



Open Data in der Forschung: Wer? Wie? Was?

Gabriele Wollnik-Korn, Open Access Week, 21. Oktober 2009, Köln



Ablauf

- Was ist Open Data in der Forschung, was sind die Vorteile und wie könnte es funktionieren?
- Wer befürwortet Open Data in der Forschung?
- Open Data – was macht die ZB MED?



Was ist Open Data

Open Data (OD)

„Open Data is an emerging term in the process of defining how scientific data maybe published and re-used without price or permission barriers“.

Forschungsdaten



Im Bereich Ernährung, Umwelt, Agrar:
Selbsterhobenes Datenmaterial, welches
immer einzig für die geplante Untersuchung
erhoben wurde

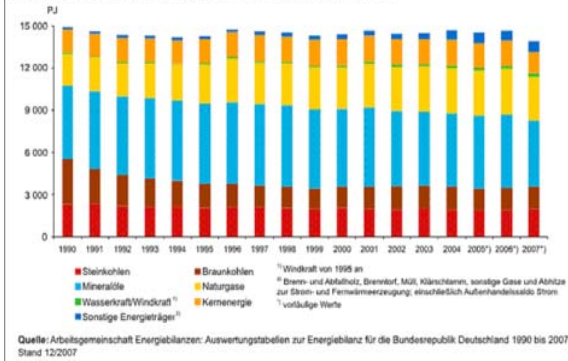
Forschungsdaten – einige Beispiele

- Zuchtwertschätzdaten
- Bodenprobedaten
- Ergebnisse der Klimaforschung
- Labordaten

... und so können Forschungsdaten aussehen

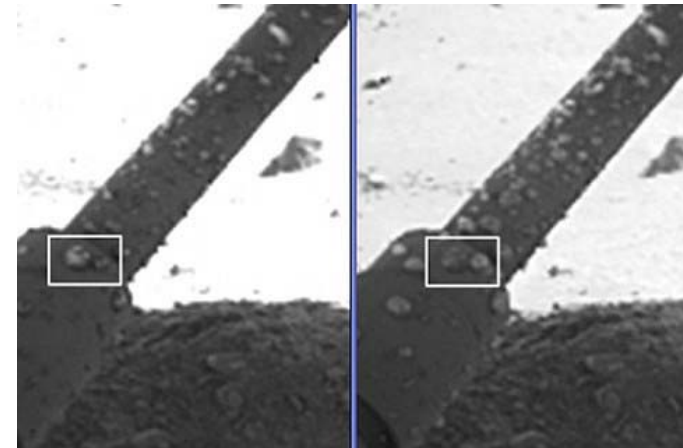


Entwicklung des Primärenergieverbrauchs in Deutschland nach Energieträgern



abelle 1: Datenumfang der Zuchtwertschätzung April 2008 bei den Rassen Fleckvieh, Gelbvieh und Braunvieh

	Baden-Württemberg	Bayern	Hessen	Österreich	Italien	Gesamt	Tiere im Pedigree
Fleckvieh							
Kühe mit linearen Beschreibungen	79.646	274.246	7.359	71.563	57.516	490.330	
Anzahl Bullen	1.274	4.876	380	2.760	2.339	11.629	1.172.188
Gelbvieh							
Kühe mit linearen Beschreibungen		4.806				4.806	
Anzahl Bullen		126				126	12.490
Braunvieh							
Kühe mit linearen Beschreibungen	29.408	45.274		53.550		128.232	
Anzahl Bullen	645	1.048		2.124		3.817	294.335



Bayrische Landesanstalt für Landwirtschaft Datum der Schätzung: 13.08.2007
Institut für Tiermichi

ZUCHTWERTSCHÄTZUNG EXTERIEUR

Name: Fleckviehbulle Ohrmarke: 2760009000000000 Geburtsdatum: 23.01.2002

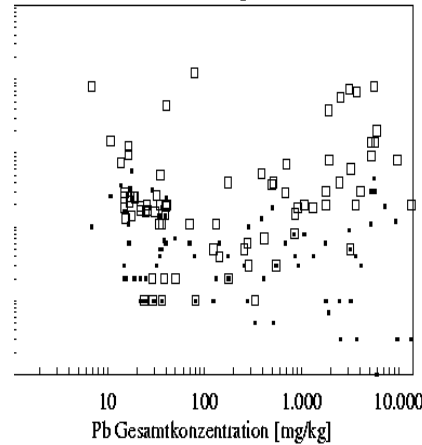
Beurteilte Tiere: 79 Sicherheit Zuchtwert Euter: 85%

