



# Datenqualität bei



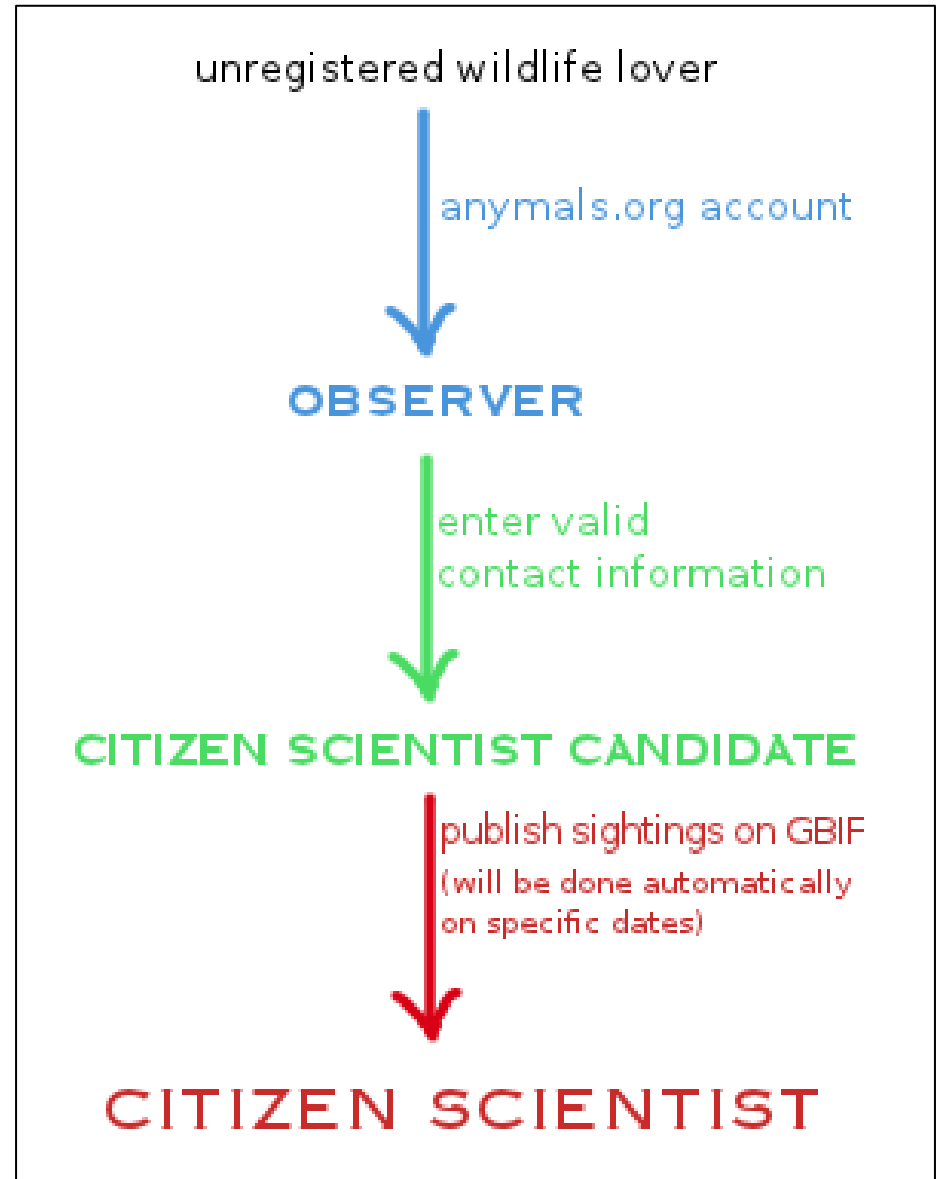
präsentiert von Falko Glöckler

# Datenqualität bei



**anymals**<sup>+plants</sup>

- User-Status
- Nur Daten angemeldeter Nutzer werden zur GBIF-Datenanbindung herangezogen



# Datenqualität bei anymals<sup>+plants</sup>



Belegfotos



Anzweifeln von Sichtungen



Forum für intensiven Nutzer-Austausch



Korrektur von Sichtungen möglich

Falsche Sichtungen löschar

dauerhafte, vorsätzliche Falschmelder sperrbar durch  
Anwendungs-ID, selbst wenn sie anonym sind

