in Abbildung 46 rechts ersichtlich. Als Darstellungsmethode wurde eine Tabelle mit Balkendiagrammen implementiert. Die Skalierung der Balken ist nicht linear, damit auch kleinere Unterschiede erkennbar sind. Die aufgeführten Arten sind auf die entsprechenden Taxon Dashboards verlinkt.

Das im Projekt Mitwelten erhobene Bird Datenset beinhaltet Beobachtungen mit hoher Confidence, die von Ornithologen und Ornithologinnen als unwahrscheinlich erachtet werden. Die via GBIF bereitgestellten Beobachtungen aus externen Quellen können eingesetzt werden, um Datenpunkte zu validieren. Im oberen Bereich des Dashboards befindet sich eine Schaltfläche, über welche die Quellen der angezeigten Daten bestimmbar ist. Es ist möglich, Daten aus dem Mitwelten Projekt, Daten aus externen Quellen oder eine Vereinigung der Daten als Quelle zu definieren.

Am unteren Bildschirmrand wird angezeigt, woher die visualisierten Daten stammen. Wenn der Mauszeiger über dem Text liegt, erscheint eine Box mit einer Auflistung von Datensets mit Links auf die jeweiligen Beschreibungen. Wenn über GBIF bezogene Daten im Dashboard visualisiert werden, beinhaltet die Box im oberen rechten Bereich einen Link zu einer Auflistung von einzelnen Beobachtungen. Abbildung 47 zeigt einen Screenshot der Box, die sowohl Mitwelten als auch externe Beobachtungen als Datenquellen aufweist.



Abbildung 47 Datenquellen Taxon Dashboard

Die Auflistung von einzelnen Beobachtungen ist in Abbildung 48 ersichtlich. Über den Link 'List of observations', der in Abbildung 47 zu sehen ist, wird auf die Ansicht navigiert. Die Auflistung der Beobachtungen ist nach Datum sortiert. Jeder Eintrag beinhaltet den wissenschaftlichen Namen der Art, Datum und Zeit der Beobachtung, das GBIF Datenset und Informationen, ob



Medieninhalte vorhanden sind. Die GBIF-Occurrence Seiten der Datenpunkte sind über Links in der letzten Spalte aufrufbar.

**MITWELTEN EXPLORE** Datasets Collection 42 Visualize Annotations Docs Timeo Wullschlegler

Occurrence List 28481 total occurrences Offset 0 Limit 50 Go

#	Name	Has Media	Eventdate	Dataset	Link
1	Turdus philomelos		01.08.2020 06:26	iNaturalist Research-grade Observations	<a href="#">Link</a>
2	Cyanistes caeruleus		02.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
3	Parus major		02.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
4	Acrocephalus schoenobaenus		02.08.2020 22:00	Données naturalistes opportunistes Faune-France - Données Natu...	<a href="#">Link</a>
5	Sitta europaea		02.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
6	Passer domesticus		02.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
7	Corvus frugilegus		02.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
8	Delichon urbicum		02.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
9	Corvus corone		02.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
10	Parus major		03.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
11	Corvus corone		03.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
12	Cyanistes caeruleus		03.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
13	Passer domesticus		03.08.2020 22:00	EOD – eBird Observation Dataset	<a href="#">Link</a>
14	Phylloscopus collybita		03.08.2020 22:00	Données naturalistes opportunistes Faune-France - Données Natu...	<a href="#">Link</a>
15	Delichon urbicum		03.08.2020 22:00	Données naturalistes opportunistes Faune-France - Données Natu...	<a href="#">Link</a>

Abbildung 48 Auflistung einzelner Datenpunkte

#### 4.5.5 Map Dashboard

Mit dem Map Dashboard können Datensets anhand der räumlichen Achse analysiert und verglichen werden. Abbildung 49 zeigt Beobachtungen des Eisvogels im Map Dashboard. Die Datenpunkte werden mit einer interaktiven Hexbinmap visualisiert. Rechts neben der Karte befindet sich die Legende und Bedienelemente zur Anpassung der Darstellung.

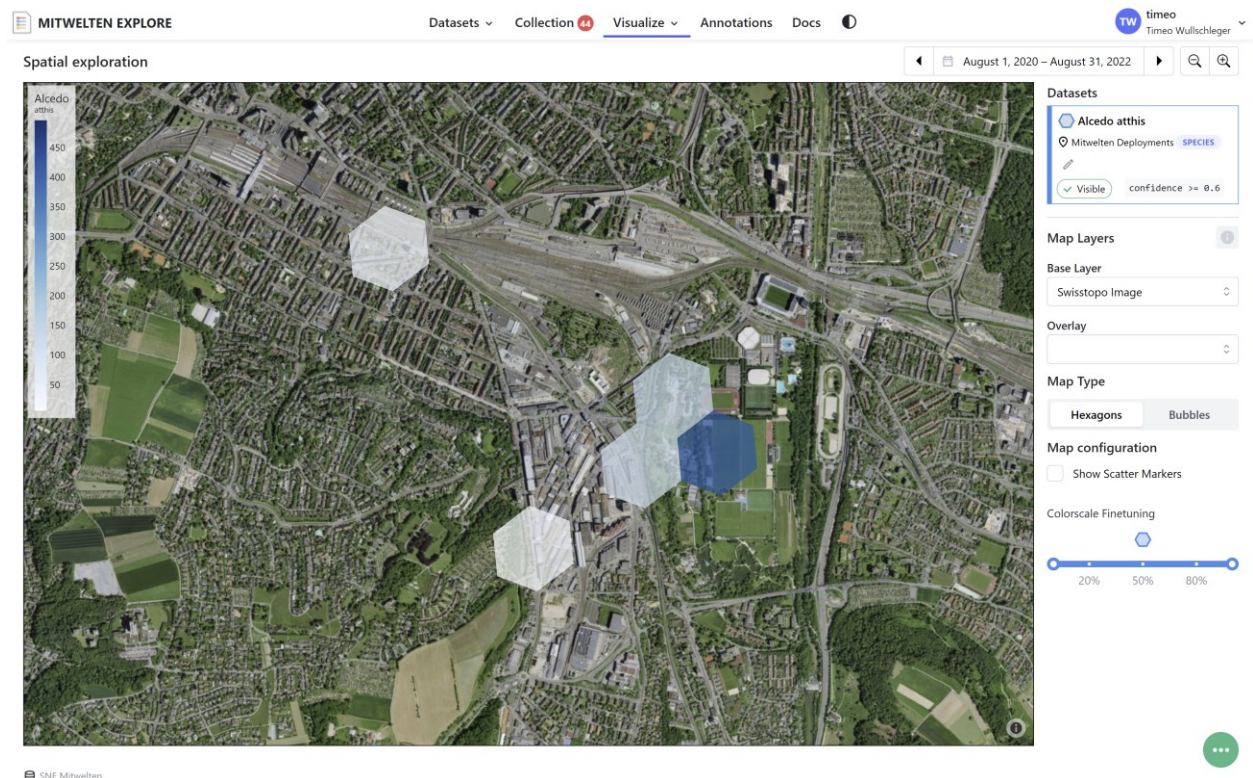


Abbildung 49 Map Dashboard

Standardmässig wird eine Hexbinmap zur Darstellung der Daten genutzt. Die originalen Datenpunkte sind ein- und ausblendbar. Die berechneten Werte pro Cluster werden mit einer linearen

Farbskala angezeigt. Das Minimum der Skala liegt beim geringsten Wert, das Maximum beim höchsten Wert. Mit einem Schieberegler kann der Datenbereich der Farbskala angepasst werden, um weniger signifikante Unterschiede hervorzuheben oder um Ausreisser zu eliminieren.

Neben der Analyse von einem einzelnen Datenset können zwei Datensets im Dashboard verglichen werden. Die berechneten Werte des ersten Datensets werden durch die Füllfarbe der Zellen dargestellt. Das zweite Datenset wird durch die Rahmenfarbe der Clusters kommuniziert. Abbildung 50 zeigt den Vergleich zwischen der Eisvogel-Beobachtungen und Beobachtungen der Klasse Aves, die alle Vögel beinhaltet. Beim oberen Eintrag in der Legende ist ein Symbol eines gefüllten Hexagons eingezeichnet, welches darauf hinweist, dass dieses Datenset mit der Füllung der Cluster visualisiert wird. Entsprechend weist das Outline-Hexagon beim unteren Eintrag auf die Kommunikation der Werte anhand der Rahmenfarbe hin. Die originalen Datenpunkte werden entsprechend Legende eingefärbt.

Wie beim Compare Dashboard werden die minimale Confidence oder die Methode zur Aggregation pro Datenset über die Legende konfiguriert. Jedes Datenset kann über die Schaltfläche 'Visible' ein- oder ausgeblendet werden.



Abbildung 50 Map Dashboard – Vergleich von zwei Datensets

Als Alternative zu der Hexbinmap wurde eine Bubble Map entwickelt. Für jeden Datenpunkt wird ein Marker auf den Map Layer gezeichnet. Der Wert eines Datenpunkt wird durch die Grösse des Markers repräsentiert. Für die Darstellung werden die Punkte pro Datenset nach Grösse sortiert, damit grosse Marker die kleineren nicht verdecken. Eine Visualisierung von PAX Daten in Form einer Bubble Map ist in Abbildung 51 ersichtlich. In gewissen Konfigurationen kann es vorkommen, dass sich Marker von verschiedenen Datensets überdecken. In diesem Fall kann die Reihenfolge der Datensets über die Schaltfläche 'Switch Positions' unterhalb der Legende umgekehrt werden.

Karten ermöglichen die Kommunikation von zusätzlichen Informationen über Eigenschaften und Charakteristiken einer Umgebung. Das Map Dashboard bietet eine Auswahl an Layer, die in der Karte darstellbar sind. Der Hintergrund der Karte wird als Base Layer bezeichnet. Ein Overlay ist typischerweise halbtransparent und wird über dem Base Layer angezeigt. Die Lebensraumkarte vom Bundesamt für Umwelt beinhaltet Informationen mit Relevanz für die Biodiversität. Die Map in Abbildung 51 nutzt Swisstopo Image als Base Layer und stellt die Lebensraumkarte als Overlay dar. Enthält ein Layer eine Legende, wird diese im unteren Bereich über der Karte angezeigt.



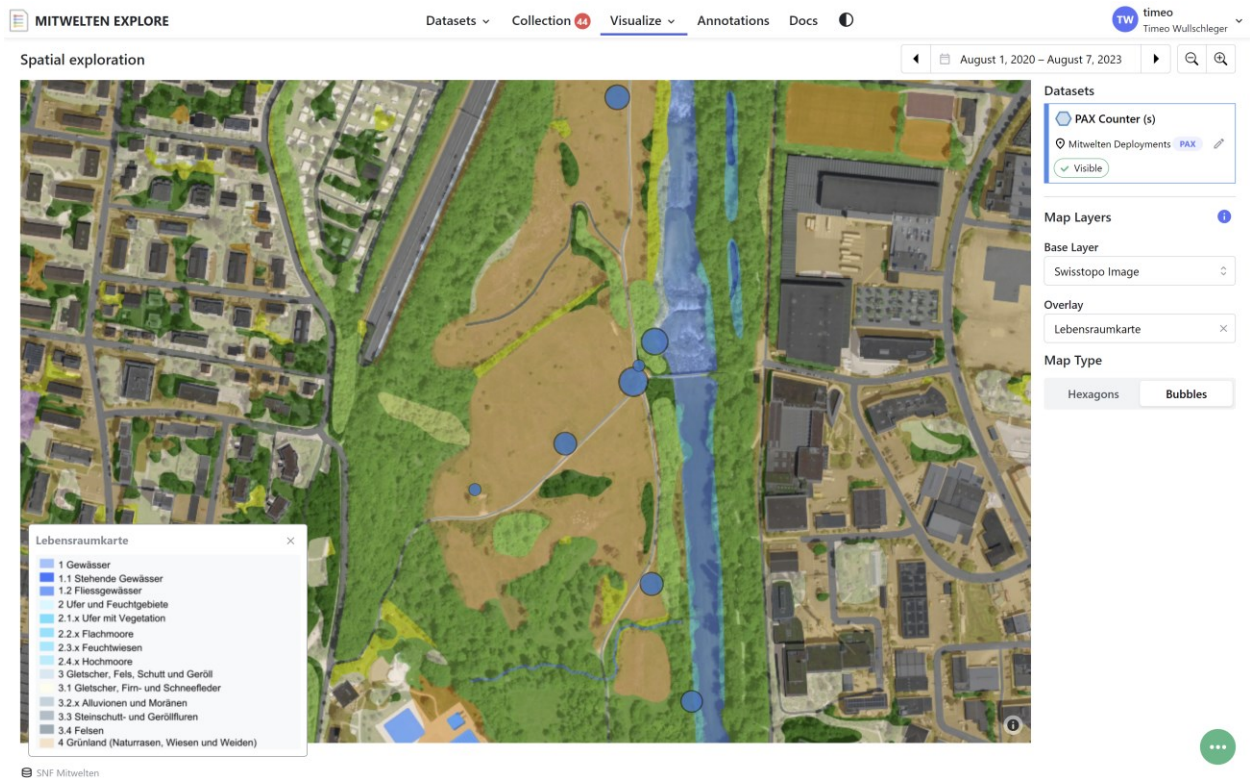


Abbildung 51 Map Dashboard - Bubble Map mit Lebensraumkarte

Datensets der Typen **Taxon** (Mitwelten und GBIF), **Pollinator** und **Distinct Birds**, die in der Collection enthalten sind, können in dem Map Dashboard visualisiert werden. Zusätzlich werden Standard Datensets für die Auswahl bereitgestellt. Sie beinhalten ein PAX Datenset, welches alle PAX Deployments beinhaltet. Weiter ist für jede Attribute zur Beschreibung der Umgebung ein Datenset vorhanden. Die Umgebungseigenschaften können für die Erforschung von Tierbeobachtungen relevant sein. Abbildung 52 zeigt die Aktivität des Eisvogels mit der Füllfarbe der Zellen und die Nähe zu Gewässer mit der Rahmenfarbe. Es ist klar ersichtlich, dass der Eisvogel hauptsächlich in Gewässernähe, wo sein bevorzugter Lebensraum ist [60], beobachtet wurde.



Abbildung 52 Map Dashboard - Eisvogel und Gewässernähe

#### 4.5.6 Deployment Dashboard

Das Deployment Dashboard wurde für die Erforschung der Biodiversität an einem bestimmten Ort oder einer Gruppe von Orten entwickelt. Kompatible Datensets sind Mitwelten Birds und Mitwelten Pollinators. Abbildung 53 zeigt das Dashboard mit Bird-Daten von einem Deployment in Reinach.

Das zentrale Element im Dashboard ist eine durchsuchbare Liste, die alle vom Deployment erkannten Arten auflistet. Jeder Eintrag ist auf das entsprechende Taxon Dashboard verlinkt. Die Einträge sind gemäss der detektierten Aktivität sortiert.

Die anteilmässige Aktivität pro Art wird mit einem Kuchendiagramm visualisiert. Aus dem Diagramm in Abbildung 53 geht hervor, dass über 75 % aller erkannten Rufe von fünf Arten stammen. Diese Visualisierung ermöglicht eine schnelle Kommunikation der Artenvielfalt an einem bestimmten Ort. Oberhalb des Charts wird die Anzahl entdeckter Arten als Kennzahl angegeben. Die Anzahl der gesamten Beobachtungen wird als zweite Kennzahl visualisiert.

