

GoINSPIRED – Kontrolle und Transformation von Artendaten aus Citizen-Science-Projekten für INSPIRE

Christian Aden¹, Anna Charlotte Kirchner¹

¹Carl von Ossietzky Universität Oldenburg · christian.aden@uni-oldenburg.de

Zusammenfassung: Die Erfassung von Artendaten wird heute mithilfe von Apps und Web-Portalen von vielen Naturschutzverbänden durchgeführt. Dabei werden Methoden des Geospatial Web genutzt, um Daten zu erfassen, zu visualisieren und zu vernetzen. In den hoheitlichen Naturschutz fließen nur wenige dieser Daten ein. Hier bestehen Sorgen bezüglich der Datenqualität der von Citizen Scientists erfassten Daten. Zudem liegen die Daten in sehr heterogenen Datenmodellen vor. Mit diesem Aufsatz soll eine Möglichkeit aufgezeigt werden, wie Citizen-Science-Daten mit webbasierten Methoden kontrolliert und in das Datenmodell *Species Distribution* der INSPIRE-Richtlinie transformiert werden können.

Schlüsselwörter: INSPIRE, Geodateninfrastruktur, Citizen Science, Datenkontrolle

Abstract: *Collecting species occurrence data can be done using a variety of software applications and web portals implemented by NGOs. They use methods of the Geospatial Web for collecting, visualizing and implementing data in spatial data infrastructures. Most of the data are not in use by public authorities. There are concerns about the quality of data collected by Citizen Scientists. In addition, the data are produced in very heterogeneous data models and provided in different formats. With this paper we want to present options for a validation of Citizen Science data and how raw data can be transformed into the data model Species Distribution of the EU INSPIRE Directive.*

Keywords: *INSPIRE, geo data infrastructure, citizen science, data controlling*

1 Hintergrund und Motivation

Die Erhebung von Daten der Tier- und Pflanzenwelt erfolgt in Deutschland durch die Naturschutzbehörden im Rahmen der gesetzlichen Anforderungen. Die Notwendigkeit dafür zeigt der auf einen Tiefstwert gesunkene NBS-Indikator für die Artenvielfalt und Landschaftsqualität im Zeitraum 2001 bis 2011 (BMUB 2015, 10; BfN 2015). Nach §3 Abs. 1 des BNatSchG sind die Landesbehörden und das Bundesamt für Naturschutz (BfN) zuständig. Nach §6 Abs. 1 BNatSchG beinhaltet es die Beobachtung von *Natur und Landschaft*, deren Zustand fortlaufend zu ermitteln und zu bewerten ist. Damit wird auch die Einhaltung der europa- und völkerrechtlichen Verpflichtungen wie der Vogelschutz- und der FFH-Richtlinie gesichert. In den Ländern erfolgt das Monitoring für die in den Anhängen der Richtlinien enthaltenen, der in Deutschland vorkommenden Arten der Roten Liste, sowie Invasive Arten und wenige andere Arten. Dafür wurden landesweite Monitoring-Programme aufgestellt. Beauftragte Biologen erfassen vor allem in den Schutzgebieten, um bei der Erfüllung von Berichtspflichten zu helfen. Ein flächendeckender Überblick über den Zustand der Natur und die Erfassung von Veränderungen, ist dagegen weitestgehend nur durch das traditionelle, sich verändernde ehrenamtliche Engagement möglich (SCHMOLL 2012). Teilnehmen können im Prinzip alle BürgerInnen. Formulare mit entsprechenden Erhebungsparametern werden von Behörden

bereitgestellt. Von Behörden und ehrenamtlichen Experten unterstützte Software nannten bereits LIPSKI et al. (2010). Bei diesen Systemen handelt es sich um zu einem Teil kostenlose Software (z. B. Multibase CS, Recorder D).

Andere Wege gingen Naturschutzverbände. Sie reagierten stärker auf die Anforderungen junger Menschen und bieten Erfassungswerkzeuge in Form von Apps und Web-Portalen. HOPPE (2012) berichtet von 40 Web-Portalen, die sich in Deutschland den Themen Artenerfassung und Artenbestimmung widmen.