Relativdruckwerte der einzelnen Merkmale

Merkmale	64	76	88	100	112	124	136
Kahnen	110						
Bornschulung	104						
Fuchswenk	109						
Euter	132						
Kreuzhöhe	110 klein						groß
Rechenlänge	106 kurz						lang
Rüchbreite	103 schmal						breit
Rumpftiefe	108 seicht						tief
Rechenneigung	90 oben						abfallend
Spitzwinkelung	93 steil						abflächig
Spitzausprägung	95 voll						trocken
Ressel	98 durchstrittig						steil
Trachten	92 niedrig						hoch
Vorwurflänge	110 kurz						lang
Sch. outerlänge	111 kurz						lang
Vorwurfaufh.	108 locker						fest
Zentralband	121 nicht ausg.						stark ausg.
Faltenboden	123 tief						hoch
Strichlänge	85 kurz						lang
Strichdicke	85 dünn						stark
Strichplatz. vorne	119 außen						innen
Str. Stellung hinten	118 nach außen						nach innen
Euterzeitigkeit	100 Reibmstr.						reiner Euter

Mänge: erwünschtes Bereich
gelesen: nichtgelesen
blutig: nichtblutig
Farbbeschreibung: 334 dunkelgelb 294 rot
819 gedeckt 159 gefleckt
Aufflecken: 424 bisaktive 115 aktiv

Mobilisation in [%] am Gesamtgehalt



Tab. 4: Nachkommenprüfung Feld Jungeber der geprüften PI-Eber (Geburtsjahre 2003-2005)

	Station							
	Alle		Landshut		Neustadt		Bergheim	
	Alle	Lebende	Alle	Lebende	Alle	Lebende	Alle	Lebende
Anzahl Eber mit Nachkommen	143	80	45	18	65	46	33	16
Nachkommen/Eber	16	22	7	11	26	32	10	4
Alter	209	206	232	229	204	203	213	218
Gewicht	141	140	149	148	139	139	141	143
Tägl. Zunahme	679	684	646	650	688	689	667	657
Speckdicke	7.7	7.7	8.5	8.7	7.5	7.5	8.0	8.4
Bemuskelungsnote	8.5	8.5	8.6	8.5	8.5	8.5	8.3	8.4
Körzuchtwert	129	130	124	128	130	131	127	121
Wertklasse I+II(%)	93.4	93.6	98.1	99.0	92.7	92.8	92.1	95.2


Abbildung 3: Ergebnisdarstellung der Zuchtwerte Exterieur

Wie liegen Forschungsdaten vor und wo kann man sie finden





Forschungsdaten ...

- ... allgegenwärtig
- ... über Forschungsinstitute verstreut
- ... von Wissenschaftlern erhoben und meist von denen selbst verwaltet
- ... häufig schlecht dokumentiert  schlecht zugänglich und nicht langfristig gesichert



Forschungsdaten sind ...

... disziplinspezifisch, heterogen und komplex

... häufig einzigartig und nicht reproduzierbar

➔ Gewinnung dieser Daten mit hohem finanziellen, technischen, intellektuellen und organisatorischen Aufwand verbunden

Forschungsdatenmanagement, aber wie



Forschungsdatenmanagement

- Im Print-Zeitalter:
 - Veröffentlichung der ausgewerteten Daten (in Auswahl) als Anhang einer Publikation
- ➔ Überprüfbarkeit der Forschungsdaten damit nicht gegeben

Forschungsdatenmanagement

- Im digitalen Zeitalter:
 - Speichern von Messdaten
 - Wertlos sobald der Wissenschaftler, der gemessen hat nicht verfügbar ist bzw. er nach einiger Zeit die Übersicht über die Messdaten verloren hat

Forschungsdatenmanagement

- Archivierung:

- „schließt das Speichern von Metadaten ein, also aller Informationen, die benötigt werden, um die reinen Messdaten zu verstehen und ggf. nach zu messen auch in Abwesenheit dessen, der die Daten zuerst gemessen und gespeichert hat“