Durch Bedienung der Schaltflächen unterhalb des Kuchendiagramms können Abschnitte mit zusätzlichen Informationen ein- oder ausgeklappt werden. Zur Verhinderung des Informationsüberflusses beim ersten Aufruf ist standardmässig nur das Kuchendiagramm ausgeklappt.

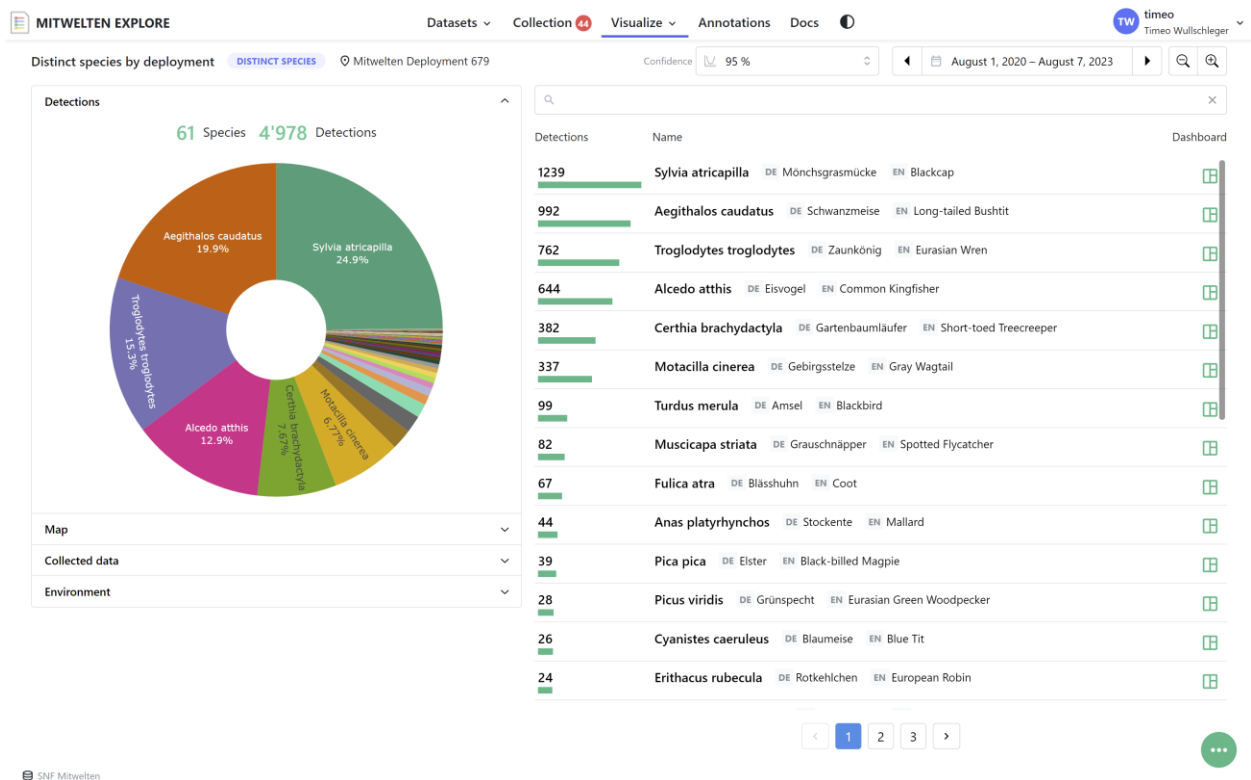


Abbildung 53 Deployment Dashboard

Im Abschnitt 'Map', ersichtlich in Abbildung 54, werden die Standorte der ausgewählten Deployments visualisiert. Zur Darstellung wurde eine Karte verwendet, die jeden Standort mit einem blauen Marker hervorhebt. Die roten Marker visualisieren Orte mit bekannten Umgebungscharakteristiken. Die Intensität der Farbe repräsentiert den Abstand zu den Deployments.

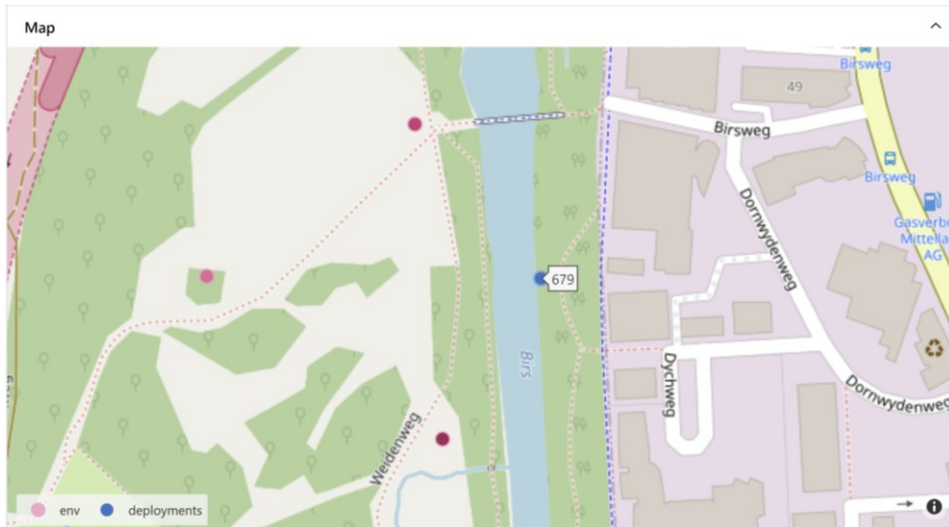


Abbildung 54 Deployment Dashboard - Map

Im Abschnitt 'Environment', der in Abbildung 55 ersichtlich ist, werden die Charakteristiken der Datenerfassungsorte visualisiert. Als Visualisierungsform wurde ein Spider Chart implementiert, der Attribute der drei nächstgelegenen Erhebungspunkte visualisiert. Die Distanz der jeweiligen Punkte zum Mittelpunkt der Deployments wird in der Legende angegeben. Die Farben der Traces entsprechen den Farben der Karte in Abbildung 54.

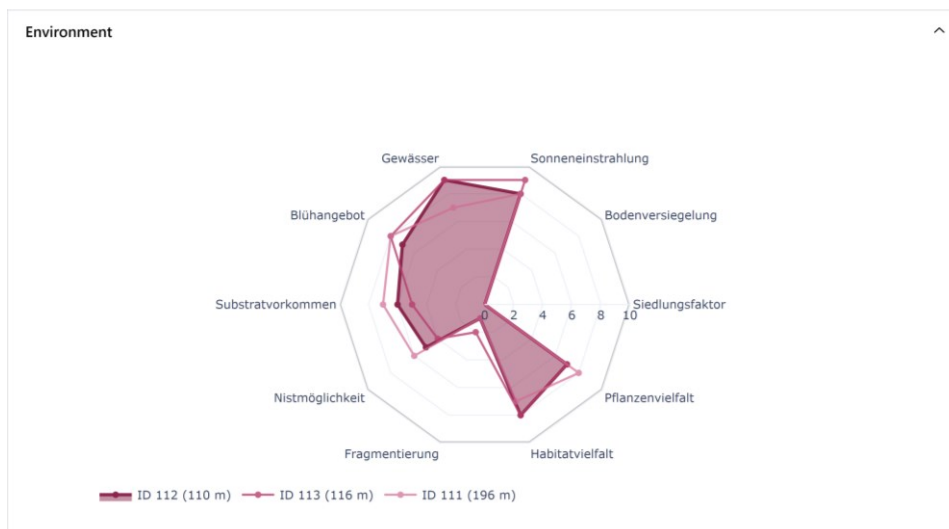


Abbildung 55 Deployment Dashboard - Environment

Bei der Interpretation von ausgewerteten Datenpunkten ist es wichtig, die Datengrundlage zu kennen. Informationen über die erhobenen Rohdaten werden im Abschnitt 'Collected Data' dargestellt. Als Rohdaten gelten bei Audio Deployments die Audioaufnahmen. In einem Bar Chart wird die tägliche Aufnahmedauer visualisiert. Oberhalb des Charts wird die gesamte Aufnahmedauer als Kennzahl angegeben. Bei Pollinator Deployments wird die Anzahl Captures pro Tag als Bar Chart visualisiert. Abbildung 56 zeigt den Abschnitt 'Collected Data' mit Daten eines Audio Deployments.

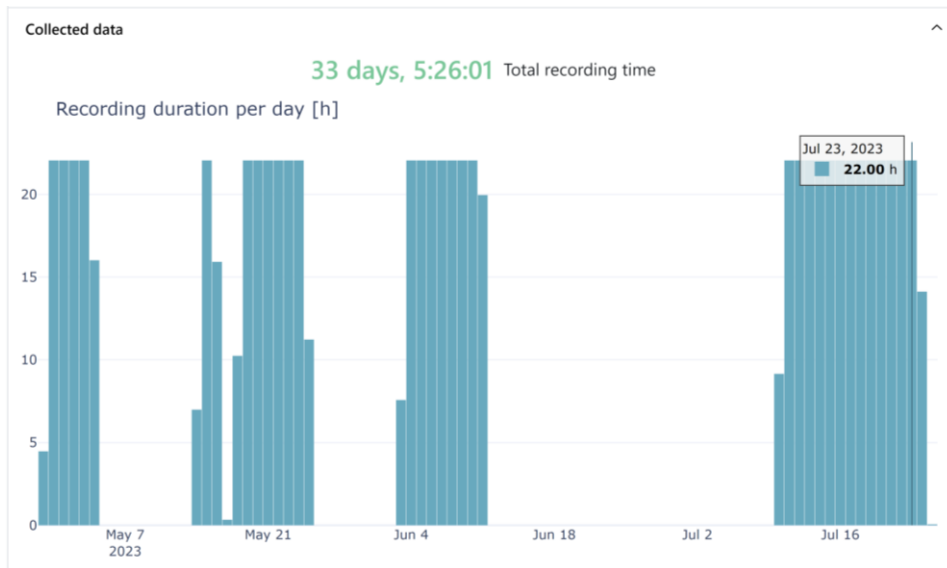


Abbildung 56 Deployment Dashboard - Erfasste Rohdaten

## 4.6 Annotationen

Anmerkungen, Kommentare oder Beobachtungen können als Annotationen im Tool gespeichert und mit anderen Nutzenden geteilt werden. Eine Annotation beinhaltet folgende Informationen:

- Ein Titel
- Ein Text im Rich Text Format
- Eine Referenz zu einem Dashboard mit den aktuell eingestellten Parametern (Zeitspanne, Confidence, Aggregationsmethoden)
- Die User ID der verfassenden Person
- Der Zeitpunkt der Verfassung

Die Zuweisung einer Annotation zu einer Person wurde über die User ID gemacht, um Änderungen des Usernames zu ermöglichen. Eine Liste der verfassten Annotationen sind unter der Route '/annotations' erreichbar. Abbildung 57 zeigt einen Screenshot der Annotationsübersicht zum Zeitpunkt der Verfassung dieser Dokumentation.

MITWELTEN EXPLORE

Datasets Collection 23 Visualize Annotations Docs

Annotations

User	Title	Text	Date	Time	Age	Actions
JT	Sperber	Schön, wie man bei Rollover die Anzahl Feststellungen pro AudioMoth einsehen kann. Sperber ist so im Gelände sehr gut möglich.	11.07.2023	05:08 UTC	11 HOURS AGO	MAP BIRDS
JT	Wenig wahrscheinliche Arten im Gundeli-Quartier	* Italiensperling (genrell CH, ausser Tessin) * Nebelkrähe * Felsentaube ist wohl Haustaube? * Cistensänger * Waldschnepfe (ausser Vogelzug Herbst, Frühling) * Zwerdommel * Alpendohle * Grauammer	11.07.2023	04:40 UTC	11 HOURS AGO	DEPLOYMENT BIRDS_BY_DEPL
TW	Vögel in der Reinacherheide	115 verschiedene Arten wurde in der Reinacherheide seid Beginn 2023 entdeckt. 25 davon jeweils nur einmal.	30.06.2023	19:55 UTC	1 WEEK AGO	DEPLOYMENT BIRDS_BY_DEPL
TW	Temperaturwerte an vier Standorten	Dreispiß ist immer etwas wärmer	23.06.2023	12:26 UTC	2 WEEKS AGO	COMPARE METEODATA METEODATA METEODATA
TW	Eisvogel & Mauersegler	Es ist auffällig, wie der Eisvogel hauptsächlich bei Geräusern aktiv ist, während der Mauersegler vor Allem in der Stadt detektiert wurde.	20.06.2023	13:39 UTC	3 WEEKS AGO	MAP BIRDS BIRDS
TA	Test	This is what I wrote.	15.06.2023	09:49 UTC	3 WEEKS AGO	TIMESERIES FAX

Abbildung 57 Screenshot Übersicht Annotationen



Durch klicken auf eine Annotation in der Übersicht wird der Inhalt geöffnet. Die Ansicht ist in Abbildung 58 ersichtlich. Im mittleren Bereich wird der formatierte Content dargestellt, auf der rechten Seite befinden sich Informationen über den Autor oder die Autorin, den Zeitpunkt der Verfassung, die im Dashboard dargestellten Datensets und die eingestellte Zeitspanne. Über die Schaltfläche Dashboard Reference kann das entsprechende Dashboard geöffnet werden. Falls die Annotation von der Person, die sie verfasst hat, betrachtet wird, befinden sich zusätzliche Schaltflächen zum Bearbeiten und zum Löschen der Annotation.

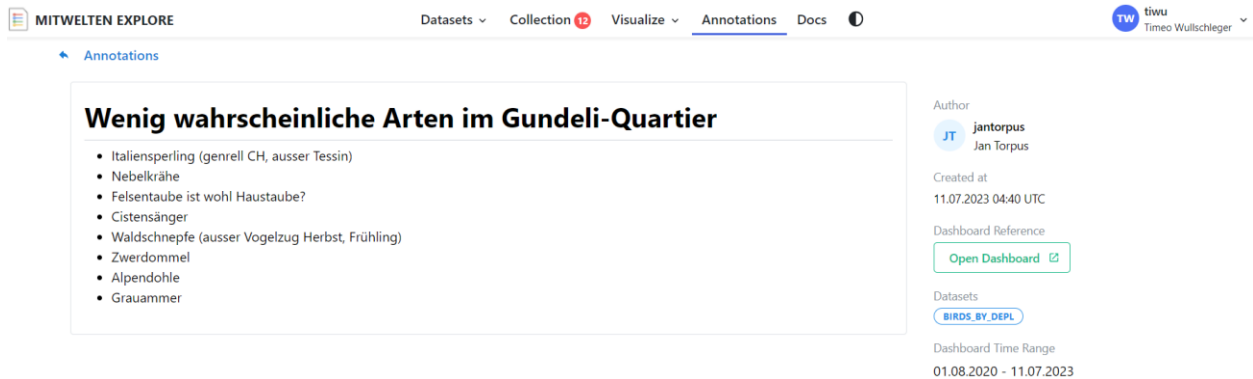


Abbildung 58 Screenshot Annotation

#### 4.6.1 Verfassen einer Annotation

In jedem Dashboard befindet sich in der unteren rechten Ecke ein schwebendes Bedienelement, über welches sich der Editor zum Verfassen einer Annotation öffnen lässt. Als Texteditor wurde eine Dash Implementation [61] vom Rich Text Editor Quill [62] eingesetzt. Damit der Editor das Dashboard nicht verdeckt, wurde eine Client Side JavaScript Funktion implementiert, die Änderungen der Grösse und der Position des Fensters ermöglicht. Der verfasste Content wird von Quill im HTML Format ausgegeben. Mit Bleach [63] werden Elemente in Tags, die nicht in der Whitelist aufgeführt sind, entfernt, um Cross Site Scripting (XSS) [64] auszuschliessen. Vor dem Übertragen der Annotation an das Backend wird der Content zu Markdown konvertiert. Bilder aus dem Web können mit Copy-Paste in den Content eingebettet werden. Bei der Anzeige des Bildes wird dieses von der Ursprungs URL geladen.