Das Erfassen von Daten zu vereinfachen und dadurch allen BürgernInnen zu ermöglichen oder Datenflüsse zu beschleunigen sind die Ziele solcher Projekte. Der Einsatz von Web-Mapping Anwendungen zur Visualisierung von Beobachtungen, sowie Social-Media-Komponenten für die Kommunikation innerhalb der „Crowd“ ist dabei eine Form des Anreizes, aber auch der Rückmeldung, die durch die Behörden nicht genutzt wird. Vorteile massenwirksamer Citizen-Science-Projekte sind die höhere Anzahl an potenziellen Meldungen und die größere Fläche die damit abgedeckt werden kann und praktisch nur durch sprachliche Hindernisse begrenzt wird. Im Gegensatz dazu stehen zeitlich und räumlich begrenzte wissenschaftliche Projekte (WERTS et al. 2012). Positiv sind zudem Kostenersparnisse und mögliche Frühwarnfunktionen für Veränderungen in Ökosystemen, insbesondere auf lokaler Ebene (JOHNSON & SIEBER 2013). Auffällig ist demgegenüber, dass nur wenige Angebote einen direkten Bezug zum hoheitlichen Monitoring aufweisen. Zwei Beispiele existieren seit einigen Jahren mit dem Web-Portal *ornitho.de* des Dachverbandes Deutscher Avifaunisten und der Artenfinder-App des Naturschutzverbandes POLLICHIA. Der DDA ist spezialisiert auf alle Arten der Vogelbeobachtung und kooperiert mit dem BfN (SUDFELDT et al. 2012). Mit der Artenfinder-App können unterschiedliche Tierarten von allen BürgerInnen erfasst und zusätzlich in einem Web-Portal bearbeitet werden. Das erfolgreiche Projekt wurde durch das Land Rheinland-Pfalz unterstützt (s. RÖLLER et al. 2014).

Aus der Vielzahl der Citizen-Science-Programme ergibt sich eine Diversität an Erfassungsmethoden und Datenmodellen, aber auch Aggregationsmethoden und Visualisierungstechniken. Wollen Behörden in Zukunft auf das Ehrenamt setzen, muss in Kooperationen erarbeitet werden, wie die heterogenen Geodaten validiert, korrigiert und in ein standardisiertes Format gebracht werden können (vgl. LIPSKI et al. 2010). Wie eine Harmonisierung von Geodaten und webbasierten Diensten auf internationaler Ebene erfolgen kann, haben die EU-Mitgliedsstaaten bei der Entwicklung der Datenmodelle für die Umsetzung der INSPIRE-Richtlinie (2007/2/EG) gezeigt. INSPIRE-Standards wurden weitestgehend auf Basis bestehender Standards des Open Geospatial Consortiums und Normen der ISO entworfen. Als Datenmodell für die Darstellung der Verteilung von Arten der Flora und Fauna liegt das Geodatenthema „Species Distribution“ (SD 2013) vor. Wie die in Citizen-Science-Projekten erhobenen Daten in das Anwendungsschema überführt werden können wurde bereits publiziert (s. WIEMANN & BERNARD 2014, DI STASO et al. 2015). Nicht behandelt wurden dabei die Attribute des Anwendungsschemas, die nur optional aufgeführt werden müssen, daher wird auch die Erweiterung des Datenschemas nicht behandelt.

Eine andere Herausforderung sind die verschiedenen Qualitäten von Monitoring-Programmen. Mit einem steigenden Anteil an Freiwilligen besteht stets die Sorge der verminderten Qualität und Glaubwürdigkeit gegenüber den generierten Daten (vgl. GILLET et al. 2012).

Auf beide Herausforderungen sollte bereits mit der Entwicklung des von der Deutschen Bundesstiftung Umwelt geförderten Systems ARDINI reagiert werden (ADEN et al. 2013). Das

Projekt sollte einen Weg der Modernisierung und Beschleunigung des Datenflusses von Naturschutzverbänden zur unteren Naturschutzbehörde in Niedersachsen durch Digitalisierung aufzeigen. Eingesetzt wurden dabei webbasierte Technologien, die Web-Mapping-Anwendungen, sowie eine Libellen- und eine Vogel-App mit artspezifischen Identifikationshilfen, Artensteckbriefen und über räumliche und zeitliche Informationen gesteuerte Plausibilitätskontrollen bereitstellen. Die Meldeparameter richten sich nach den analogen Meldebögen des Niedersächsischen Landesbetriebs für Wasserwirtschaft, Küsten- und Naturschutz. Sämtliche Arbeiten können einschließlich der Verwendung digitaler Karten auch offline durchgeführt werden. Mit dem Web-Portal können die Daten in Karten betrachtet und gefiltert werden. Durch die Exportfunktionen werden Formate wie GML, Shapefile bedient. Weiterhin ist leider keine Importschnittstelle für digitale und auf Plausibilität kontrollierte Artendaten in Niedersachsen vorhanden.

