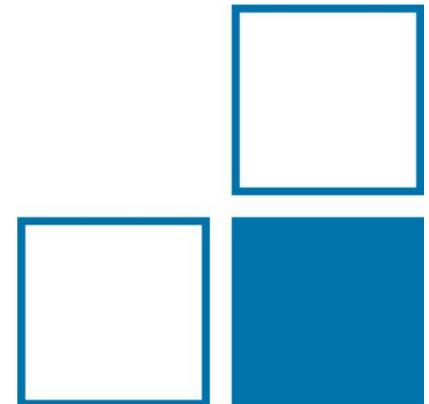


Implementierung der FAIR-Prinzipien im Forschungsdatenmanagement

Eine Terminologie-basierte Strategie für die inhaltliche
Beschreibung numerischer Faktendatensätze

Giacomo Lanza, Joachim Erich Meier
PTB – Wissenschaftliche Bibliotheken



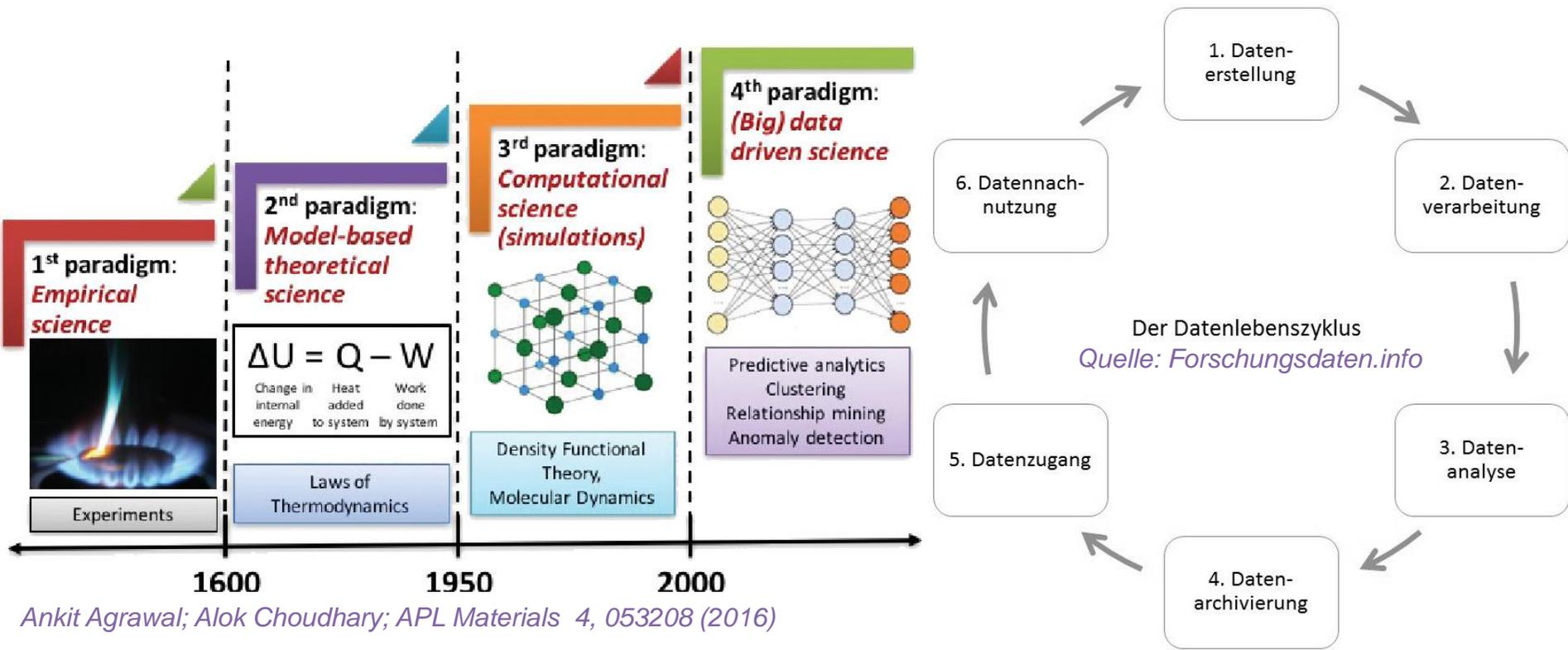
- Problemstellung
 - Datennachnutzung; FAIR data
 - Metadaten
 - Numerische Faktendaten