Abbildung 59 zeigt den schwebenden Editor zum Verfassen von Annotationen. Die Schaltfläche zum Publizieren kann erst betätigt werden, wenn ein Titel vorhanden ist. Eine Vorschau der Annotation wird im Tab Preview angezeigt.

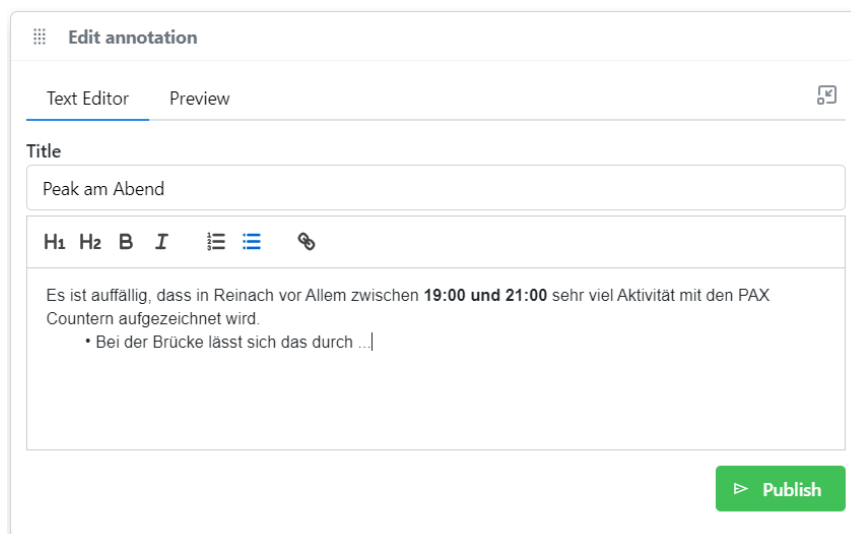


Abbildung 59 Screenshot Annotationseditor

#### 4.6.2 Bearbeiten von Annotationen

Eine Annotation kann durch den Autor oder die Autorin bearbeitet werden. Der Content wird von Markdown wiederum zu HTML konvertiert, das mit einem Quill Editor bearbeitet wird. Mit einem PUT Request an das Backend wird der Content aktualisiert. Der Screenshot in Abbildung 60 zeigt den Editor zum Bearbeiten einer Annotation.

### Temperaturwerte an vier Standorten

H1 H2 B I

Die Temperatur am Dreispitz ist jeweils um 1.5 °C wärmer als an den anderen Standorten.

- Könnte die Versiegelung

Author

TW
tiwu
Timeo Wullschleger

Created at
23.06.2023 12:26 UTC

Dashboard Reference

Open Dashboard

Datasets

METEODATA
METEODATA
METEODATA
METEODATA
METEODATA

Dashboard Time Range
15.10.2021 - 16.01.2023

Cancel
Save

Abbildung 60 Screenshot Bearbeiten einer Annotation

#### 4.6.3 Löschen von Annotationen

Personen, die eine Annotation verfasst haben sind berechtigt zum Löschen der entsprechenden Annotation. Im Backend wird überprüft, ob die User ID im JWT der eingeloggten Person mit der ID der verfassenden Person übereinstimmt. Von anderen Personen verfasste Annotationen können nur durch Personen mit der User-Role `explore_admin` gelöscht werden.

## 4.7 Anleitung und Dokumentation im Tool

Eine Dokumentation, die alle Abschnitte und Funktionen des Tools abdeckt, ist unter der Route '/docs' erreichbar. Die Dokumentation enthält eine Anleitung, die zum Ziel hat, dass Anwendende die Funktionen kennenlernen und das Tool selbstständig nutzen können. Die Dokumentation ist in mehrere Themen gegliedert.

Für die Navigation zwischen den Seiten wurden Tabs eingesetzt, die den Hash in der URL entsprechend setzen. Die Dokumentation der Dashboards ist entsprechend unter der URL '/docs#dashboards' aufzufinden. Ein Screenshot der Dokumentation der Datensets ist in Abbildung 61 ersichtlich.

Pro Kapitel wurde ein Markdown File geschrieben, welches den Seiteninhalt enthält. Bei der Navigation wird das File gelesen und formatiert angezeigt. Alle in der Dokumentation enthaltenen Bilder sind im Assets Folder im Tool gespeichert und werden lokal geladen. In verschiedenen Seiten des Tools wurde auf das entsprechende Thema verlinkt, um die Nutzenden über Funktionen aufzuklären.

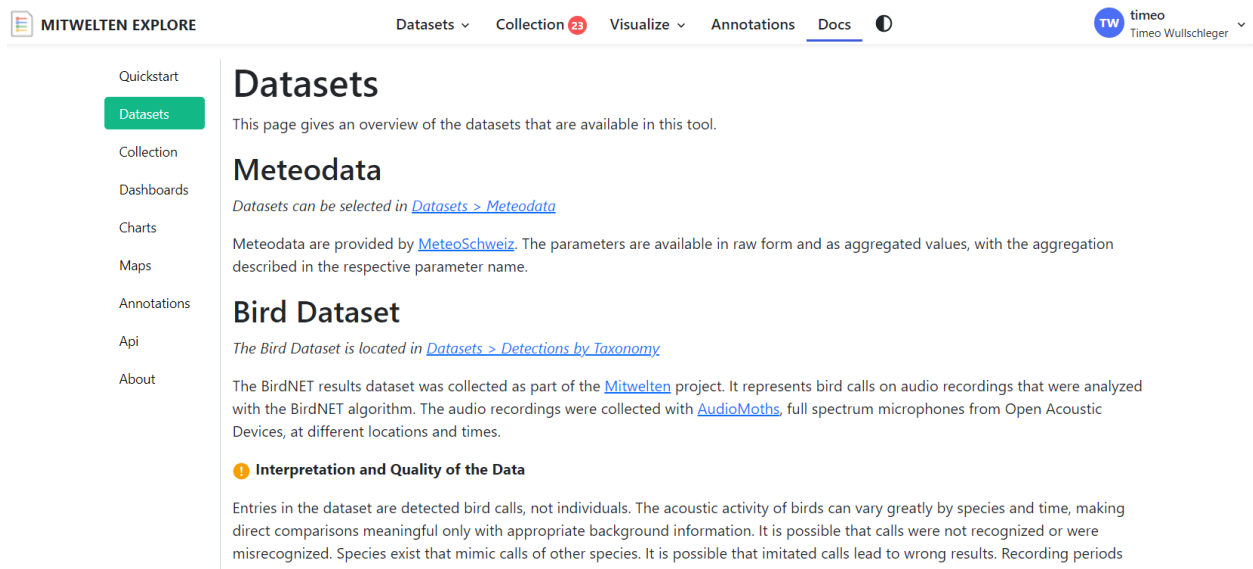


Abbildung 61 Screenshot Dokumentation im Tool



## 4.8 Landing Page

Eine Landing Page ist die erste Seite, die eine Person beim Aufrufen einer Website sieht. Sie ist wichtig, um einen guten ersten Eindruck zu vermitteln. Die Landing Page der Mitwelten Explore Applikation ist in Abbildung 62 dargestellt. Sie ist ohne Registrierung zugänglich und ist über eine Google Websuche auffindbar.

Die Landing Page wurde minimalistisch gestaltet. Im oberen linken Bereich befindet sich der Titel und ein Hinweis, dass man das Tool nach einer Registrierung nutzen kann. Direkt unterhalb befinden sich Links zum Einloggen und zum Registrieren.

Der rechte Bereich zeigt eine Slideshow mit Screenshots der Dashboards. Über den Bildern befindet sich jeweils eine Beschriftung und eine Beschreibung, was man im entsprechenden Dashboard erkunden kann. Im unteren Bereich des Bildschirms befinden sich Links auf die Webseite der FHNW, die Webseite des Projekts Mitwelten und auf das GitHub Repository dieser Projektarbeit. Der Hintergrund wurde mit einem dezenten Farbverlauf versehen. Die verwendeten Farben harmonisieren mit den Bildern, wodurch ein ansprechendes visuelles Gesamtkonzept angestrebt wird. Die Navigationsleiste ist bis auf den Inhalt identisch mit der Leiste in der Applikation, wodurch die interne Konsistenz erhöht wird. Funktional betrachtet ist sie auf der Landing Page nicht zwingend notwendig.

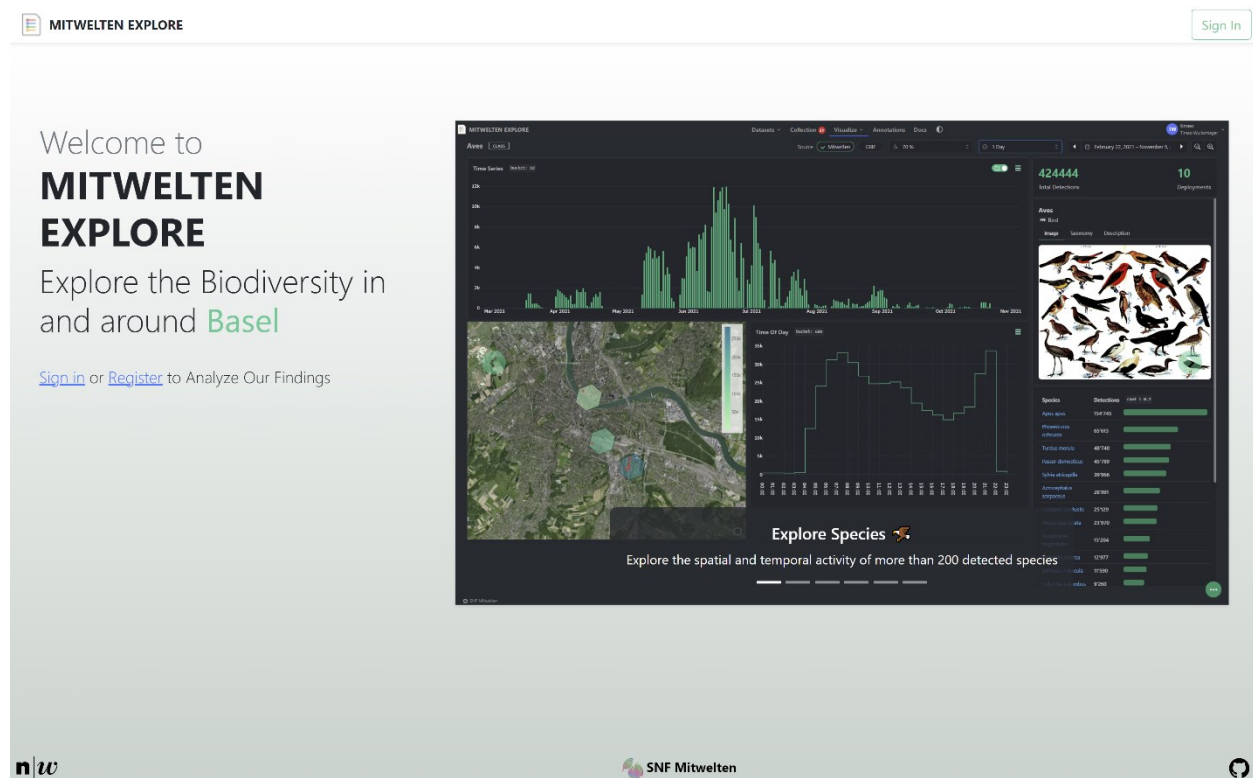


Abbildung 62 Landing Page Explore Tool

Der Anteil mobiler Geräte für den Zugriff auf Websites lag im Juli 2023 weltweit bei knapp 56 % [65]. Im Gegensatz zu den Dashboards, die explizit für die Nutzung auf Desktop Geräten konzipiert sind, ist die Landing Page responsive. Die Wahrscheinlichkeit, dass jemand mit einem Smartphone auf die Seite zugreift, ist hoch.

Abbildung 63 zeigt die Landing Page in der Mobile-Ansicht. Die Elemente sind vertikal angeordnet, damit die Bildschirmfläche optimal ausgenutzt wird. Die Texte über den Bildern werden nicht gezeigt, da sie den relevanten Inhalt überdecken würden. Die Slideshow kann auf Touch-Geräten mit Wischen nach links oder rechts gesteuert werden.

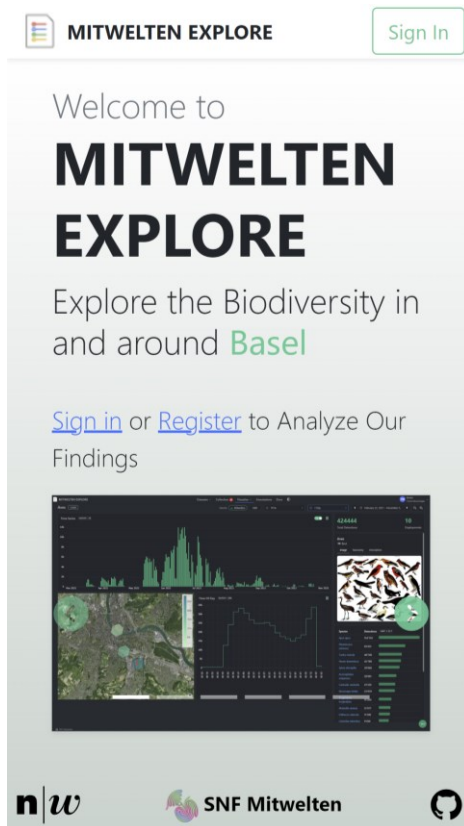


Abbildung 63 Landing Page Mobile Device

Die Applikation soll gefunden werden, wenn eine Person im Internet danach sucht. Mit knapp 93 % weltweitem Marktanteil ist Google die meistgenutzte Suchmaschine. Bing liegt mit 2.9 % auf dem zweiten Platz [66]. Damit eine Webseite in den Resultaten einer Suchanfrage enthalten ist, muss sie verifiziert und indiziert werden. In der Google Search Console wurde die Applikation als neue Property erfasst. Für die Verifizierung der Landing Page wurde ein HTML-Tag bereitgestellt, das in den Header des HTML Files kopiert wurde. Nach erfolgter Verifizierung wurde eine Indexierung beantragt, die nach einigen Stunden bestätigt wurde. Der Prozess wurde analog für eine Indexierung bei Bing durchgeführt. Die Indexierung wurde mit einer Websuche nach dem Begriff 'Mitwelten Explore' getestet. Sowohl bei Google als auch bei Bing erschien der Eintrag an oberster Stelle. Weitere Suchmaschinen wie DuckDuckGo oder Yahoo listen die Webseite ebenfalls in den Suchresultaten auf. Sie übernehmen Indexierungen von grossen Suchmaschinen. DuckDuckGo nutzt unter Anderem die Bing Search Engine zum indexieren von Webseiten [67].