2 Ziele

Ausgehend von den Ergebnissen des Projekts ARDINI, war es das Ziel, den Datenfluss von der Feldarbeit bis in amtliche Geodatenbanken weiter zu beschleunigen. In Anbetracht der Tatsache, dass die mit ARDINI erfassten Daten aufgrund fehlender Infrastrukturen nicht in die Niedersächsischen Monitoring-Systeme integriert werden, bietet es sich an, diese direkt in das Datenmodell für das Geodaten Thema Species Distribution zu transformieren, praktisch im Bottom-up-Verfahren. Das vollständige Datenmodell des Anwendungsschemas, einschließlich der Erweiterung soll damit abgedeckt werden. Zudem sollen weitere, auch automatisierte Kontrollmethoden angeboten werden, damit eine möglichst hohe Qualität der Daten erreicht wird. Zielanwender des Systems sind in erster Instanz NGOs.

3 Methoden

Mit dem GeoCMS HotSpot besteht ein webbasiertes System, das in erster Linie der Herstellung und Administration von Geo Web Services und Web-GIS, aber auch der Einrichtung von Datenerfassungsportalen dient (ADEN 2012).

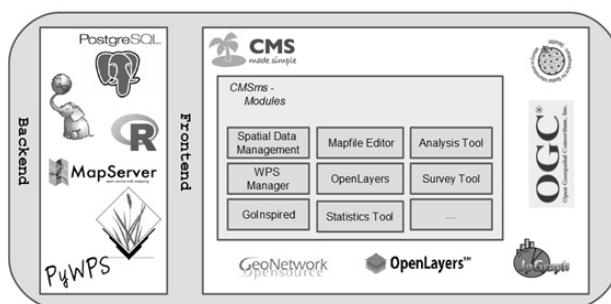


Abb. 1:
Software und Module des
GeoCMS HotSpot

Die Basis stellt das ein in PHP entwickeltes Content Management System das auf Modulen aufbaut Es wird auf einem Linux-OS mit dem DBMS PostgreSQL/PostGIS betrieben. Notwendige Softwareprodukte wie der UMN Mapserver, GRASS-GIS, OpenLayers und andere

wurden bereits in verschiedenen Modulen des GeoCMS eingesetzt. Um mit dem bestehenden System einen vollständigen Datenfluss von der Felddatenerfassung bis in das Datenmodell des Zielschemas zu ermöglichen, sollten bereits bestehende Module des GeoCMS verwendet werden. Sie wurden in vorherigen Arbeiten entwickelt. Insbesondere sind es die Module für das Geodaten-Management, Datenerfassungen und Analysemethoden. Ein Überblick über die im System enthaltenen Open-Source-Softwareprodukte und entwickelte Module wird in Abbildung 1 dargestellt.

Validierung von „user generated content“

Erst wenn Daten nachweislich geprüft wurden, können sie von Behörden für naturschutzfachliche Fragestellungen verwendet werden. In frei zugänglichen Citizen-Science-Projekten werden häufig die von ELWOOD et al. (2013) aufgeführten Methoden der Qualitätskontrolle eingesetzt. Am wichtigsten erscheint ihnen das Crowdsourcing, eine Methode bei der die Korrektheit der Daten über die Masse an Informationen, auf der anderen Seite durch Überarbeitung oder Bewertung durch die Community sichergestellt wird. Eine zweite Methode ist das Heranziehen von regionalen Experten, die ihnen zugewiesene Dateneingänge validieren. Als dritte Option sehen ELWOOD et al. (2013) GIS-Methoden, die via automatischer Kontrolle auf Beziehungen zwischen Objekten und Zuständen sowie zeitlichen Angaben aufsetzen. Neben diesen Methoden soll eine automatisierte Kontrolle auf Basis einer Habitatmodellierung umgesetzt werden, wie von KIRCHNER & ADEN (2016) beschrieben.

Schema-Mapping

Ein Schema-Mapping wird üblicherweise dann durchgeführt, wenn heterogene Daten zusammen verarbeitet werden sollen. Es handelt sich dabei um ein Vorgehen, bei dem die Beziehungen zwischen den Elementen und Daten eines Schemas eines Datenmodells mit dem Schema eines zweiten Datenmodells untersucht und formal spezifiziert werden (CALVANESE et al. 2013). Zu Beginn wird die Schema-Analyse durchgeführt, bei der wesentliche Informationen über Struktur und Dateiformate, verwendete Elemente und Dateninhalte der Ausgangs- und Zielschemata erfasst werden. Mit dem Mapping werden die Elemente der Schemata verglichen. Assoziationen innerhalb und zwischen den Schemata werden gesucht und die korrespondierenden Elemente einschließlich der Unterschiede in Syntax und Semantik herausgestellt. Auf Grundlage der Korrespondenzen werden schließlich Transformationsregeln aufgestellt, die mit entsprechenden Funktionen Anwendung finden.