- Lösungsansatz und Realisierung
 - Metadatenschema; Kontrollierte Vokabulare
 - Demonstrator
 - Akteure und Aufgaben

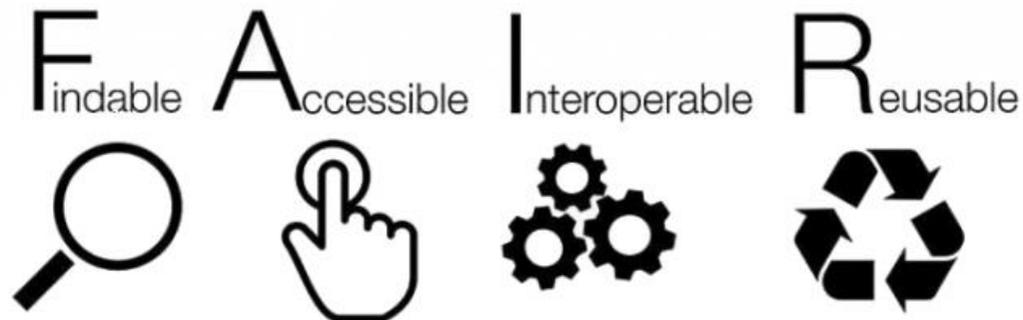
- Fazit und Vision

Motivation

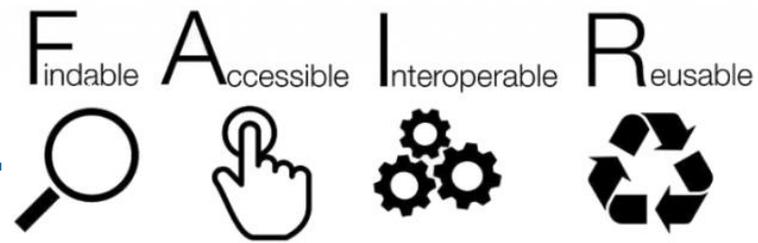
- Hohe Datenverfügbarkeit
- Datenbasierte Forschung
 - Metastudien; Recherche; Datenkuratierung; künstliche Intelligenz...



- Zunehmende Datenmengen
- Pflicht (Gute Wissenschaftliche Praxis; Zeitschriften...)
- **Nachnutzbarkeit und Reproduzierbarkeit angestrebt**