Der in den Suchresultaten angezeigte Text wird von dem Description-Tag im HTML Header übernommen. Als Description wurde eine Beschreibung der Applikation in zwei Sätzen verfasst. Über den 'og:image' Tag wurde ein Bild des Taxon Dashboards definiert, welches beim Teilen der URL in sozialen Medien oder in Chats als Vorschau angezeigt wird. Abbildung 64 zeigt die Link Vorschau in Microsoft Teams.

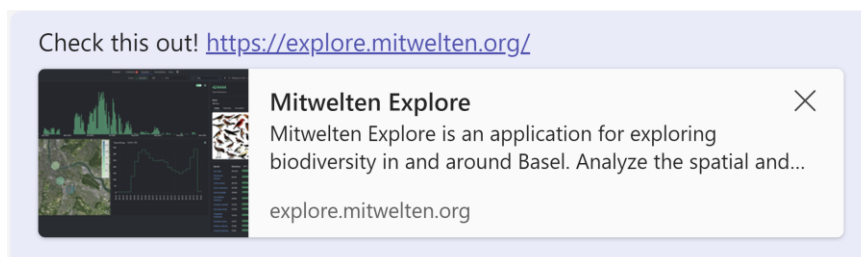


Abbildung 64 Link Preview in MS Teams

Verschiedene Anbieter, darunter Bing, bieten AI-gestützte Funktionen zur Zusammenfassung von Inhalten oder für die Extrahierung von Informationen aus Webseiten an. Der Bing Chatbot basiert auf dem LLM (Large Language Model) GPT-4 und kann auf Online Ressourcen zugreifen.

Die Antwort des Bing Chatbots auf die Frage 'what is mitwelten explore ?' ist in Abbildung 65 ersichtlich. Die Informationen stammen aus der Landing Page. Grundsätzlich wurden die Informationen gut interpretiert. Die aufgelisteten Aktionen, die in der Applikation möglich sind, wurden möglicherweise aus den Bildbeschriftungen abgeleitet. Es ist nicht klar, welcher Anteil an Informationen effektiv aus der Landing Page extrahiert wurde und wie viel aus dem Kontext generiert wurde.



Abbildung 65 Zusammenfassung der Mitwelten Explore Applikation - Bing Chat [68]

## 4.9 Datenvisualisierung

Sämtliche Visualisierungen in den Dashboards wurden mit Plotly Graph Objects entwickelt. Mit vereinheitlichten Figure-Layouts wird eine anwendungsinterne Konsistenz angestrebt. Die Layouts basieren, abhängig vom aktiven Dashboard Theme, auf den minimalistischen Templates `plotly_white` oder `plotly_dark`. Die Konstruktion der Darstellungen erfolgt in Callback Funktionen, wo ausgewählte Darstellungsoptionen und die Einstellung des Dashboard Themes bekannt sind.

Die Elemente in den Dashboards werden neu geladen, wenn der Zeitbereich oder andere Parameter modifiziert werden. Ein Loading Overlay informiert die nutzende Person über den Ladevorgang, sobald der Vorgang länger als 400 ms dauert. Die meisten Grafiken benötigen für die Aktualisierung weniger als eine Sekunde. Bei interaktiven Hexbinmaps werden die Cluster bei Änderungen des Zoom Levels neu berechnet. Damit die Interaktionen auch bei hohen Datenmengen flüssig erscheinen, befindet sich der Indikator am unteren Rand und nicht im Zentrum des Charts. Abbildung 66 zeigt das entwickelte Loading Overlay im unteren rechten Bereich eines Zeitreihendiagramms.

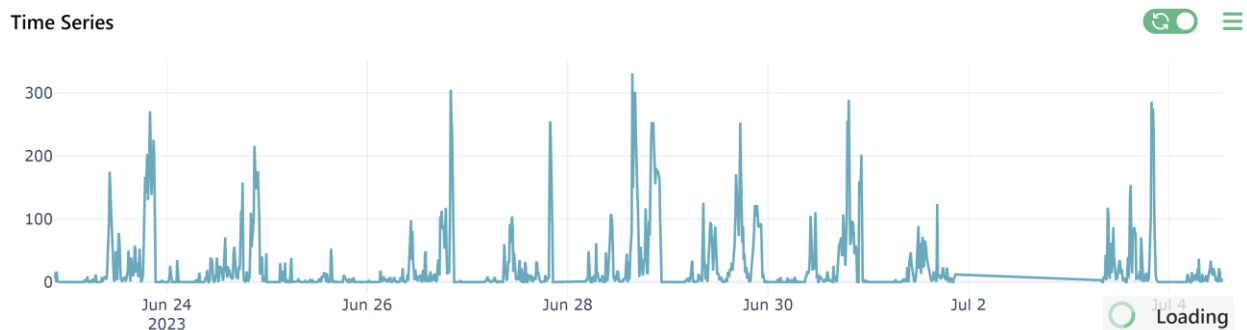


Abbildung 66 Indikation des Ladevorgangs eines Line Charts

### 4.9.1 Color Palette

Für die Visualisierung von mehreren Datensets in einer Visualisierung wird die Farbe eingesetzt, um sie zu unterscheiden. Die Farben sollen sich so stark wie möglich unterscheiden, damit die Zugehörigkeit auf den ersten Blick erkennbar ist. Sie sollen ästhetisch sein und zum Gesamtbild passen. Grelle und sehr dunkle Farben sollen vermieden werden um den Kontrast der Darstellung im Gleichgewicht zu halten. Nicht alle Menschen können alle Farben unterscheiden. Gewisse Farbkombinationen, wie Grün-Rot, sind suboptimal.

Abbildung 67 zeigt die Farbpalette zur Visualisierung von mehreren Traces in einer Darstellung. Die Palette wurde von Adobe [69] entwickelt.

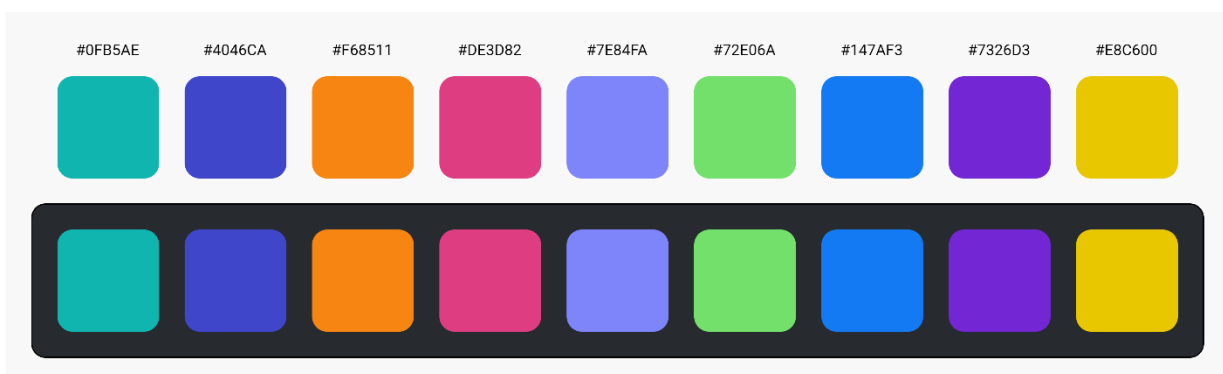


Abbildung 67 Datenvisualisierung - Kategorische Color Palette

Die ersten sechs Farben sind optimiert für Personen mit eingeschränktem Farbsehvermögen. Durch die mittlere Helligkeit eignen sich die Farben sowohl für den Einsatz im hellen als auch im dunklen Theme.

Diese Dokumentation enthält Abbildungen von Datenvisualisierungen mit unterschiedlichen Farbpaletten. Während der Entwicklung wurden verschiedenste Kombinationen getestet und analysiert.

#### 4.9.2 Time Series Chart

Zur Visualisierung von Zeitreihendaten wurde eine Komponente entwickelt, die verschiedene Darstellungstypen ermöglicht. Die Applikation beinhaltet sowohl interpolierbare als auch nicht-interpolierbare Datensätze, was eine flexible Visualisierung erfordert. Über das Burger Menu oberhalb des Charts auf der rechten Seite sind die Einstellungen erreichbar. Abbildung 68 zeigt die Schaltfläche und das geöffnete Menu.

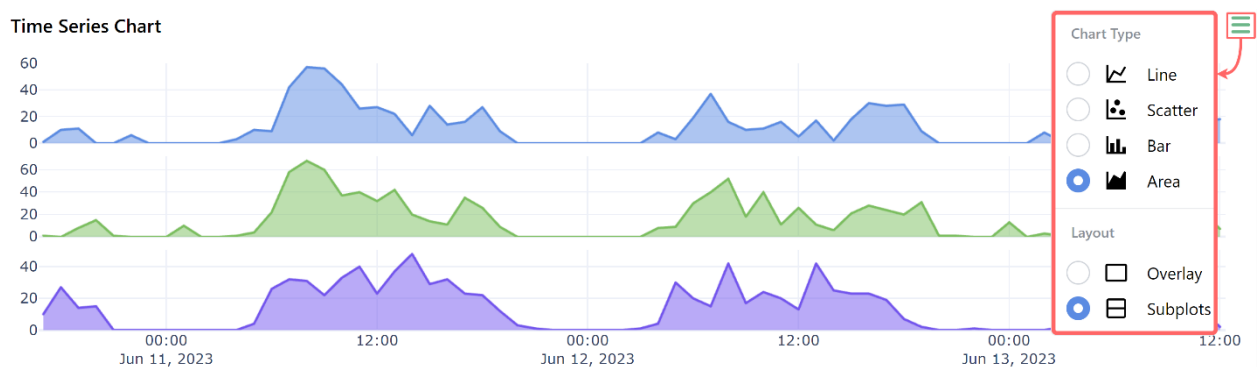


Abbildung 68 Darstellungsoptionen Time Series Chart

Im ersten Abschnitt des Menus ist der Visualisierungstyp wählbar. Beinhaltet die Visualisierung mehrere Traces, befindet sich in einem zweiten Abschnitt eine Auswahl zwischen Overlay- und Subplot Layout. Abbildung 69 zeigt die Layout Typen. Im Overlay Layout werden alle Traces übereinander visualisiert. Eine einheitliche Skalierung kann durch die Normalisierung von Datenreihen erreicht werden. Das Subplot Layout enthält pro Datensatz einen Subplot. Die Zeitachse wird gemeinsam genutzt. Eine Normalisierung ist in diesem Layout Typ nicht notwendig, da die Range der y-Achsen automatisch festgelegt wird.

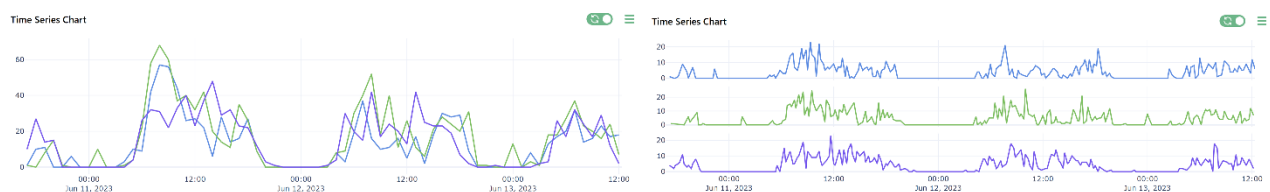


Abbildung 69 Chart Layout: Overlay (l) und Subplots (r)

#### 4.9.3 Hexbinmap

Die Hexbinmap wird im Map- und Taxon Dashboard eingesetzt. Es wurde eine All-In-One Komponente entwickelt, welche die Daten zwischenspeichert und die Interaktivität der Visualisierung gewährleistet.

Es existieren verschiedene Ansätze für die Implementierung einer Hexbinmap zur Darstellung von räumlichen Daten. Die Festlegung der Grössen und Positionen der Cluster ist dabei sehr wichtig. Die Hexbinmap Funktion von Plotly berechnet die Hexagons basierend auf den Datenpunkten und einem Parameter, der die maximale Anzahl horizontaler Zellen festlegt [70]. Die Abhängigkeit der Datenpunkte hat zur Folge, dass die Zellen je nach Ansicht an verschiedenen Orten und in verschiedenen Grössen berechnet werden.

Das Open Source Geo-Indexing System H3 von Uber [71] teilt die gesamte Erdoberfläche in Hexagonale Zellen ein. Insgesamt 15 Auflösungsstufen sind möglich. Jedes Cluster referenziert eine fixe Position und wird über eine ID referenziert. Eine Zellen ID besteht aus den ID's der darüberliegenden Zellen. Das bringt den Vorteil, dass für Zellen mit einer gröberen Auflösung anstelle einer Berechnung lediglich ein Lookup gemacht werden muss. Für die Darstellung der Hexbinmap wurde das H3-Konzept umgesetzt.

Für die Konstruktion der Cluster wird in einem ersten Schritt das Zoomlevel und die entsprechende Cluster Grösse aus den Koordinaten des Datensets berechnet. Anschliessend wird für jeden Datenpunkt eine Cluster ID berechnet. Für jedes vorhandene Cluster wird ein aggregierter Wert aus den zugehörigen Datenpunkten berechnet. Zur Visualisierung der Zellen auf der Karte wird aus den ID's ein GeoJSON Objekt generiert. Die Choroplethmapbox Funktion von Plotly stellt die Zellen dar und färbt sie entsprechend den aggregierten Werten ein.

Bei der Modifikation des Zoom Levels werden die Cluster kleiner bzw. grösser. Zoom-Events lösen eine Callback Funktion aus, welche die Visualisierung beim Erreichen von bestimmten Stufen aktualisieren. Aktualisierungen werden mit einem Loading Overlay kommuniziert. Die Interaktivität der Karte ist während den Ladevorgängen nicht eingeschränkt.

#### 4.9.4 FFT

Eine FFT zeigt Zeitreihendaten im Frequenzbereich. Als Visualisierungsmethode wurde ein Line Plot mit logarithmischen Achsen entwickelt. Die darzustellenden Resultate einer FFT Berechnung bestehen aus zwei Arrays. Das erste Array beinhaltet die Amplitudenwerte, das zweite repräsentiert die Frequenzen. Da es sich bei den Daten nicht um hochfrequente Signale handelt, wird anstelle der Frequenzangabe in Hertz die Dauer der Schwingung in Sekunden berechnet. Die Sekundenangaben werden für die Darstellung in leserfreundliche Zeitintervalle konvertiert.

Die Anzahl Datenpunkte einer FFT entspricht der Hälfte der Originalpunkte. Eine Zeitreihe mit einer Auflösung von 10 Minuten besteht aus 52'560 Datenpunkten pro Jahr. Wird davon eine FFT berechnet, besteht diese aus bis zu 26'280 Datenpunkten. Visualisierungen von grossen Datenmengen können sehr lange dauern und einen negativen Einfluss auf die gesamte Performance einer Applikation haben.

Zur Verringerung der zu visualisierenden Datenmenge wurde eine Funktion zur Aggregierung entwickelt. Die Punkte werden auf 256 Bins verteilt. Die Grösse der Bins ist logarithmisch. Pro Bin wird der Maximalwert der Amplitude berechnet und dargestellt. Abbildung 70 zeigt auf der linken Seite die rohen Datenpunkte einer FFT und auf der rechten Seite eine aggregierte FFT. Die Aggregierung hat den Nachteil, dass eine leichte Verschiebung der Frequenzen auftreten kann. Markante Informationen sind trotzdem klar erkennbar.