4 Ergebnisse

Bevor die eigentliche Entwicklung des Moduls *GoInspired* für die Transformation der Citizen-Science-Daten umgesetzt werden konnte, wurde das Schema-Mapping zwischen dem ARDINI- und dem Species Distribution Datenschema durchgeführt. Welche Validierungsmethoden umgesetzt und wie die Daten auf Basis des Schema-Mappings und daraus abgeleiteter Transformationsfunktionen konform gegenüber dem Anwendungsschema Species Distribution bereitgestellt werden können, wird im Folgenden beschrieben.

Datenschema *Species Distribution* und Informationsquellen in ARDINI

Daten zum Vorkommen von Arten werden mit dem INSPIRE-Datenmodell *Species Distribution* in *Feature Collections* in Form von Polygonen vorgehalten (SD 2013). Bei dem Objekttyp, genannt *SpeciesDistributionDataSet*, handelt es sich um alle Geometrien (*Collection*) der Verbreitungseinheiten bzw. *SpeciesDistributionUnits*. Die Identifikation des *DataSets*, die Beschreibung der räumlichen Ausdehnung und die zeitliche Gültigkeit erfolgt über die darin enthaltenen Attribute (*inspireId*, *name*, *domainExtent*, *beginLifespanVersion*, *endLifespanVersion*). Die *Collection* besitzt zwei Assoziative Rollen. Mit der Rolle „*documentBasis*“ wird ein Verweis zu Dokumenten bereitgestellt, mit denen die Herstellung der Daten erläutert wird (Kampagne, rechtliche Grundlage). Die Assoziative Rolle „*member*“ verweist auf die in der *Collection* enthaltene Objektklasse *SpeciesDistributionUnits*. Diese enthält Attribute die deren Eigenschaften wie die Geometrie (*geometry*), das identifizierende Merkmal (*inspireId*), den Namen der Art (*speciesName*), die zeitliche Gültigkeit der Daten (*beginLifespanVersion*, *endLifespanVersion*), sowie das Vorkommen der Spezies (*distributionInfo*) beschreiben. Nicht zwingend besteht die Assoziation zu einem räumlichen Objekt, das Informationen über die Ausdehnung der Fläche des Vorkommens der Spezies in der jeweiligen *SpeciesDistributionUnit* enthält (*AbstractFeature*) und nach ISO 19136 in Form von GML angegeben werden müsste.

Mit dem komplexen Attribut *SpeciesNameType* wird der Name der Spezies durch einen Bezeichner (*referenceSpeciesId*) und die Angabe einer Referenz zu einem taxonomischen Schlüssel (*referenceSpeciesScheme*) eindeutig beschrieben. Hier werden die EU-Nomen als präferierte Referenz für die Namen der Spezies (Taxon) verwendet. Zusätzlich können mit der lokalen Kennung (*localSpeciesId*) Verweise zu einem lokal verwendeten Namen (*localSpeciesName*) sowie der eingesetzten Taxonomie (*localSpeciesScheme*) angegeben werden. Die Verknüpfung zwischen der lokal und einer der präferierten Taxonomien erfolgt über ein eigenes Attribut (*qualifier*), die Angabe des darin enthaltenen Namens der Spezies ebenfalls (*referenceSpeciesName*).

Die Beschreibung des Vorkommens einer Spezies über das entsprechende Attribut (*distributionInfo*) kann über den Datentyp *DistributionInfoType* tiefer gehend erfolgen. Hier sind Angaben zur Populationsdichte in Klassen (*occurrenceCategory*), den Erfassungszeitraum (*collectedFrom*, *collectedTo*), die Größe der Population (*populationSize*), die Art des Vorkommens der Spezies (*populationType*), den Status der Ansässigkeit (*residencyStatus*) und das Vorliegen einer Gefährdungsstufe (*sensitiveInfo*) möglich. Die Größe der Population (*RangeType*) kann um Informationen zur Zählmethode (*countingMethod*), ergänzt werden. Welche Schwerpunkte bei der Zählung vorlagen wird durch Codelisten (*Article17CountingUnitValue*, *GeneralCountingUnitValue*) angegeben (*countingUnitValue*). Die Individuenzahl der beobachteten Art in einer Verteilungseinheit kann eingrenzend genannt werden (*lowerBound*, *upperBound*).