Archivierung geschieht heute selten und oft unzulänglich

Forschungsdatenmanagement - Fazit

- ➔ Partielle Nutzung der Datenbestände, da sie in der Regel nur einer begrenzten Gruppe von Wissenschaftlern bekannt und zugänglich sind
- ➔ Verlust vieler Forschungsdaten nach wenigen Jahren

Forschungsdatenmanagement - Fazit

- ➔ Öffentliche Zugänglichmachung: selten
- ➔ Nachnutzungspotenzial geht verloren
- ➔ Beeinträchtigung der Forschungseffizienz

Die erforderliche Mehrarbeit der Aufbereitung, der Dokumentation und Qualitätssicherung findet scheinbar keine Anerkennung im Wissenschaftsbetrieb



Wer & wie fördert Open Data in der Forschung

Förderer von OD in der Forschung:

- CODATA
- CSTP
- NIH
- Welcome Trust
- OECD
- DFG
- Leibniz-Gemeinschaft
- Helmholtz-Gemeinschaft
- Fraunhofer-Gesellschaft
- Max-Planck-Gesellschaft

Wie sieht die Förderung aus ...

- „OECD Principles and Guidelines for Access to Research Data from Public Funding“
(April 2007)



Warum Open Data



- Forschungseffektivität = Vermeidung unnötiger Doppelarbeit
- Verifizierbarkeit u. Überprüfbarkeit datenbasierter Forschungsergebnisse
- Möglichkeit der Auswertung und Verwertung der Daten für Folgeprojekte
- Beschleunigung des Forschungsprozesses durch Data Sharing

Warum Open Data



- Gewinn neuer Erkenntnisse durch Zusammenführung von Daten aus verschiedenen Quellen
- Förderung der öffentlichen und wirtschaftlichen Nachnutzung von Daten
- Transparenz der Forschung

Open Data – Anreize

- Institutionalisierung der Datenveröffentlichung
- Eindeutige Auffindbarkeit, Zitierbarkeit
- Formalisierung der Veröffentlichung von Daten
- Bekanntmachung der Existenz der Daten

Anforderungen an die Veröffentlichung von Forschungsdaten

- Sicherstellung der langfristigen Auffindbarkeit durch dauerhafte Adressen (DOIs) und deren eigenständige Zitierbarkeit
- Metadatenerfassung der Daten bzw. Datensammlungen auf der Basis von Standards (z.B.: Dublin Core, oder ISO 690-2)

Anforderungen an die Veröffentlichung von Forschungsdaten

- Honorierung der Bereitstellung nachnutzbarer Daten als wissenschaftliche Leistung
- Schaffung von Rahmenbedingungen welche den Datenmissbrauch vermeiden
- Berücksichtigung disziplinspezifischer Anforderungen



Was macht die ZB MED

- Dialog mit WissenschaftlerIn
- DOI-Vergabe

**➔ ZB MED ist kompetenter Dienstleister
für die Wissenschaftler/Innen**

Dialog konkret ...

- Anforderungen und Bedürfnisse der WissenschaftlerInnen in EUA ermitteln
 - Wie ist der Online-Zugang zu Forschungsdaten?
 - Wie ist die Nachfrage nach Langzeitarchivierung?
 - Wie ist die Bereitschaft eigene Forschungsdaten zur Verfügung zu stellen?

Ein mögliches Szenario ...

- Wissenschaftlerin Frau Sedimentum hat im Rahmen ihrer Forschung für das Umwelt-Analyse-Zentrum Daten über Bodenanalysen in Bonn gesammelt
- Das Umwelt-Analyse-Zentrum prüft, evaluiert, speichert und verwaltet die Daten
- Zusätzlich registriert Frau Sedimentum ihre Forschungsdaten auch bei der ZB MED

Ein mögliches Szenario ...

- Herr Obst liest in einer Fachzeitschrift über die Datensammlung von Frau Sedimentum. Er möchte deren Daten unter einem anderen Gesichtspunkt auswerten.
- In seiner Publikation „Gesunder Boden – gesunde Pflanzen. Ertragssteigerung der Apfelernte unter Berücksichtigung des Nährstoffgehalts im Boden,“ zitiert Herr Obst die Daten von Frau Sedimentum durch deren DOI

**Danke
für Ihre Aufmerksamkeit!**