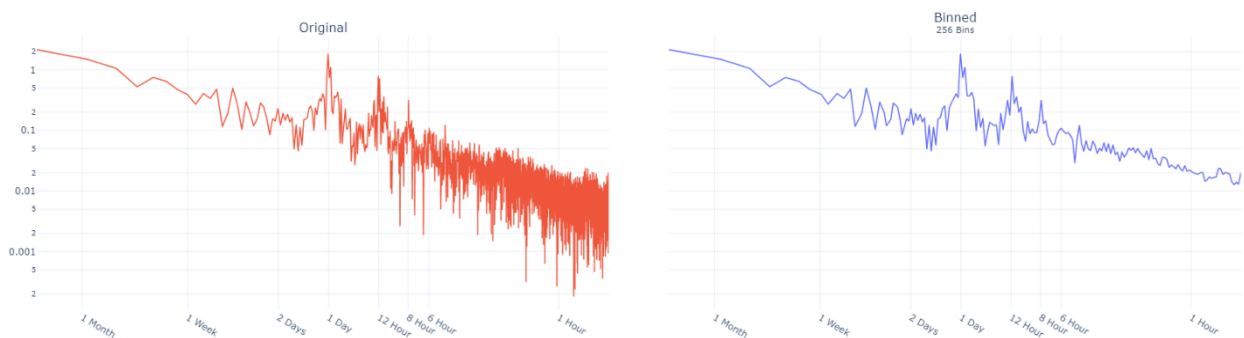


Abbildung 70 Datenvisualisierung - FFT Binning



#### 4.9.5 Korrelationsmatrix

Die Visualisierung der Korrelationsmatrix ist im Compare Dashboard enthalten, um Zusammenhänge zwischen verschiedenen Datensets zu analysieren. Die Korrelationsmatrix beinhaltet den Pearson Korrelationskoeffizienten für jedes Datenset-Paar. Die quadratische Matrix hat N-1 Spalten, wobei N die Anzahl Datensets repräsentiert. Bei vier Datensets ist die Dimension 3x3, da jedes Datenset mit den drei restlichen Datensets vergleichbar ist. Zur Visualisierung der Matrix wurde eine Heatmap implementiert, die in Abbildung 71 ersichtlich ist.

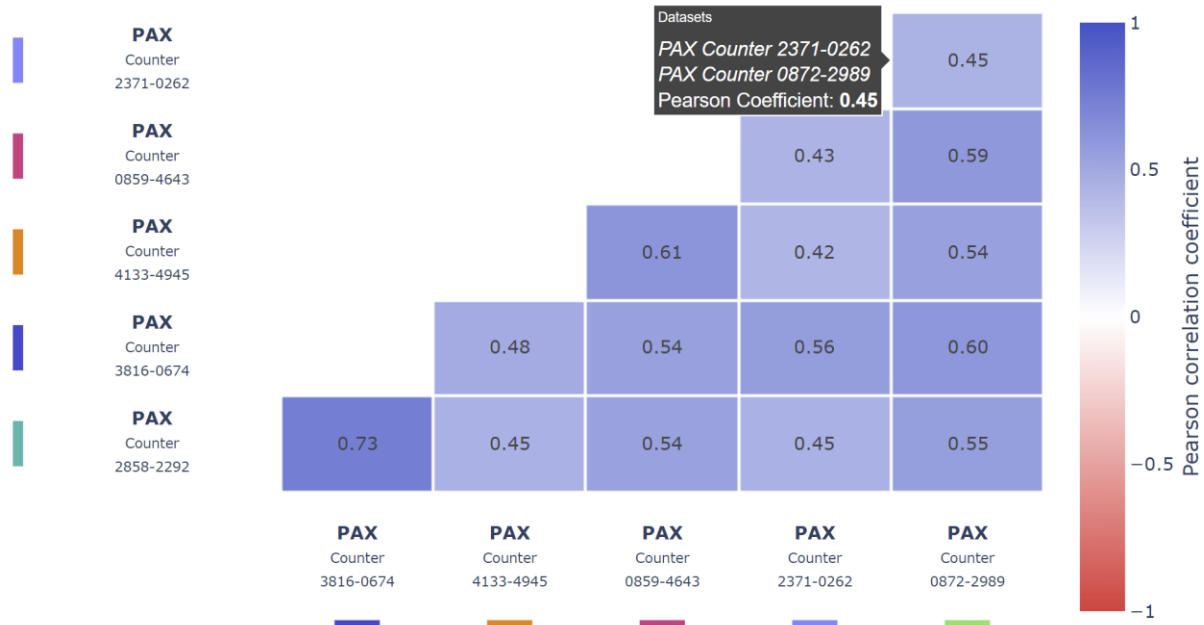


Abbildung 71 Datenvisualisierung - Korrelationsmatrix

Die Zellen sind, den Korrelationskoeffizienten entsprechend, mit einer divergierenden Farbskala eingefärbt. Starke positive oder negative Korrelationen werden somit hervorgehoben. Die Zellen sind zusätzlich mit den berechneten Werten beschriftet, damit weniger signifikante Unterschiede erkennbar sind. Die Spalten und die Zeilen sind mit den Namen der Datensets beschriftet und farblich gekennzeichnet. Insgesamt besteht die Visualisierung aus fünf Subplots, wovon vier zur Beschriftung dienen. Die graue Informationsbox in Abbildung 71 erscheint, sobald sich der Mauszeiger über einer Zelle befindet.

#### 4.9.6 Radar Plot

Der Radar Plot wird im Deployment Dashboard eingesetzt, um die Charakteristiken der Umgebung eines Deployments zu visualisieren. Zum Verfassungszeitpunkt dieser Dokumentation wurden zehn Attribute zur Beschreibung der Umgebung vorgeschlagen. Die gezeigten Werte wurden geschätzt und entsprechen nicht zwingend den realen Gegebenheiten. Ein Eintrag aus dem Datenset zur Beschreibung der Umgebungseigenschaften besteht aus den Koordinaten des Erhebungspunkts und Werten zur Beschreibung der Attribute. Die Werte sind ganzzahlig und liegen zwischen 0 und 10.

Ein Radar Plot beinhaltet pro Attribute eine Achse. Die Achsen sind radial angeordnet. Die auf den entsprechenden Achsen eingezeichneten Datenpunkte werden durch eine Linie verbunden, womit eine Form entsteht. Die Form stellt das 'Profil', der Umgebungseigenschaften dar. Es ist schnell erkennbar, ob starke Eigenschaften eher im linken oder im rechten Bereich der Visualisierung liegen.

Abbildung 72 zeigt den Radar Plot von einer Installation in den Merian Gärten. Die drei Erhebungspunkte mit der geringsten Distanz zum Deployment sind dargestellt. Eine Distanzangabe



in Metern ist in der Legende enthalten. Die relevanteste Information ist der Eintrag mit der geringsten Distanz. Er wird mit einer Farbfüllung hervorgehoben. Die zwei weiteren Einträge sind halbtransparent eingezeichnet.

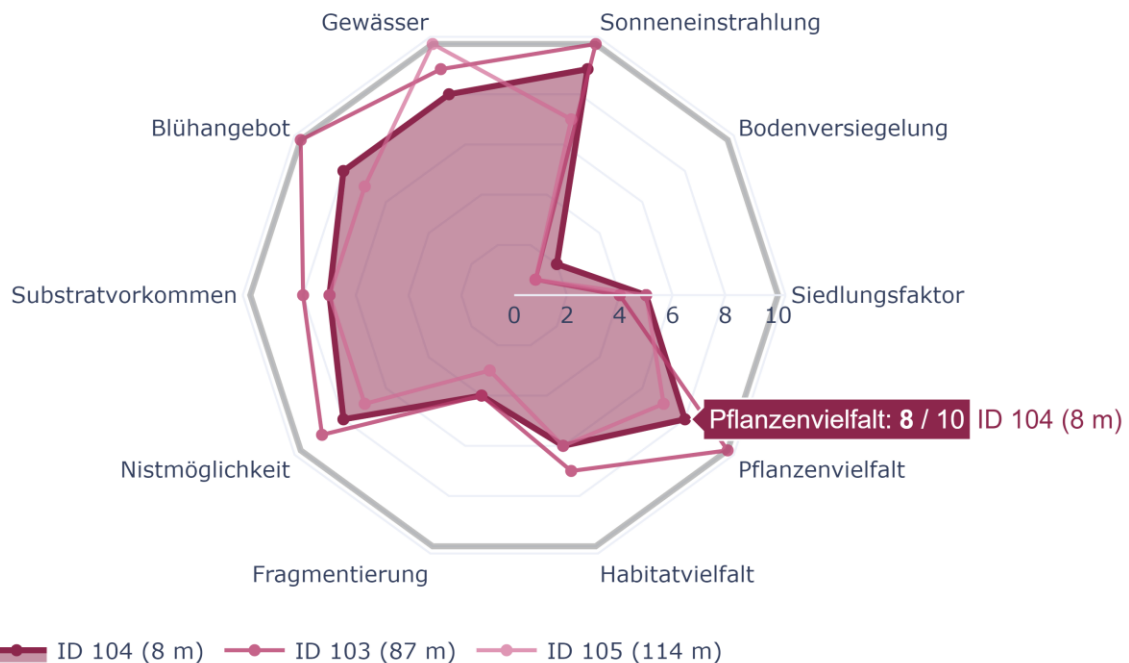


Abbildung 72 Datenvisualisierung - Radar Plot

Die Visualisierten Einträge stehen in keiner Abhängigkeit zueinander. Bei der Exploration der erhobenen Daten können Umgebungscharakteristiken von Orten, die nicht direkt neben dem Deployment liegen, auch relevant sein.

#### 4.10 User Management

Aus einer Vielzahl von User Management Systemen wurden Keycloak [72] und Auth0 [73] als potentielle Lösungen identifiziert und verglichen. Auth0 bietet Authentication-as-a-Service und ist kostenlos für Projekte mit bis zu 7'000 Nutzenden. Gewisse erforderliche Funktionen, zum Beispiel Role Management, sind in der kostenlosen Version jedoch nicht verfügbar. Keycloak ist ein Open Source IAM System, welches lokal betrieben werden kann und alle wichtigen Funktionen beinhaltet.

Als User Management System wurde Keycloak eingesetzt. Die Applikation läuft in einem Docker Container im Mitwelten Backend. Die Daten werden in einer PostgreSQL Datenbank, die auf dem selben Backends läuft, gespeichert. Als Mail Provider wurde ein Gmail Konto konfiguriert. Die Registrierung erfordert eine Bestätigung der E-Mail Adresse. Falls eine registrierte Person ihr Passwort vergisst, hat sie die Möglichkeit, eine E-Mail mit einem Link zur Rücksetzung des Passworts anzufordern.

Nach der erfolgten Anmeldung in Keycloak wird im Client ein Cookie gesetzt, welches den JWT beinhaltet. In dem FastAPI Server des Tools durchläuft jeder Request eine Middleware, die prüft, ob ein gültiger Token im Header oder als Cookie enthalten ist. Falls kein gültiger Token vorhanden ist, wird der Client auf Keycloak weitergeleitet, um sich anzumelden.

In regelmässigen Abständen wird die Gültigkeitsdauer des Tokens überprüft. Falls die Dauer unter einen definierten Bereich fällt, wird der Token erneuert. Für die Anzeige des Namens der eingeloggten Person wird das gesetzte Cookie decodiert.

#### 4.10.1 Rollen

In der Admin Konsole von Keycloak können Rollen an Personen zugewiesen werden, um Berechtigungen individuell zu vergeben. Die Berechtigungen werden nach dem additiven Prinzip vergeben. Im additiven Prinzip werden Rechte durch Rollen vergeben, Verweigerungen gibt es nicht. Zum Zeitpunkt der Verfassung dieser Dokumentation wurden folgende Rollen definiert:

- Die Rolle `public` wird jeder Person direkt nach der Registrierung automatisch zugewiesen und verschafft keine speziellen Berechtigungen.
- Die Rolle `internal` wird den Mitgliedern des Projekts Mitwelten manuell zugewiesen. Sie verleiht in anderen Tools Leseberechtigungen auf Rohdaten wie Bilder aus der Bestäuber Studie oder Audioaufnahmen. Schreibberechtigungen werden für das Management von Deployments erteilt.
- Die Rolle `explore_admin` wurde für das Explore Tool definiert. Sie berechtigt zum löschen von allen Annotationen.

### 4.11 Datenmanagement

#### 4.11.1 Erweiterung Mitwelten Datenbank

Damit Informationen zur Taxonomie in der Datenbank vorhanden sind, wurden zwei neue Tabellen erstellt, die in Abbildung 73 ersichtlich sind. Jedes Taxa, welches im Projekt Mitwelten erfasst wurde, ist in der Table `taxonomy_data` eingetragen. Die deutschen und englischen Namen wurden anhand des GBIF Taxonomy Backbone Datenset übersetzt. Für eine systemübergreifende einheitliche Indexierung der Taxa zu erreichen wurde der GBIF Taxon Key als `datum_id` gespeichert. Für jedes Taxa ist die URL auf ein Referenzbild von Wikipedia aufgeführt. Jedes vorhandene Taxon ist in der Table `taxonomy_tree` enthalten, die Informationen zum Stammbaum enthält. Als IDs wurden GBIF Taxon Key eingesetzt.

taxonomy_data		taxonomy_tree	
<code>datum_id</code>	<code>int8</code>	<code>species_id</code>	<code>int8</code>
<code>label_sci</code>	<code>varchar</code>	<code>genus_id</code>	<code>int8</code>
<code>label_de</code>	<code>varchar</code>	<code>family_id</code>	<code>int8</code>
<code>label_en</code>	<code>varchar</code>	<code>order_id</code>	<code>int8</code>
<code>image_url</code>	<code>varchar</code>	<code>class_id</code>	<code>int8</code>
<code>created_at</code>	<code>timestampz</code>	<code>phylum_id</code>	<code>int8</code>
<code>updated_at</code>	<code>timestampz</code>	<code>kingdom_id</code>	<code>int8</code>

Abbildung 73 Mitwelten Datenbank - Taxonomie Daten

Abfragen von Tierbeobachtungen werden anhand des GBIF Taxon Keys gemacht. Für die Spalten mit den wissenschaftlichen Namen in der Tabelle `'birdnet_results'` und für die Spalte `'label_sci'` in `'taxonomy_data'` wurden Indexe generiert. Für eine Abfrage anhand des GBIF Keys werden die Einträge nach den wissenschaftlichen Namen gefiltert, auf die der Taxon Key referenziert.

Die Nutzerdaten von Keycloak sind in einer unabhängigen PostgreSQL Datenbank gespeichert. Name, Benutzername und Email Adresse werden in der Table `user_entity` gespeichert. Jeder Person wird bei der Registrierung eine eindeutige ID zugewiesen, die auch im JWT aufgeführt ist. Um aus der Mitwelten Datenbank auf die User-Informationen zuzugreifen wurde ein Foreign Data Wrapper mit Leseberechtigung auf die Table erstellt. Abbildung 74 zeigt die neuen Tabellen

in der Mitwelten Datenbank, die auf die Nutzerinformationen zugreifen. Die Collection-Inhalte der Nutzenden sind zusammen mit der User-ID in der Table `user_collections` gespeichert. Jede Annotation wurde von einer registrierten Person erfasst. Für die Anzeige der Information, wer eine Annotation verfasst hat, wird der Name aus der Keycloak Table abgefragt.