Mit der Erweiterung des Anwendungsschemas kann schließlich über einen Verweis auf die zugrunde liegenden Daten aus den Beobachtungen im Feld und auf weitere Metadaten zu diesen zugegriffen werden (SD 2013, 45 ff.). Die Erweiterung erfolgt über die Subklasse *SpeciesDistributionUnit* die von der gleichnamigen Klasse *SpeciesDistributionUnit* erbt. Sie enthält unter anderem das Attribut *occurrenceRecordIdentifier* das eine Kennung für Daten aus Beobachtungen als Wert enthalten kann (z. B. Darwin Core Triple mit *InstitutionCode*,

CollectionCode, *CatalogNumber*). Über die assoziative Rolle *metadata* besteht eine Verbindung zum räumlichen Objekttyp *SourceInformation*, der Metadaten zur Quelle der Daten (*source*) enthält. Die Quelle der Daten wird weiter aufgeschlüsselt durch Attribute zu der Daten haltenden Einrichtung (*institutionName*, *institutionAddress*) und Datenbank (*source-Database*), mit Angaben zu den Methoden der Erfassung (*sourceMethod*) des Datentyps *sourceMethodType*.

Mit dem Schema-Mapping wurde schnell deutlich, dass nur wenige Attribute aus ARDINI-Datenerfassungen eine direkte Beziehung oder identische Eigenschaften aufweisen. Tabelle 1 zeigt eine Liste der Attribute und Objekte für die Quellen gefunden wurden. Korrespondenzen zeigten ausschließlich solche Daten auf, die entweder Koordinaten des Fundes, zeitliche Angaben (Datum) oder Informationen zur Anzahl und dem Namen der Art enthalten. Betrachtet man neben den Dateneingängen (DB-Beobachtungen) die Informationen, die speziell in der LibellenApp vorhanden sind, so sind auch Informationen zum Schutzstatus (z. B. Rote Liste) sowie der Häufigkeit des Vorkommens (z. B. häufig, selten) vorhanden. Hieraus konnten weitere Informationen abgeleitet werden. Als mögliche weitere Quelle wurden Metadaten angesehen. Die zweifache Angabe von Informationen zu einer NGO könnte damit vermieden werden, hier handelt es sich um Angaben zu Adressen und zur Laufzeit und Gültigkeit der Daten. Formulare werden für die Attribute herangezogen, die in Form von Text eingegeben werden müssen. Die mit dem Anwendungsschema vorgegebenen Codelisten wurden in einer Datenbank zusammengeführt und stehen als Listenauswahl zur Verfügung. In zwei Fällen können Web-Dienste herangezogen werden.

Transformationen und Datenquellen

Tabelle 1: Attribute und Objekte des Zielschemas und mögliche Quellen

| Attribut | Quelle | Objektklasse / Attribut |
|---|--------------------------------|---|
| beginLifeSpanVersion | Metadaten DB-Beobachtungen | LI_ProcessStep / date (für SpeciesDistributionDataSet) (ggf. manuelle Eingabe) datum_start (für SepeciesDistributionUnit) |
| endLifeSpanVersion | Metadaten DB-Beobachtungen | EX_Extent / temporalExtent (für SpeciesDistributionDataSet) (ggf. manuelle Eingabe) datum_ende (für SpeciesDistributionUnit) |
| institutionName | Metadaten | CI_ResponsibleParty / organisationName (ggf. manuelle Eingabe) |
| adminUnit | Metadaten | CI_Address / administrativeArea (manuelle Eingabe) |
| addressArea | Metadaten | CI_Address / city (ggf. manuelle Eingabe) |
| postCode | Metadaten | CI_Address / postalCode (ggf. manuelle Eingabe) |
| thoroughfare | Metadaten | CI_Address / thoroughfare (ggf. manuelle Eingabe) |
| localSpeciesName | DB-Beobachtungen | name_deutsch, name_wissenschaft |
| localSpeciesId | DB-Beobachtungen | artnummer |
| CountingMethod, (Codeliste: CountingMethodValue) | DB-Beobachtungen | zaehllart (Vögel) |
| SensitiveInfo | DB-Artensteckbriefe | rl_d, rl_nds, fh (Libellen) |
| OccurrenceCategory (Codeliste: Occurrence-Category-Value) | DB-Artensteckbriefe (Libellen) | pl_bestand, pl_bestand_boerde, pl_bestand_tiefland_west, pl_bestand_tiefland_west, pl_bestand_nds, |
| CountingMethod, (Codeliste: CountingMethodValue) | DB-Beobachtungen | zaehllart (Libellen), wenn DB-Beobachtungen erweitert wird |