Anforderungen



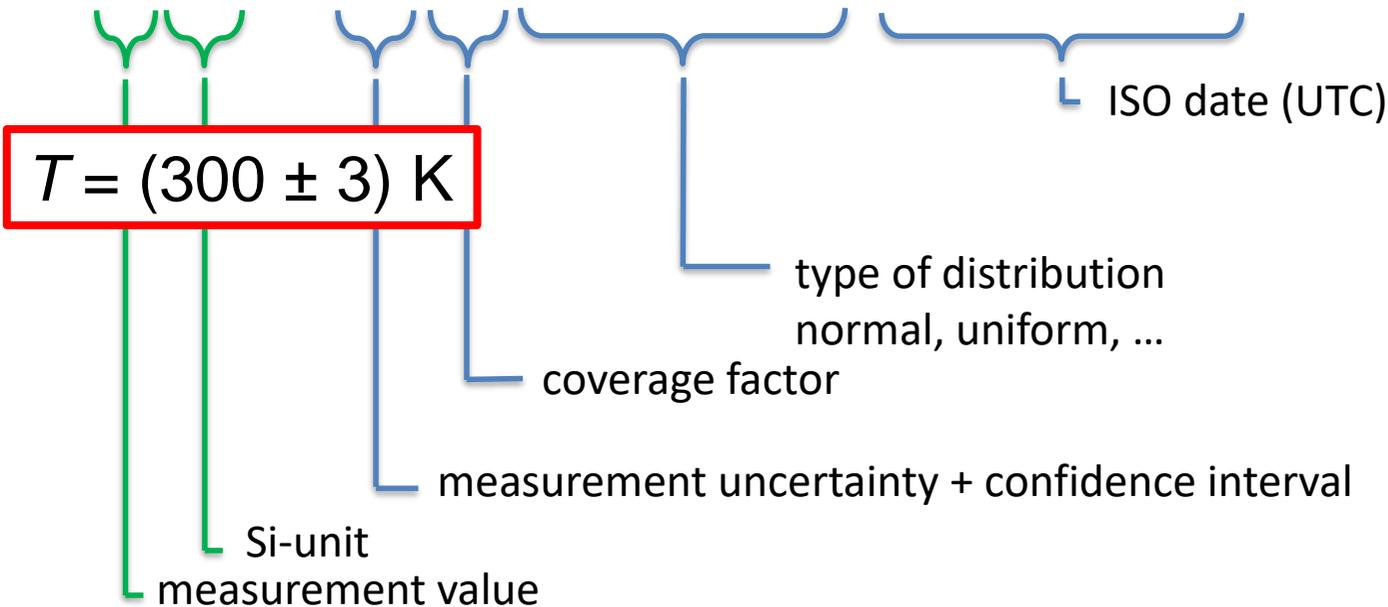
- Dateiformat: langfristig lesbar, offen
- Identifikatoren: global eindeutig, langlebig
- Metadaten: präzise, indexiert, relevant, langlebig, standardisiert (Fachvokabular)
- Übertragungsprotokoll: offen, universell
- Lizenz: eindeutig, dokumentiert

- Kein einheitlicher Standard
 - DataCite: <title>, <subject>, <keywords> ...
 - Fachspezifische Ansätze = „Insellösungen“

 - Gemeinsame Sprache von Datenportalen und Suchmaschinen
 - Präzise Angaben für eine spätere Suche notwendig!
- Auffindbarkeit sehr erschwert!
- Erweiterter Metadatenstandard benötigt
- fachübergreifend gültig, praxisnah anpassbar, maschinenoperabel

Numerische Faktendaten

$Y = y \text{ [SI]} \pm U_{95}(k) \text{ distribution Timestamp}$



Vorschlag: **D-SI**

```
<si:real>
  <si:label>temperature</si:label>
  <si:value>300</si:value>
  <si:unit>K</si:unit>
  <si:uncertainty>3</si:uncertainty>
  <si:coverageFactor>2</si:coverageFactor>
</si:real>
```

- Feldbezeichnungen ← Metadatenschema / Namensraum
 - Merkmale (Größe, Einheit, Unsicherheit, Wertebereich ...)
 - Datengenerierung (Messung / Simulation / Modell + Parameter)
 - Versuchsobjekt (Probentyp; Proben-ID; chemische Identität; Auftraggeber)

- Zulässige Feldwerte ← kontrollierte Vokabulare / Thesauri
 - je Feld (Größen; Einheiten; Materialien; Messverfahren; Naturkonstanten ...)

Metadatenschema

- Feldbezeichnungen; hierarchische Anordnung und Regel zum Befüllen
- Startpunkt: **D-SI**
- syntaxunabhängig (.xml, .json, .yaml ...)

Feld	Beschreibung
block	Zusammenhängender Tabellenbereich
block/npoints	Anzahl Messungen
block/quantity	In einer Datei enthaltene Größe
block/quantity/role	"variable", "parameter" oder "measurand"
block/quantity/name	Identifikator der Größe
block/quantity/description	Ausführliche Bezeichnung
block/quantity/list/label	Kurzbezeichnung
block/quantity/list/real[label="min"]/value	Minimum
block/quantity/list/real[label="max"]/value	Maximum
block/quantity/list/unit	Identifikator der Einheit
block/quantity/list/uncertainty	Erweiterte Messunsicherheit
block/quantity/list/coverageFactor	Erweiterungsfaktor (default=2)