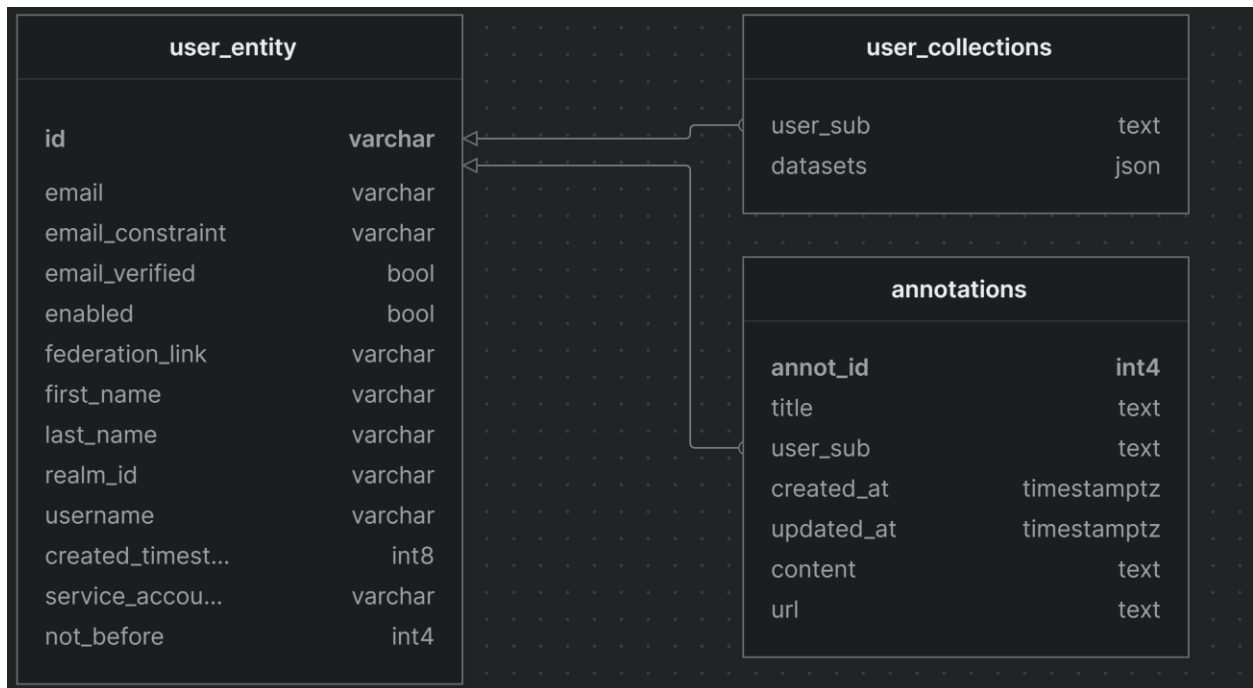


Abbildung 74 Datenbankschema Mitwelten - Annotations und Collections

Für die Speicherung der Umgebungscharakteristiken wurde eine neue Table erstellt. Sie beinhaltet die Koordinaten des Ortes und die Zeit der Erhebung. Für die Werte sind zehn Spalten vorhanden. Die Bezeichnungen und die Beschreibungen der Attribute sind in einem JSON File gespeichert.

#### 4.11.2 Caching von Daten aus externen Quellen

Beobachtungen aus externen Quellen, die via GBIF bezogen wurden und Meteodaten wurden in der zusätzlichen Datenbank 'mitwelten\_cache' gespeichert. Abbildung 75 zeigt das Datenbankschema. Informationen zu den Meteodaten sind in drei Tabellen unterteilt. Jeder Messpunkt referenziert auf die Station, wo er erfasst wurde. Die Parameterbeschreibungen und die Einheiten der Messpunkte sind in der Tabelle 'parameter' aufgeführt. Die Tabelle 'meteodata' wurde mit TimescaleDB in eine zeitlich indexierte Hypertable umgewandelt, um die einzelnen Zeitreihen möglichst effizient abzufragen.

Beobachtungen aus externen Quellen, die über GBIF bezogen wurden, sind in der Tabelle 'gbif' gespeichert. Neben den GBIF Taxon Keys für alle Taxonomie Ebenen beinhalten die Einträge Informationen zur Zeit und zum Ort der Beobachtung. Das Datenset, aus welchem eine Beobachtung stammt, ist mit dem Namen und als URL enthalten. Gewisse Beobachtungen enthalten Medieninhalte wie Bilder, Audioaufnahmen oder Videos. Die Inhalte werden, falls vorhanden, in der Spalte 'media' verlinkt.

Auf GBIF kann jede Beobachtung mit allen dazu vorhandenen Informationen angezeigt werden. Über den 'key' wird die GBIF-ID einer Beobachtung referenziert. Die URL auf die Beobachtungsseite kann anhand des keys generiert werden.

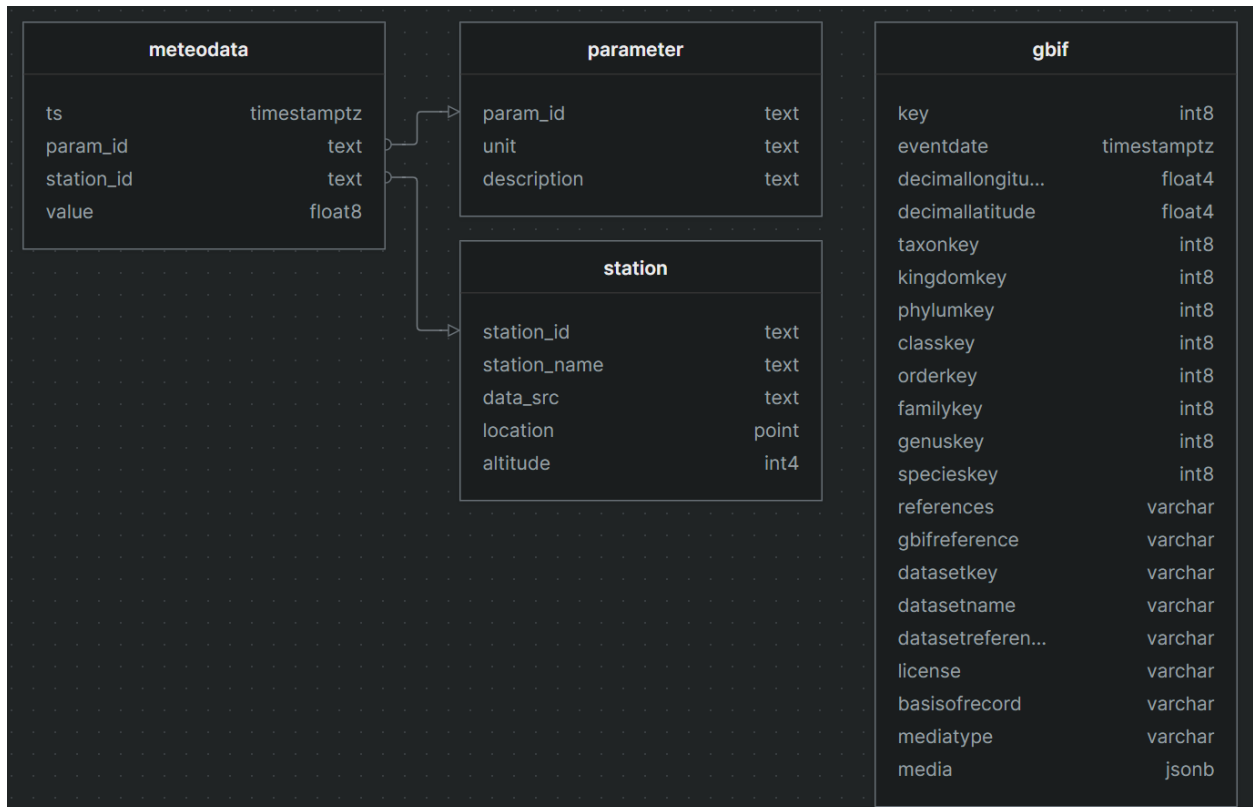


Abbildung 75 Datenbankschema Mitwelten Cache

Die Meteodaten werden vom IDAWEB bezogen und als komprimierte ZIP Dateien heruntergeladen. Ein Datensatz besteht aus zwei Dateien. Informationen zu den Stationen und den Parametern sind als Metadaten in einem CSV File aufgeführt. Das zweite CSV File enthält eine Liste von Messwerten mit den Erfassungszeitpunkten.

#### 4.11.2.1 Einfügen von Meteodaten

Für das Einfügen der Datenpakete in die Datenbank wurde ein CLI Tool [74] entwickelt. Neben der ZIP Datei muss für die Ausführung der Pfad zu einer Konfigurationsdatei in den Argumenten enthalten sein. Die Konfigurationsdatei beinhaltet alle Informationen für den Datenbankzugriff.

In einem ersten Schritt wird die Datei mit den Metadaten geöffnet und geparkt. Alle Stationen und Parameter werden, falls noch nicht vorhanden, in die entsprechenden Tabellen eingefügt. Anschliessend wird die CSV Datei mit den Messwerten geöffnet und geparkt. Der Schreibvorgang wird in Batchoperationen durchgeführt.

#### 4.11.2.2 Aktualisieren der GBIF Beobachtungen

Die Beobachtungen aus externen Quellen werden von der GBIF Occurrence API bezogen. Für die Aktualisierung der zwischengespeicherten Daten wurde ein Jupyter Notebook [74] entworfen.

Parameter zur räumlichen Eingrenzung, die Taxonomiegruppe und die Zeitspanne der Beobachtungen werden in einer ersten Zelle festgelegt. Als Mittelpunkt wurde das Zentrum der Merian Gärten definiert. Als maximale Entfernung in x- und y Richtung wurden 20 KM definiert. Aus diesen Informationen generiert das Skript ein Raster, welches den Bereich in mehrere Zellen aufteilt. Die Auflösung kann bei Bedarf angepasst werden.

Die Anzahl Beobachtungen pro Zelle wird anschliessend mit parallelen Requests abgefragt und in einer Grafik visualisiert. Abbildung 76 zeigt auf der linken Seite das generierte Raster. Auf der rechten Seite wird die Summe von Vogel Beobachtungen pro Zelle visualisiert. Eine Zelle darf,

bedingt durch eine Limite von GBIF, maximal 100'000 Beobachtungen enthalten. Bei einer Überschreitung dieses Wertes wird ein Fehler angezeigt.

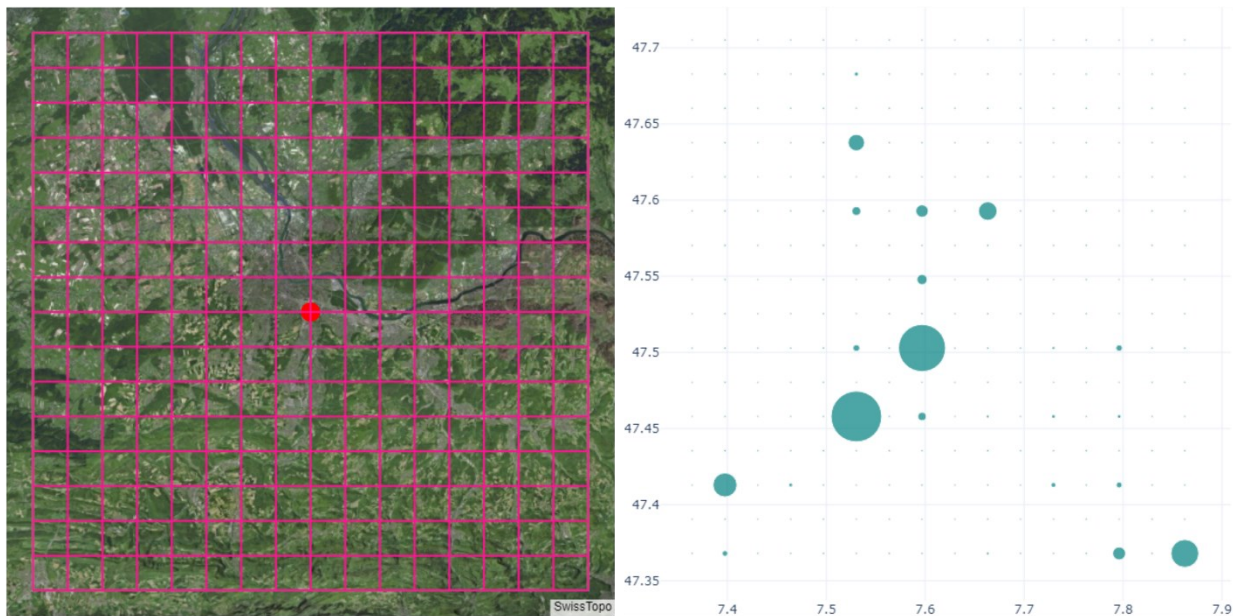


Abbildung 76 GBIF Cache Update - Grid & Observations per Cell

Die Beobachtungen werden zellenweise in mehreren Threads heruntergeladen, validiert und geparkt. Bei der Validierung wird geprüft, ob der Erfassungszeitpunkt einer Beobachtung einem realen Datum entspricht. Diese Überprüfung wird gemacht, da es Einträge auf GBIF mit nicht existierenden Daten gibt. Die Herkunft einer Beobachtung wird über eine Datenset-ID referenziert. Für jede eindeutige Datenset-ID wird anschliessend eine Abfrage durchgeführt, um den Namen des Datensets zu ermitteln. Sobald alle Informationen bekannt sind, können die Werte in die Datenbank eingefügt werden. Für jede Beobachtung wird geprüft, ob sie bereits existiert. Der Schreibvorgang erfolgt in mehreren Batches.

#### 4.11.3 Bereitstellung der Daten

Für die Bereitstellung von Daten wurde die REST API im Mitwelten Backend um die entsprechenden Routen ergänzt. Für die Endpunkte wurde eine standardisierte Grundstruktur festgelegt. Zum Filtern der Datenpunkte wurden einheitliche Parameter festgelegt. Der Erfassungszeitraum ist durch die Parameter `from` und `to` für alle Datensets eingrenzbar. Die ergänzten Routen sind in der Mitwelten Data API Doc dokumentiert.

#### 4.11.4 Caching

Um die Dashboards möglichst effizient zu gestalten wurde ein Caching von gewissen Requests an die Mitwelten Daten API und an die Wikipedia API implementiert. Die Daten werden in einer Redis Datenbank zwischengespeichert, die auf dem Mitwelten Backend in einem Docker Container läuft. Für das Management des Caches wurde die Python Library `requests-cache` [75] eingesetzt.

Die Gültigkeitsdauer von zwischengespeicherten Anfragen beträgt bei komplexen Abfragen 2 Stunden. Damit die Werte von aktiven PAX Deployments in der Applikation zeitnah verfügbar sind, werden die entsprechenden Requests für eine kürzere Dauer zwischengespeichert. Abfragen von Informationen wie Annotationen, die sich regelmässig ändern, werden nicht im Cache gespeichert. Generell werden alle Requests, die eine Authentifizierung erfordern, vom Cache ausgeschlossen.

#### 4.12 Datenbezug für externe Tools

Die Explore Anwendung enthält Funktionen zur explorativen Visualisierung und Analyse Datensätzen. Für Fachpersonen aus dem Bereich Biologie oder aus verwandten Gebieten soll die Möglichkeit bestehen, die Daten in unabhängigen Tools zu explorieren. Für solche Analysen wird in vielen Fällen Python oder R genutzt. Eine Dokumentation der Mitwelten Data API wird automatisch generiert und ist unter der Route '/docs' verfügbar. Für jede Route werden erforderliche und optionale Query-Parameter aufgeführt.