Tabelle 1 (Fortsetzung)

| Attribut | Quelle | Objektklasse / Attribut |
|--|-------------------------|---|
| collectedFrom | DB-Beobachtungen | datum_start (erstes in jeweiliger Verteilungseinheit) |
| collectedTo | DB-Beobachtungen | datum_ende (letztes in jeweiliger Verteilungseinheit) |
| PopulationSize (lowerBound, upperBound) | DB-Beobachtungen | anzahl_ges (Vögel), Summe der Individuen oder maximale Summe der nach Klassen erfolgten Eingabe (Libelle) |
| domainExtent | DB-Beobachtungen | longitude, latitude (Transformation) |
| geometry | DB-Beobachtungen | longitude, latitude (Transformation) |
| localSpeciesScheme | Listenauswahl | Informationen zu Arten der Steckbriefe liegen vor (u. a. Erstautoren) |
| ReferenceSpeciesScheme (Codelist: ReferenceSpecies-SchemeValue) | Listenauswahl | EU-Nomen, EUNIS oder NatureDirectives Codelisten |
| referenceSpeciesName | Webservice, Text | Web-Portale (EU-Nomen, EUNIS) oder Natura2000 |
| referenceSpeciesId | Webservice, Text, Liste | Web-Portale (EU-Nomen, EUNIS) oder Datenbank (Natura2000), auch Erweiterung der lokalen Datenbank um Referenztabellen |
| Qualifier | Listenauswahl | Codeliste: QualifierValue |
| CountingUnit (CountingUnitValue) | Listenauswahl | Codeliste: Article17CountingUnitValue oder GeneralCountingUnitValue |
| sourceDatabase | Text | ... oder Abfrage lokaler Datenbanken und Listenauswahl |
| methodValue (SourceMethodValue) | Listenauswahl | Codeliste: Article17SourceMethodValue oder GeneralCountingMethodValue, auch automatisch (randomObservation) |
| PopulationType | Listenauswahl | Codeliste (PopulationTypeValue) |
| residencyStatus | Listenauswahl | Codeliste (ResidencyStatusValue) |

Validierung der Daten

Im System eingegangene Daten können mithilfe von fünf unterschiedlichen Methoden validiert werden. Zwei Methoden sind bereits in die Apps integriert worden. Die Habitat- und die Zeitraum-Kontrolle (s. o.) können im GeoCMS durch zusätzliche Geodaten und zeitliche Informationen ergänzt werden. Alle Methoden (s. Tabelle 2) wurden mithilfe von PHP, PostGIS-Funktionen und WPS umgesetzt.

Mit der SDM-Kontrollmethode *HotSpotSDM* wird ein WPS auf Basis von pyWPS durchgeführt. Eine Erläuterung der Methode findet sich in KIRCHNER & ADEN (2016) in diesem AGIT-Journal.

Tabelle 2: Validierungsmethoden im Modul *GoInspired*

| Methode | Umsetzung |
|----------|--|
| Experten | Kennzeichnung: plausibel/nicht plausibel Eingabe zusätzlicher INSPIRE-Attribute (z. B. Populationstyp) per Formular |
| Crowd | Mindestanzahl an positiven Bewertungen, Eingabe per Formular |
| Zeitraum | Abgleich des Zeitpunktes der Erfassung mit dem saisonalen Auftreten der Art |
| Habitat | Ort der Erfassung, PostGIS |
| SDM | Web Processing Service zur Durchführung von Habitatmodellierungen |