# Datenqualität bei



animals<sup>+</sup>plants

~ 9000 Sichtungen

> 1500 Arten

~ 500 Zweifel




> 160 Foren-Beiträge




Stand: Okt. 2012




## Ideen zur Auswertung:




- Zweifel je Nutzer-Status
- Schwellenwert von Zweifeln je Sichtung
- Einbeziehung der Zweifel in den Eingabe-Prozess; als Lernen aus den Zweifeln anderer Sichtungen (z.B. erneute Nachfrage, ob die bereits angezweifelte Art aufgenommen werden soll)




Aktuelles




  **Tom** zweifelt an der Sichtung **Feldahorn** (*Acer campestre*)  vor 1 Tag, 19 Stunden




  **Tom** zweifelt an der Sichtung **Orca** (*Orcinus orca*)  vor 2 Tage, 23 Stunden




  **Andreas** zweifelt an der Sichtung **Acker-Hahnenfuß** (*Ranunculus arvensis*)  vor 3 Tage, 3 Stunden

  **Andreas** mag die Sichtung **Graureiher** (*Ardea cinerea*)  vor 2 Wochen

  **Tom** mag die Sichtung **Bacchusreiher** (*Ardeola bacchus*)  vor 2 Wochen

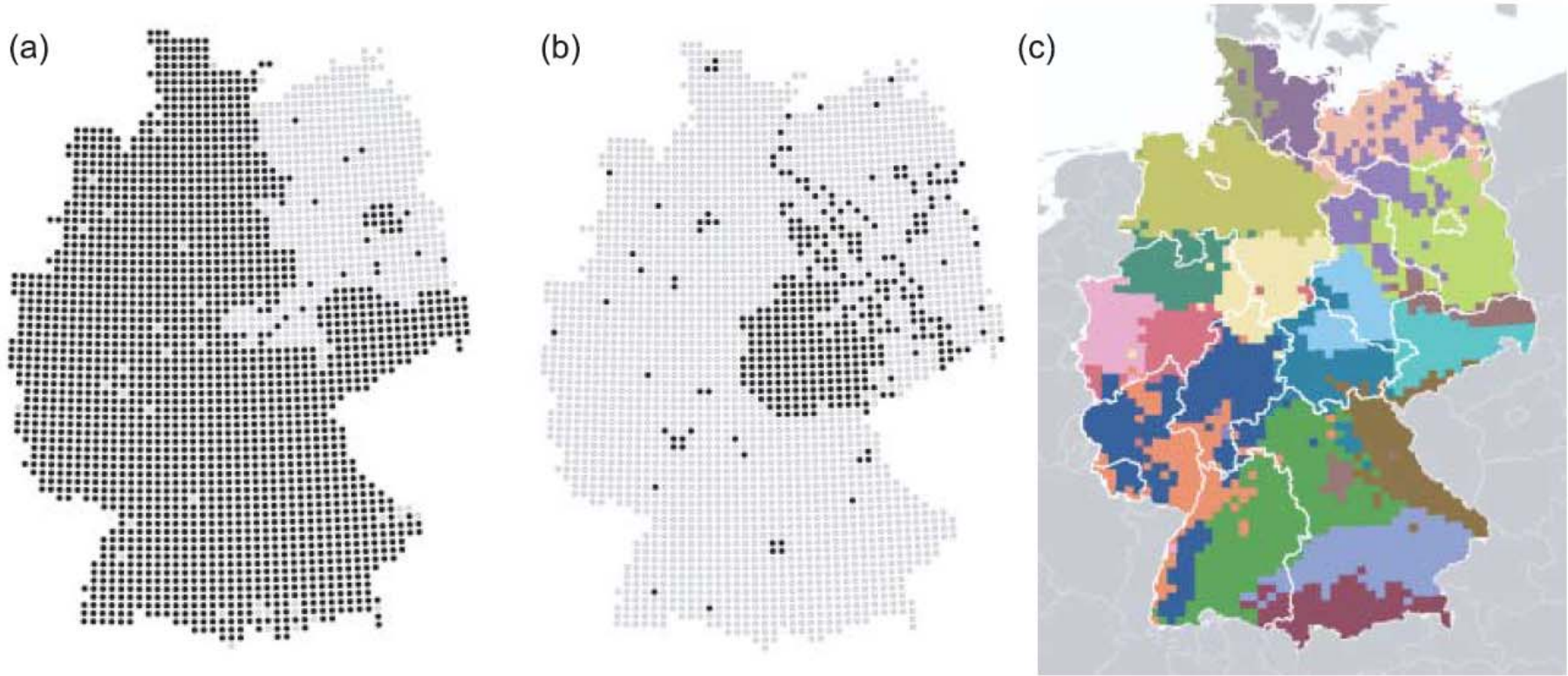
  **Andreas** mag die Sichtung **Bacchusreiher** (*Ardeola bacchus*)  vor 2 Wochen