Es wurden verschiedene Codebeispiele entwickelt, um Abfragen der Mitwelten Data API in Python, R und JavaScript zu demonstrieren.

Die Python-Beispiele wurden in Form von Jupyter Notebooks erstellt. Als Library für die Visualisierungen wurde Plotly verwendet. Da GitHub keine interaktiven Visualisierungen in Jupyter Notebooks anzeigt, wurden die Grafiken statisch eingebettet. Die Funktionen für Datenabfragen von Vogel- und Bestäuber Detektionen wurden in einem Helper File definiert. Das Pollinator Beispiel demonstriert den Gebrauch der entsprechenden Funktionen.

Mehrere RMarkdown Notebooks wurden entwickelt, die Datenabfragen und Visualisierungen in R beinhalten. Die Notebooks wurden als Markdown Files exportiert, damit sie auf GitHub formatiert dargestellt werden. Für einfache Visualisierungen wurde ggplot2 eingesetzt, Kartendarstellungen wurden mit Plotly generiert. Die Codebeispiele in R und Python sind im Repository Mitwelten Explore Data Management [74] verfügbar.

Auf der Plattform ObservableHQ wurden mehrere Notebooks entworfen, um die Abfrage und Visualisierung von Daten mit JavaScript zu demonstrieren. Die Beispiele sind öffentlich und können interaktiv genutzt werden. Die Notebooks sind im Anhang verlinkt.

### **4.13 Mögliche Erweiterungen des Tools**

Alle geforderten Funktionen und auch eine Reihe von zusätzlichen Funktionen wurden in der vorhandenen Zeit implementiert und getestet. Mögliche Erweiterungen der Applikation werden folgend beschrieben.

#### **4.13.1 Bereitstellung von Rohdaten**

In der Applikation werden nur ausgewertete Datenpunkte der Bird- und Pollinator Studie visualisiert. Für eine Validierung dieser Datenpunkte ist eine Bereitstellung der Rohdaten erforderlich. Es ist nicht auszuschliessen, dass sich sensible Personendaten in den Rohdaten befinden. Audioaufnahmen können persönliche Gespräche enthalten und in den Bildern können Personen erkennbar sein. Die Daten dürften also nur für eine bestimmte Nutzergruppe erfolgen. Ein Verweis auf die Originaldaten aus einem Dashboard kann identisch zu den GBIF Referenzen implementiert werden. Für jedes Dashboard würde eine Liste generiert werden, die auf die Rohdaten verweist.

Für die Bereitstellung der Audioaufnahmen wäre es möglich, die Ausschnitte der Files zu streamen, in welchen die Vögel detektiert wurden.

#### **4.13.2 Datenexport**

In den Dashboards werden Ansichten auf Datensets mit interaktiven Grafiken visualisiert. Die Applikation könnte um eine Funktion erweitert werden, die es nutzenden ermöglicht, die angezeigten Datenpunkte als Datei herunterzuladen.

#### **4.13.3 Bat Detections**

Neben den Vogel- und Bestäuber Detektionen könnten Detektionen von Fledermäusen in das Tool integriert werden. Im Projekt Mitwelten existieren gewisse Audioaufnahmen mit einem sehr breiten Frequenzbereich. Falls eine automatisierte Auswertung der Daten möglich ist, könnten die Resultate in der Applikation aufgeführt werden.

#### **4.13.4 Environmental Confidence Factor**

Für die Validierung von detektierten Spezies könnte für jede Art, Gattung oder Familie auf Basis von Fachwissen und zusätzlichen Hintergrundinformationen ein Faktor berechnet werden, der die Wahrscheinlichkeit einer korrekten Erkennung beschreibt. Somit wäre es möglich, fehlerhaft ausgewertete Daten zu identifizieren oder seltene Arten hervorzuheben.

#### **4.13.5 Dynamisches Hinzufügen von Map Layern**

In der Applikation wird eine Auswahl von relevanten Map Layern bereitgestellt. Eine mögliche Erweiterung besteht in der Darstellung von benutzerdefinierten Map Layern. Diese Funktionalität erfordert ein Eingabeelement für die URL des entsprechenden Layers.

#### **4.13.6 Medieninhalte für Annotationen**

Der Inhalt einer Annotation besteht aus einem Text im Rich Text Format. Als Erweiterung könnte das Anfügen von Dateien wie PDF Dokumenten, Bildern, Videos oder Audio Files ermöglicht werden. Die Speicherung der Daten könnte in S3 erfolgen, wo auch die Rohdaten liegen.

Für diese Erweiterung ist es sehr wichtig klar festzulegen, wer Inhalte hochladen darf und wer für die Inhalte verantwortlich ist. Falls jede angemeldete Person Inhalte publizieren darf, ist eine Überprüfung der Inhalte vor der Freischaltung von Vorteil, um rechtliche Konflikte zu vermeiden.



## 5 Fazit

Im Rahmen dieser Projektarbeit wurde erforscht, wie ein Tool die Exploration von ökologisch relevanten Daten unterstützen kann. Als Hauptprodukt wurde eine Applikation zur Visualisierung und Analyse von Daten konzipiert, entwickelt und getestet.

Die Hauptgebiete dieser Projektarbeit umschliessen das Datenmanagement, UX/UI Design, Datenvisualisierung und User Management. In einer ersten Projektphase wurden die Grundlagen dieser Gebiete erarbeitet und dokumentiert.

In der zweiten Phase wurde analysiert, wie die vorhandenen Daten gefiltert werden können, um Teilmengen daraus zu erkunden und zu vergleichen. Für die Erforschung von weiteren Einflussfaktoren auf die Biodiversität, die Kontextualisierung der Daten und für deren Validierung wurden mit einer Recherche zusätzliche Datenquellen gefunden. Für die Aggregation, Analyse und die effiziente Bereitstellung von Daten wurden Konzepte entworfen und getestet. Für die Bedienung der Applikation wurde ein verständlicher Workflow erarbeitet, der an die Bedienung eines Onlineshops erinnert. Die erarbeiteten UX/UI Konzepte wurden beim Design der Dashboard Layouts angewendet, um eine möglichst intuitive Nutzung zu ermöglichen. Damit Forschende ihre Erkenntnisse mit anderen Personen teilen können, wurde ein Konzept zur Verfassung von Annotationen erarbeitet. Ein IAM-System wurde eingeführt, um eine Anmeldung mit persönlichen Profilen zu ermöglichen. Neben der Explore Applikation wurde es als SSO Lösung für alle wichtigen Mitwelten-Applikationen eingesetzt, was sowohl die Benutzungsfreundlichkeit und auch die Sicherheit erhöht.

In der letzten Projektphase wurde die Applikation nach dem iterativen Entwicklungsprinzip implementiert. Bereits in frühen Phasen der Entwicklung wurden erste Versionen von mehreren Personen getestet, um Feedback zu erhalten. Für die Bereitstellung der Daten aus dem Mitwelten Projekt wurde die bestehende API ergänzt. Die öffentliche API ermöglicht die Analyse von Datensätzen in unabhängigen Tools. Datenabfragen in Python, R und JavaScript wurden anhand von Codebeispielen demonstriert.

Das Ergebnis dieses Projekts ist eine umfassende, effizient funktionierende Applikation, die es ermöglicht, Biodiversität zu erkunden, Einflussfaktoren zu identifizieren und Erkenntnisse mit anderen zu teilen. Die Applikation vereint Daten aus verschiedenen Quellen, um ein umfassendes Verständnis der Biodiversität zu ermöglichen.

Fachpersonen aus dem Projekt Mitwelten konnten die Applikation bereits während der Entwicklung produktiv nutzen und waren mit der Funktionalität zufrieden. Die Rückmeldungen aus dem Mitwelten-Team waren durchweg positiv. Mehrere Personen aus meinem privaten Umfeld haben die Applikation getestet und fanden sich ohne zusätzliche Instruktionen problemlos zurecht.

Insgesamt wurden die in der Aufgabenstellung festgelegten Ziele erreicht und das Projekt wurde erfolgreich abgeschlossen.

## 6 Literaturverzeichnis

- [1] Coursera, "Principles of UX/UI Design," *Coursera*, 2023. <https://www.coursera.org/learn/principles-of-ux-ui-design?> (accessed Sep. 05, 2023).
- [2] J. Yablonski, *Laws of UX: Using Psychology to Design Better Products & Services*. O'Reilly Media, Inc., 2020.
- [3] J. Nielsen, "F-Shaped Pattern For Reading Web Content (original eyetracking research)," *Nielsen Norman Group*, Apr. 16, 2006. <https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/> (accessed Aug. 02, 2023).
- [4] J. Cao, "Designing Websites That Mirror How Our Eyes Work (Part 2)," *TNW | Dd*, Apr. 15, 2015. <https://thenextweb.com/news/designing-websites-that-mirror-how-our-eyes-work-part-2> (accessed Aug. 02, 2023).
- [5] K. Pernice, "F-Shaped Pattern of Reading on the Web: Misunderstood, But Still Relevant (Even on Mobile)," *Nielsen Norman Group*, Nov. 12, 2017. <https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/> (accessed Aug. 02, 2023).
- [6] F. Vassilatos, "Dashboard Design UX Patterns Best Practices," *Pencil & Paper*, Jul. 11, 2023. <https://pencilandpaper.io/articles/ux-pattern-analysis-data-dashboards/> (accessed Aug. 02, 2023).
- [7] A. Harley, "Visibility of System Status," *Nielsen Norman Group*, Jun. 03, 2018. <https://www.nngroup.com/articles/visibility-system-status/> (accessed Aug. 02, 2023).
- [8] S. Farrell, "Navigation: You Are Here," *Nielsen Norman Group*, Oct. 25, 2015. <https://www.nngroup.com/articles/navigation-you-are-here/> (accessed Aug. 03, 2023).
- [9] *Jakob's Law of Internet User Experience (2 min. video) (Video)*, (Aug. 18, 2017). Accessed: Aug. 02, 2023. [Online Video]. Available: <https://www.nngroup.com/videos/jakobs-law-internet-ux/>
- [10] R. Krause, "Maintain Consistency and Adhere to Standards (Usability Heuristic #4)," *Nielsen Norman Group*, Jan. 10, 2021. <https://www.nngroup.com/articles/consistency-and-standards/> (accessed Aug. 02, 2023).
- [11] M. Rosala, "User Control and Freedom (Usability Heuristic #3)," *Nielsen Norman Group*, Nov. 29, 2020. <https://www.nngroup.com/articles/user-control-and-freedom/> (accessed Aug. 02, 2023).
- [12] A. Joyce, "Help and Documentation: The 10th Usability Heuristic," *Nielsen Norman Group*, Dec. 13, 2020. <https://www.nngroup.com/articles/help-and-documentation/> (accessed Aug. 02, 2023).
- [13] T. Fessenden, "Aesthetic and Minimalist Design (Usability Heuristic #8)," *Nielsen Norman Group*, Jan. 24, 2021. <https://www.nngroup.com/articles/aesthetic-minimalist-design/> (accessed Aug. 03, 2023).
- [14] J. Schwabisch, *Better Data Visualizations*. Columbia University Press, 2021.
- [15] C. O. Wilke, *Fundamentals of Data Visualization*. O'Reilly and Associates, 2019. Accessed: Aug. 11, 2023. [Online]. Available: <https://clauswilke.com/dataviz/visualizing-proportions.html>
- [16] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007. doi: 10.1109/MCSE.2007.55.
- [17] Plotly, "plotly.py." Plotly, Jul. 28, 2023. Accessed: Jul. 28, 2023. [Online]. Available: <https://github.com/plotly/plotly.py>
- [18] M. Waskom, "seaborn: statistical data visualization." Jul. 28, 2023. Accessed: Jul. 28, 2023. [Online]. Available: <https://github.com/mwaskom/seaborn>
- [19] M. Narang, "Top 10 Python Libraries for Data Visualization," Jul. 14, 2023. <https://www.knowledgehut.com/blog/business-intelligence-and-visualization/python-data-visualization-libraries> (accessed Jul. 28, 2023).
- [20] Plotly, "plotly/plotly.js." Plotly, Jul. 28, 2023. Accessed: Jul. 28, 2023. [Online]. Available: <https://github.com/plotly/plotly.js>
- [21] Mozilla, "WebGL: 2D and 3D graphics for the web - Web APIs | MDN," Jul. 12, 2023. [https://developer.mozilla.org/en-US/docs/Web/API/WebGL\\_API](https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API) (accessed Jul. 28, 2023).
- [22] IBM Cloud Education, "Structured vs. Unstructured Data: What's the Difference?," *IBM Blog*, Jun. 29, 2021. <https://www.ibm.com/blog/structured-vs-unstructured-data/> (accessed Sep. 04, 2023).