Erhebung anlegen, Validierung und Transformation definieren

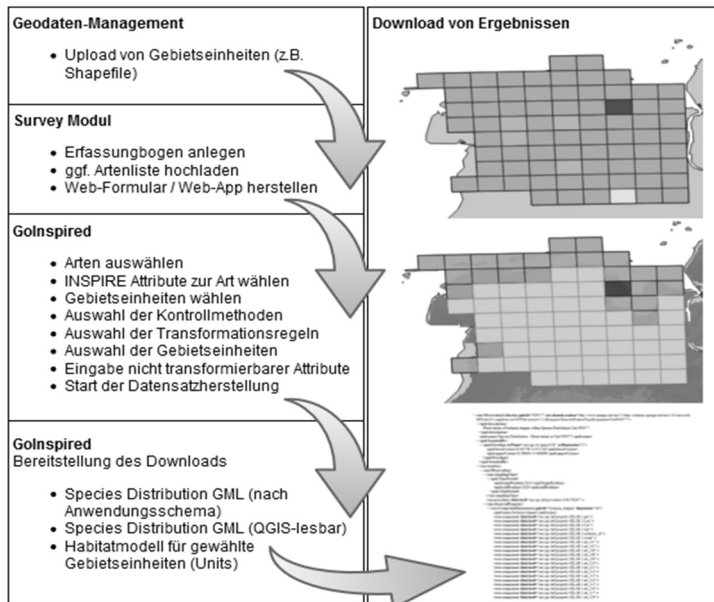


Abb. 2:
Workflow des
Moduls *GoInspired*

Für die Erhebung von Artendaten muss zunächst ein Erfassungsbogen mit dem Survey-Modul angelegt werden (s. Abbildung 2), der um eine Artenliste ergänzt werden kann (Name der Art, wiss. Name, Schutzstatus, lokale ID, Referenz ID u. a.). Dadurch werden Attribute des Zielschemas bedient. Das Formular wird automatisiert in Kombination mit einer Web-Mapping-Komponente angelegt. Ebenso kann eine Web-App hergestellt werden. Adressen für den Zugang werden direkt ausgegeben und können an die Teilnehmer der Erhebung versendet werden. Speziell für die ARDINI-LibellenApp wurde eine Schnittstelle in das Survey-Modul integriert, die dafür sorgt, dass Daten in die dafür im System angelegte Erhebungsdatenbank integriert werden. Insgesamt stehen damit drei Erhebungskomponenten zur Verfügung. Auf Basis von OSM-Daten kann nun im Feld mit der App oder an einem lokalen Rechner der Fund in der Karte markiert und das entsprechende Formular ausgefüllt und versendet werden. Liegen schließlich Daten zu einer Erhebung vor, kann der Herstellungsprozess INSPIRE-konformer Datensätze beginnen.

Im Modul *GoInspired* wird dafür zunächst die Art gewählt, deren Verteilung dargestellt werden soll. Zu den Angaben aus der Artenliste können nun weitere Informationen aus vorgegeben Wertelisten gewählt werden (Beziehung nationaler und internationaler Referenzen zueinander, zeitliche Gültigkeit u. a.). Anschließend wird der Geodatensatz ausgewählt, der die Verteilungseinheiten darstellt. Dieses muss ein Polygondatensatz sein, der im Vorfeld in das GeoCMS geladen wurde. Schließlich werden die Kontrollmethoden per Checkbox ausgewählt, die bei der Datenauswahl beachtet werden sollen.

Im Folgenden werden die offenen Transformationsregeln gewählt (Zählmethode und Einheiten, Vorkommenskategorie u. a.). Die Attribute die mit dem Erhebungsbogen erfasst werden können, werden den Attributen des Zielschemas zur Auswahl gegenübergestellt. Vorliegende Werte werden bei der späteren Transformation herangezogen. Die Anzahl der Funde kann

automatisch angegeben werden (maximale/minimale Anzahl Funde an einem Tag im Erfassungszeitraum). Nicht bediente Attribute müssen, sofern verpflichtend, manuell eingegeben werden und sind je nach Untersuchungsgebiet für alle Einheiten verwendbar (Populationstyp, Vorkommenskategorie, Zählmethode).

Nach der Durchführung der Transformationsfunktionen stehen den Nutzern zwei GML-Dateien mit allen nach SD (2013) geforderten Attributen zur Verfügung. Das Ergebnis des SDM-Prozesses, wird in Form eines Shapefiles angeboten. Die Attributwerte der Erfassungen sind über Verweise innerhalb der GML-Dateien über einen Sensor Observation Service zugänglich.

5 Diskussion und Ausblick

Es wurde am Beispiel der ARDINI Erhebungskomponenten gezeigt, dass Daten aus Citizen-Science-Projekten direkt in das Datenmodell *Species Distribution* transformiert werden können. Durch die Verwendung des Datenschemas einer Erfassung, das nicht direkt mit dem Zielschema vergleichbar ist, können viele Attribute nicht automatisiert aus bestehenden Quellen herangezogen werden. Sie müssen manuell eingegeben oder in Auswahllisten angeboten werden. Da dieses jedoch webbasiert durchgeführt werden kann, ist eine effektive Bearbeitung möglich. Da in erster Linie ein Bottom-up-Verfahren verfolgt wird und das GeoCMS HotSpot für den Einsatz auf regionaler Ebene vorgesehen ist, könnten lokal agierende NGOs ihre Daten mit diesem Vorgehen und dem Zielschema dauerhaft vergleichbar anbieten. Tests werden ab dem Sommer 2016 in Zusammenarbeit mit dem NABU Oldenburg durchgeführt. Datenkontrollen können die Auswahl der für den Prozess geeigneten Daten erleichtern und die Qualität sichern. Die Habitatmodellierung als Validierungsmethode wird weiterhin untersucht, es sind teilweise artspezifische Anforderungen zu erfüllen.