  **Sabrina Madison Becker** mag die Sichtung **Kornweihe** (*Circus cyaneus*)  vor 2 Wochen, 2 Tage

  **Tom** zweifelt an der Sichtung **Little Sparrowhawk** (*Accipiter minullus*)  vor 2 Wochen, 3 Tage

# Beispiel für künstliche Muster in Verbreitungskarten aufgrund von unterschiedlichen taxonomischen Konzepten

präsentiert von Falko Glöckler



**Figure 6** The figure shows the occurrences of *Tripleurospermum perforatum* (a) and of *Tripleurospermum maritimum* (b) in the FLORKART data base across Germany. The distribution patterns reflect political boundaries, which are at the same time the boundaries of organizational units of the mapping project. In western Germany, *T. perforatum* was interpreted as a species in its own right; in eastern Germany it was apparently considered a subspecies of *T. maritimum*. Clustering eight Isomap components of FLORKART resulted in the pattern shown to the right (c), where the grid cells are colour coded according to their cluster affiliation. The map is overlaid with the state boundaries of Germany (white lines). This figure shows that in many cases both the clusters and the state boundaries coincide.

Entnommen aus:

Mahecha & Schmidlein 2008, Revealing biogeographical patterns by nonlinear ordinations and derived anisotropic spatial filters, *Global Ecology and Biogeography*, **17**, 284–296

# Plausibilitätsprüfung bereits an GBIF angebundener Daten

vorge stellt am Beispiel

## Biodiversity Datasets Assessment Tool



Universidad  
de Navarra

Dept. Zoology & Ecology

präsentiert von Falko Glöckler

# Biodiversity Datasets Assessment Tool



Universidad  
de Navarra

Dept. Zoology & Ecology



- Visualisierungs-Tool für Daten in GBIF
- Daten(bank)-Manager können Bias u. Fehler aus Nutzersicht sehen
- deskriptive Statistik in Qualitätskontrollen und „Fitness-for-Use“ einbeziehen

<http://www.unav.es/unzyec/mzna/biddsat/>

# Weitere nützliche Tools u. Services zur Prüfung u. Analyse von GBIF Daten



- OpenUp! Quality Toolkit  
<http://services.bgbm.org/DataQualityToolkit/>  
(Datenquelle kann vor GBIF-Anbindung getestet werden)
- Google Refine  
<http://code.google.com/p/google-refine/>  
Tabellenkalkulation, Datentransformation, Stapelverarbeitung, basiert auf wiederverwendbare Regeln
- GBIF Name Parser  
<http://tools.gbif.org/#nameparser>  
Webservice zum Prüfen der Verfügbarkeit von Taxonnamen im taxonomischen Backbone von GBIF
- Analyse-Routinen selbst bauen: z.B. GBIF via Statistik-Software „R“ direkt auswerten  
[http://spatial-analyst.net/?title=Retrieval and analysis of GBIF records in R](http://spatial-analyst.net/?title=Retrieval+and+analysis+of+GBIF+records+in+R)

# Weitere nützliche Tools u. Services zur Prüfung u. Analyse von GBIF Daten



- GBIF Checklist Bank  
<http://ecat-dev.gbif.org/api/clb>
- Viele weitere GBIF Tools und Webservices  
<http://tools.gbif.org/>
- Freies, sehr modular aufgebautes Statistik-Tool „R“  
<http://www.r-project.org>