- [23] PostgreSQL, "9.16. JSON Functions and Operators," *PostgreSQL Documentation*, Aug. 10, 2023. <https://www.postgresql.org/docs/15/functions-json.html> (accessed Sep. 04, 2023).
- [24] M. B. Jones, J. Bradley, and N. Sakimura, "JSON Web Token (JWT)," Internet Engineering Task Force, Request for Comments RFC 7519, May 2015. doi: 10.17487/RFC7519.
- [25] "BirdNET Sound ID – The easiest way to identify birds by sound." <https://birdnet.cornell.edu/> (accessed Jun. 26, 2023).
- [26] "AudioMoth | Open Acoustic Devices," *openacousticdevices*. <https://www.openacousticdevices.info/audiomoth> (accessed Sep. 21, 2021).
- [27] GBIF Secretariat, "GBIF Backbone Taxonomy", doi: 10.15468/39omei.
- [28] MediaWiki, "API:REST API - MediaWiki." [https://www.mediawiki.org/wiki/API:REST\\_API](https://www.mediawiki.org/wiki/API:REST_API) (accessed Jul. 10, 2023).
- [29] "Hauptseite - Wikispecies." <https://species.wikimedia.org/wiki/Hauptseite> (accessed Jul. 10, 2023).
- [30] Bundesamt für Meteorologie und Klimatologie, "MeteoSchweiz." <https://www.meteoschweiz.admin.ch/> (accessed May 18, 2023).
- [31] "Datenportal für Lehre und Forschung (IDAweb) - MeteoSchweiz." <https://www.meteoschweiz.admin.ch/service-und-publikationen/service/wetter-und-klimaprodukte/datenportal-fuer-lehre-und-forschung.html> (accessed May 18, 2023).
- [32] GBIF Occurrences, "Occurrence Metrics." [https://www.gbif.org/occurrence/charts?taxon\\_key=212&year=2020,\\*&year=\\*,2022&advanced=1&occurrence\\_status=present&gadm\\_gid=CHE.4\\_1&gadm\\_gid=CHE.5\\_1](https://www.gbif.org/occurrence/charts?taxon_key=212&year=2020,*&year=*,2022&advanced=1&occurrence_status=present&gadm_gid=CHE.4_1&gadm_gid=CHE.5_1) (accessed Jun. 26, 2023).
- [33] T. Auer *et al.*, "EOD – eBird Observation Dataset", doi: 10.15468/aomfmb.
- [34] "BirdNET – Apps bei Google Play." [https://play.google.com/store/apps/details?id=de.tu\\_chemnitz.mi.kahst.birdnet&hl=de\\_CH](https://play.google.com/store/apps/details?id=de.tu_chemnitz.mi.kahst.birdnet&hl=de_CH) (accessed Jun. 26, 2023).
- [35] iNaturalist, "iNaturalist," *iNaturalist*. <https://www.inaturalist.org/> (accessed Jul. 19, 2022).
- [36] GBIF, "GBIF occurrence API." <https://www.gbif.org/developer/occurrence> (accessed Jul. 10, 2023).
- [37] Swisstopo, "Geodaten und Applikationen," *Bundesamt für Landestopografie swisstopo*. <https://www.swisstopo.admin.ch/de/geodata.html> (accessed Jul. 07, 2023).
- [38] B. für U. B. | O. fédéral de l'environnement O. | U. federale dell'ambiente UFAM, "Umwelt- und Geodaten des BAFU." <https://www.bafu.admin.ch/bafu/de/home/daten--indikatoren--karten/umwelt--und-geodaten-des-bafu.html> (accessed Jul. 07, 2023).
- [39] Delarze *et al.*, *Lebensräume der Schweiz*. 2015.
- [40] geo.admin.ch, "GeoAdmin API." <https://api3.geo.admin.ch/index.html> (accessed Jul. 07, 2023).
- [41] F. Vassilatos, "Filter UX Design Patterns - Analysis & Best Practices," *Pencil & Paper*, Apr. 17, 2023. <https://pencilandpaper.io/articles/ux-pattern-analysis-enterprise-filtering/> (accessed Aug. 16, 2023).
- [42] N. Srivastava, J. Newn, and E. Velloso, "Combining Low and Mid-Level Gaze Features for Desktop Activity Recognition," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 4, p. 189:1-189:27, Dec. 2018, doi: 10.1145/3287067.
- [43] N. Babich, "Floating Action Button in UX Design," *Medium*, May 30, 2020. <https://uxplanet.org/floating-action-button-in-ux-design-7dd06e49144e> (accessed Aug. 22, 2023).
- [44] Laerd Statistics, "Pearson's product moment correlation. Statistical tutorials and software guides.," 2020. <https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php> (accessed Jul. 26, 2023).
- [45] NTI Audio, "FFT," *FFT*, 2021. <https://www.nti-audio.com/de/service/wissen/fast-fourier-transformation-fft> (accessed Jul. 28, 2023).
- [46] plotly, "Dash." Plotly, Jul. 17, 2023. Accessed: Jul. 17, 2023. [Online]. Available: <https://github.com/plotly/dash>
- [47] streamlit, "Welcome to Streamlit 🍷." Streamlit, Sep. 08, 2023. Accessed: Sep. 08, 2023. [Online]. Available: <https://github.com/streamlit/streamlit>
- [48] Tableau, "Tableau Cloud | Schnelle, flexible und benutzerfreundliche Analytics in der Cloud." <https://www.tableau.com/de-de/products/cloud-bi> (accessed Sep. 08, 2023).

- [49]Pallets, "Welcome to Flask — Flask Documentation (1.1.x)," 2020. <https://flask.palletsprojects.com/en/1.1.x/> (accessed Aug. 03, 2020).
- [50]facebook, "React." Meta, Jul. 17, 2023. Accessed: Jul. 17, 2023. [Online]. Available: <https://github.com/facebook/react>
- [51]facultyai, "dash-bootstrap-components." Faculty, Jul. 13, 2023. Accessed: Jul. 17, 2023. [Online]. Available: <https://github.com/facultyai/dash-bootstrap-components>
- [52]S. Vijay, "snehilvj/dash-mantine-components." Jul. 16, 2023. Accessed: Jul. 17, 2023. [Online]. Available: <https://github.com/snehilvj/dash-mantine-components>
- [53]mantedev, "Mantine." Mantine, Jul. 17, 2023. Accessed: Jul. 17, 2023. [Online]. Available: <https://github.com/mantedev/mantine>
- [54]A. M. Ward, "Dash Bootstrap Templates." Jul. 16, 2023. Accessed: Jul. 17, 2023. [Online]. Available: <https://github.com/AnnMarieW/dash-bootstrap-templates>
- [55]FastAPI, "FastAPI," 2023. <https://fastapi.tiangolo.com/> (accessed Jul. 10, 2023).
- [56]encode, "encode/uvicorn." Encode, Jul. 12, 2023. Accessed: Jul. 12, 2023. [Online]. Available: <https://github.com/encode/uvicorn>
- [57]wullt, "Mitwelten Explore." Mitwelten, Jun. 03, 2023. Accessed: Sep. 05, 2023. [Online]. Available: <https://github.com/mitwelten/mitwelten-explore>
- [58]Grafana, "Grafana: The open observability platform," *Grafana Labs*, 2020. <https://grafana.com/> (accessed Aug. 04, 2020).
- [59]"Wikipedia:Summary style," *Wikipedia*. Jul. 29, 2023. Accessed: Aug. 07, 2023. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Wikipedia:Summary\\_style&oldid=1167666430](https://en.wikipedia.org/w/index.php?title=Wikipedia:Summary_style&oldid=1167666430)
- [60]vogelwarte.ch, "Eisvogel | Schweizerische Vogelwarte," *vogelwarte.ch*. <https://www.vogelwarte.ch/de/voegel/voegel-der-schweiz/eisvogel> (accessed Aug. 09, 2023).
- [61]pandamodium, "Dash Quill." Jun. 01, 2023. Accessed: Jul. 11, 2023. [Online]. Available: <https://github.com/pandamodium/dash-quill>
- [62]quilljs, "Quill Rich Text Editor." Quill, Jul. 11, 2023. Accessed: Jul. 11, 2023. [Online]. Available: <https://github.com/quilljs/quill>
- [63]mozilla, "Bleach." Mozilla, Jul. 08, 2023. Accessed: Jul. 11, 2023. [Online]. Available: <https://github.com/mozilla/bleach>
- [64]P. Schmitz and S. Luber, "Was ist Cross-Site-Scripting (XSS)?," Mar. 29, 2018. <https://www.security-insider.de/was-ist-cross-site-scripting-xss-a-699660/> (accessed Jul. 11, 2023).
- [65]StatCounter, "Desktop vs Mobile vs Tablet Market Share Worldwide," *StatCounter Global Stats*, Aug. 15, 2023. <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/worldwide> (accessed Aug. 15, 2023).
- [66]StatCounter, "Search Engine Market Share Worldwide," *StatCounter Global Stats*, Aug. 15, 2023. <https://gs.statcounter.com/search-engine-market-share> (accessed Aug. 15, 2023).
- [67]M. Duò, "Submitting Your Website to Search Engines (When You Should Do It And How)," *Kinsta®*, Aug. 10, 2023. <https://kinsta.com/blog/submit-website-to-search-engines/> (accessed Aug. 15, 2023).
- [68]Bing (Microsoft Bing search), "Bing Chat response to the message 'what is mitwelten explore?,'" Aug. 15, 2023. [Online]. Available: <https://www.bing.com/search?q=biodiversity%20around%20basel>
- [69]Adobe Spectrum, "Spectrum, Adobe's design system," *Spectrum Design*, Apr. 06, 2022. <https://spectrum.adobe.com/page/color-for-data-visualization/> (accessed Aug. 14, 2023).
- [70]Plotly, "Hexbin," 2023. <https://plotly.com/python/hexbin-mapbox/> (accessed Jul. 10, 2023).
- [71]"H3: A Hexagonal Hierarchical Geospatial Indexing System." Uber Open Source, Jul. 08, 2023. Accessed: Jul. 10, 2023. [Online]. Available: <https://github.com/uber/h3>
- [72]Keycloak, "Keycloak - Open Source Identity and Access Management." <https://www.keycloak.org/> (accessed May 18, 2023).
- [73]Auth0, "Auth0," *Auth0 Docs*. <https://auth0.com/docs/> (accessed May 18, 2023).
- [74]wullt, "Mitwelten Explore Data Management." Mitwelten, Sep. 05, 2023. Accessed: Sep. 06, 2023. [Online]. Available: <https://github.com/mitwelten/mitwelten-explore-data-management>
- [75]"requests-cache/requests-cache." requests-cache, Aug. 28, 2023. Accessed: Aug. 28, 2023. [Online]. Available: <https://github.com/requests-cache/requests-cache>

## 7 Abbildungsverzeichnis

Abbildung 1 Z- und F-Pattern	7
Abbildung 2 Visualisierung von Zeitreihendaten	10
Abbildung 3 Scatter Map Plot	10
Abbildung 4 Visualisierung der Punktedichte mittels Heatmap	11
Abbildung 5 Hexbinmap: Punkte pro Zelle	12
Abbildung 6 Hexbinmap: Punkte pro Zelle, Zoom	12
Abbildung 7 Bubble Map	13
Abbildung 8 Visualisierung von Proportionen - Aktivste Vogelarten Deployment 679	13
Abbildung 9 Bereitstellung von Daten: REST API	16
Abbildung 10 Taxonomie Kategorien Bestäuber Datenset	20
Abbildung 11 MeteoSchweiz Stationen in den Gebieten der Mitwelten Feldstudien [30]	22
Abbildung 12 GBIF Resultate Klasse Aves Kt. Basel Stadt & Kt. Basel Land, 2020-2022 [32]	23
Abbildung 13 Kartenausschnitt Lebensraumkarte [38]	24
Abbildung 14 Layout Design Entwurf - Titel und Filter	27
Abbildung 15 Prominente Bereiche auf dem Bildschirm [42]	28
Abbildung 16 Dashboard Layout Entwurf	29
Abbildung 17 Dashboard Layout - FAB User Actions	29
Abbildung 18 Datenbankschema Mitwelten	30
Abbildung 19 Zeitliche Aggregation von PAX Messwerten	31
Abbildung 20 Pearson Korrelation	32
Abbildung 21 Darstellung eines Signals im Zeit- und Frequenzbereich [45]	33
Abbildung 22 Zeitreihe: Temperaturwerte über zwei Jahre	33
Abbildung 23 FFT von Temperaturwerten	34
Abbildung 24 Architektur Mitwelten Explore	38
Abbildung 25 Navigationselemente Explore Tool	40
Abbildung 26 Datenauswahl - Taxa	41
Abbildung 27 Datenauswahl - Taxa: Info Modal	42
Abbildung 28 Datenauswahl - Pollinators	43
Abbildung 29 Datenauswahl - Bird Diversity	43
Abbildung 30 Datenauswahl - Meteodaten	44
Abbildung 31 Datenauswahl - PAX	45
Abbildung 32 Datenauswahl - ENV Measurements	45
Abbildung 33 Notifications Collection Manipulation	46
Abbildung 34 Screenshot Collection	47
Abbildung 35 Eingabe des anzuzeigenden Zeitraums	47
Abbildung 36 Aktualisierung des Zeitbereichs bei Zoom	48
Abbildung 37 Auswahl Aggregation, Confidence und Zeitfenster	48
Abbildung 38 Dashboard Aktionen -Teilen einer Ansicht	49
Abbildung 39 Times Series Dashboard	50
Abbildung 40 Statistische Aggregate und FFT im Time Series Dashboard	51
Abbildung 41 Compare Dashboard	51
Abbildung 42 Compare Dashboard - Statistische Aggregate	52
Abbildung 43 Compare Dashboard - Korrelationsmatrix und FFT	52
Abbildung 44 Compare Dashboard - Konfiguration der Datensets	53
Abbildung 45 Taxon Dashboard	54
Abbildung 46 Taxon Dashboard - Erweiterte Informationen	55
Abbildung 47 Datenquellen Taxon Dashboard	55
Abbildung 48 Auflistung einzelner Datenpunkte	56
Abbildung 49 Map Dashboard	56
Abbildung 50 Map Dashboard - Vergleich von zwei Datensets	57
Abbildung 51 Map Dashboard - Bubble Map mit Lebensraumkarte	58
Abbildung 52 Map Dashboard - Eisvogel und Gewässernähe	58
Abbildung 53 Deployment Dashboard	59
Abbildung 54 Deployment Dashboard - Map	60

Abbildung 55 Deployment Dashboard - Environment	60
Abbildung 56 Deployment Dashboard - Erfasste Rohdaten	61
Abbildung 57 Screenshot Übersicht Annotationen	61
Abbildung 58 Screenshot Annotation	62
Abbildung 59 Screenshot Annotationseditor	62
Abbildung 60 Screenshot Bearbeiten einer Annotation	63
Abbildung 61 Screenshot Dokumentation im Tool	64
Abbildung 62 Landing Page Explore Tool	65
Abbildung 63 Landing Page Mobile Device	66
Abbildung 64 Link Preview in MS Teams	66
Abbildung 65 Zusammenfassung der Mitwelten Explore Applikation - Bing Chat [68]	67
Abbildung 66 Indikation des Ladevorgangs eines Line Charts	68
Abbildung 67 Datenvisualisierung - Kategorische Color Palette	68
Abbildung 68 Darstellungsoptionen Time Series Chart	69
Abbildung 69 Chart Layout: Overlay (l) und Subplots (r)	69
Abbildung 70 Datenvisualisierung - FFT Binning	70
Abbildung 71 Datenvisualisierung - Korrelationsmatrix	71
Abbildung 72 Datenvisualisierung - Radar Plot	72
Abbildung 73 Mitwelten Datenbank - Taxonomie Daten	73
Abbildung 74 Datenbankschema Mitwelten - Annotations und Collections	74
Abbildung 75 Datenbankschema Mitwelten Cache	75
Abbildung 76 GBIF Cache Update - Grid & Observations per Cell	76

## 8 Tabellenverzeichnis

Tabelle 1 Vorschlag Attribute zur Beschreibung von Umgebungsfaktoren	24
Tabelle 2 Callback Requests	35

## 9 Codeausschnittverzeichnis

Codeausschnitt 1 Callback Beispiel	35
Codeausschnitt 2 Clientside Callback Beispiel	36
Codeausschnitt 3 Ordnerstruktur Mitwelten Explore	39



## 10 Ehrlichkeitserklärung

Hiermit bestätige ich, dass die eingereichte Projektarbeit mit dem Titel «Visual Exploration for Urban Biodiversity Monitoring» das Resultat meiner persönlichen Erarbeitung der Themen ist. Ich habe keine anderen als die angegebenen Quellen benutzt.

Ort, Datum Windisch, 08.09.2023 Unterschrift \_\_\_\_\_

## 11 Anhang

Die im Rahmen dieser Projektarbeit entwickelte Software ist auf GitHub abgelegt.

### 11.1 Repositories

Source Code der Mitwelten Explore Applikation:

<https://github.com/mitwelten/mitwelten-explore>

Mitwelten Explore Data Management:

<https://github.com/mitwelten/mitwelten-explore-data-management>

Mitwelten DB Backend:

<https://github.com/mitwelten/mitwelten-db-backend>

### 11.2 Datenbezug Mitwelten Data API

API Dokumentation:

<https://data.mitwelten.org/api/v3/docs>

Python Beispiele:

<https://github.com/mitwelten/mitwelten-explore-data-management/tree/master/api-usage/python>

R Beispiele:

<https://github.com/mitwelten/mitwelten-explore-data-management/tree/master/api-usage/R>

JavaScript Notebooks :

<https://observablehq.com/@timeo-wullschleger/mitwelten-detected-birds>

<https://observablehq.com/@timeo-wullschleger/mitwelten-most-active-birds>

<https://observablehq.com/@timeo-wullschleger/mitwelten-birds-calendar>