Literatur

- ADEN, C. (2012), Current approaches for networks between NGOs, citizen scientists and public authorities using mobile devices for environmental data collection and online portals for data processing. In: JOBST, M. (Ed.), Service Oriented Mapping 2012. Jobst Media Verlag, Wien, 335-352.
- ADEN, C., KASTNER, F., LOESBROCK, J. & KROHN-GRIMBERGHE, S. (2013), Neue Ansätze digitaler Artenerfassung für den ehrenamtlichen Naturschutz – Ergebnisse der Entwicklung mobiler Lösungen in Niedersachsen. In: Naturschutz und Landschaftsplanung – Zeitschrift für angewandte Ökologie, 45 (3), 101-107.
- BFN – BUNDESAMT FÜR NATURSCHUTZ (2015), Artenschutzreport 2015.
http://www.bfn.de/fileadmin/BfN/presse/2015/Dokumente/Artenschutzreport_Download.pdf (21.12.2015).
- BMUB – Bundesministerium für Umwelt, Naturschutz Bau und Reaktorsicherheit (2015), Naturschutz-Offensive 2020.
http://www.bmub.bund.de/fileadmin/Daten_BMU/Pool/Broschueren/naturschutz-offensive_2020_broschuere_bf.pdf (21.12.2015).

- CALVANESE, D., DE GIACOMO, G., LENZERINI, M. & VARDI, M. Y. (2013), On simplification of schema mappings. *Journal of Computer and System Sciences*, 79, 816-834.
- DI STASO, U., MATTASSONI, M. & DE AMICIS, R. (2015), INSPIRE and VGI: Species Distribution in Europe. In: GARTNER, G. & HUANG, H. (Eds.), *Proceedings of the 1st ICA European Symposium on Cartography, EuroCarto 2015*. Wien, 160-170.
- ELWOOD, S., GOODCHILD, M. F. & SUI, D. (2013), *Prospects for VGI Research and the Emerging Fourth Paradigm. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, Springer.
- GILLET, D. J., PONDELLA, D. J. II, FREIWALD, J., SCHIFF, K. C., CASELLE, J. E., SHUMAN, C. & WEISBERG, S. B. (2012), Comparing volunteer and professionally collected monitoring data from the rocky subtidal reefs of Southern California, USA. *Environ Monit Assess*, 184, 3239-3257.
- HOPPE, A. (2012), Neue Lösungen zur Datenerfassung im ehrenamtlichen Naturschutz: Ersatz, Transformation oder Ergänzung „alter Tugenden“? In: FROHN, H.-W., ROSEBROCK, J. (Bearb.), *Ehrenamtliche Kartierungen für den Naturschutz – Historische Analysen, aktuelle Situation und Zukunftspotenziale*. *Naturschutz und Biologische Vielfalt*, 123, 243-271.
- JOHNSON, P. A. & SIEBER, R. E. (2013), Situating the Adoption of VGI by Government. In: SUI et al. (Hrsg.), *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*.
- KIRCHNER, A. C. & ADEN, C. (2016), HotSpotSDM – ein WPS-Prozess für die Modellierung ökologischer Artenfunde. *AGIT – Journal für Angewandte Geoinformatik*, 2-2016, 592-601.
- LIPSKI, A., RÜTER, S., RUSCHKOWSKI, E. v. & HACHMANN, R. (2010), Digitale Artenerfassung im ehrenamtlichen Naturschutz. *Naturschutz und Landschaftsplanung*, 42 (8), 235-242.
- RÖLLER, O., SCHOTTHÖFER, A. & SCHRÖTER, S. (2014), ArtenAnalyse – ein WebGIS zur räumlichen und zeitlichen Auswertung von Tier- und Pflanzendaten in Rheinland-Pfalz. *Natur und Landschaft – Zeitschrift für Naturschutz und Landschaftspflege*, 7/2014, 321-324.
- SCHMOLL, F. (2012), Das Ehrenamt. Vergangenheit und Zukunft einer tragenden Säule des Naturschutzes. In: FROHN, H.-W. & ROSEBROCK, J. (Hrsg.), *Ehrenamtliche Kartierungen für den Naturschutz*. *Naturschutz und Biologische Vielfalt*, 123. BfN, Bonn-Bad Godesberg, 17-35.
- SD – Species Distribution (2013), *INSPIRE Data Specification on Species Distribution – Technical Guidelines*. European Commission Joint Research Centre. 126 S.
- SUDFELDT, C., BAIRLEIN, F., DRÖSCHMEISTER, R., KÖNIG, C., LANGGEMACH, T. & WAHL, J. (2012), *Vögel in Deutschland – 2012*. DDA, BfN, LAG VSW, Münster, 56 S.
- WERTS, J. D., MIKHAILOVA, E. A., POST, C. J. & SHARP, J. L. (2012), An Integrated WebGIS Framework for Volunteered Geographic Information and Social Media in Soil and Water Conservation. *Environmental Management*, 49, 816-832.
- WIEMANN, S. & BERNARND, L. (2014), Linking crowdsourced observations with INSPIRE. In: HUERTA, SCHADE & GRANELL (Eds.), *Connecting a Digital Europe through Location and Place*. *Proceedings of the AGILE'2014 International Conference on Geographic Information Science*.