- Ambiguitäten lösend (Homonyme; Synonyme; Sprachveränderung; falsche Freunde; Fachjargon)
- sprachunabhängig → Persistente Identifikatoren
- maschinenlesbar → Format (.rdf, .owl, .json ...)
→ Mirrors; Resolver (PID → pURL)
- Dienen auch als „Wissensspeicher“
(Symbol; Formel; Definition; feste Werte ...)
- Beziehungen (z. B. Größe ↔ Einheit)

Kontrollierte Vokabulare



< > Print Export	
 International Electrotechnical Commission	
IEC 61360 - Common Data Dictionary (CDD - V2.0014.0016)	
<u>UNIT</u>	
Code:	0112/2///62720#UAA185
Version:	001
Revision:	02
IRDI:	0112/2///62720#UAA185#001
Preferred name:	kelvin
Synonymous name:	
Short name:	K
Definition:	unit of thermodynamic temperature is the fraction 1/273,16 of the thermodynamic temperature of the triple point of water
Note:	base unit of the International System of Units (SI)
Remark:	UN/ECE code: KEL
Definition source:	ISO 80000-12:2009 \ ISO 80000-5:2007
Definition class:	
Unit structure:	K
Unit in text:	kelvin
Unit in SGML/XML:	K
Formula:	

qty:temperature

Type: PID-BasicInfoType-Metrology

[Digital Object View](#) [JSON View](#) [Versions View](#) [Show Relationships](#)

identifier

21.T11148/29481f51f16208d8170a

Short Name *

qty:temperature

Please use printable ascii characters without blank

Names in different languages

ar	درجة الحرارة المطلقة
ISO 639-1 (two-letter) code for the language in which the name is given.	Name of the object (quantity / unit / chemical element / constant) in the chosen language.
fa	دماي ترموديناميكي
ISO 639-1 (two-letter) code for the language in which the name is given.	Name of the object (quantity / unit / chemical element / constant) in the chosen language.
hi	ऊष्मगतिकीय तापमान
ISO 639-1 (two-letter) code for the language in which the name is given.	Name of the object (quantity / unit / chemical element / constant) in the chosen language.
zh	热力学温度
ISO 639-1 (two-letter) code for the language in which the name is given.	Name of the object (quantity / unit / chemical element / constant) in the chosen language.
ja	熱力学温度
ISO 639-1 (two-letter) code for the language in which the name is given.	Name of the object (quantity / unit / chemical element / constant) in the chosen language.
ms	Suhu termodinamik

<quantity>

<name>qty:temperature</name>

<si:list>

<si:real>

<si:value si:label="min">300</si:value>

<si:value si:label="max">1000</si:value>

</si:real>

<si:unit>unit:K</si:unit>

</si:list>

</quantity>

Suchen

Bezeichnung Sortieren Filter hinzufügen

Autor Sortieren

Messmethode Sortieren

Datum Sortieren

Vor

Suchen

Bezeichnung	Messmethode	Autor	Datum	Verknüpfung	Anzeigen	Bearbeiten
Bilancio energetico cottura degli spaghetti	Kalorimetrie	Robert Koch	1799-01-01	LINK	Anzeigen	Bearbeiten
Mesure de la température superficielle d'une lampe	Thermométrie	Joseph Louis Gay-Lussac	1840-08-10	LINK	Anzeigen	Bearbeiten
Bestimmung der natürlichen Schwingungsfrequenz eines Schaltkreises	elektrotechnische Frequenzmessung	Dr. Fuzz	1866-10-21	LINK	Anzeigen	Bearbeiten
Charakterisierung der elektrischen Leitfähigkeit einer Stahlmischung	Widerstandsmessung	Stan Laurel	1950-05-19	LINK	Anzeigen	Bearbeiten
Charakterisierung der	Kalorimetrie	Oliver Hardy	1950-05-20	LINK	Anzeigen	Bearbeiten

- Forschungsdaten → erhöhte, bis zielsichere Auffindbarkeit
- Numerische Daten in Artikeln / Berichten
(z.B. digitaler Kalibrierschein)
- Sensor-zu-Sensor-Kommunikation

Universell anwendbar (Wissenschaft + Industrie)

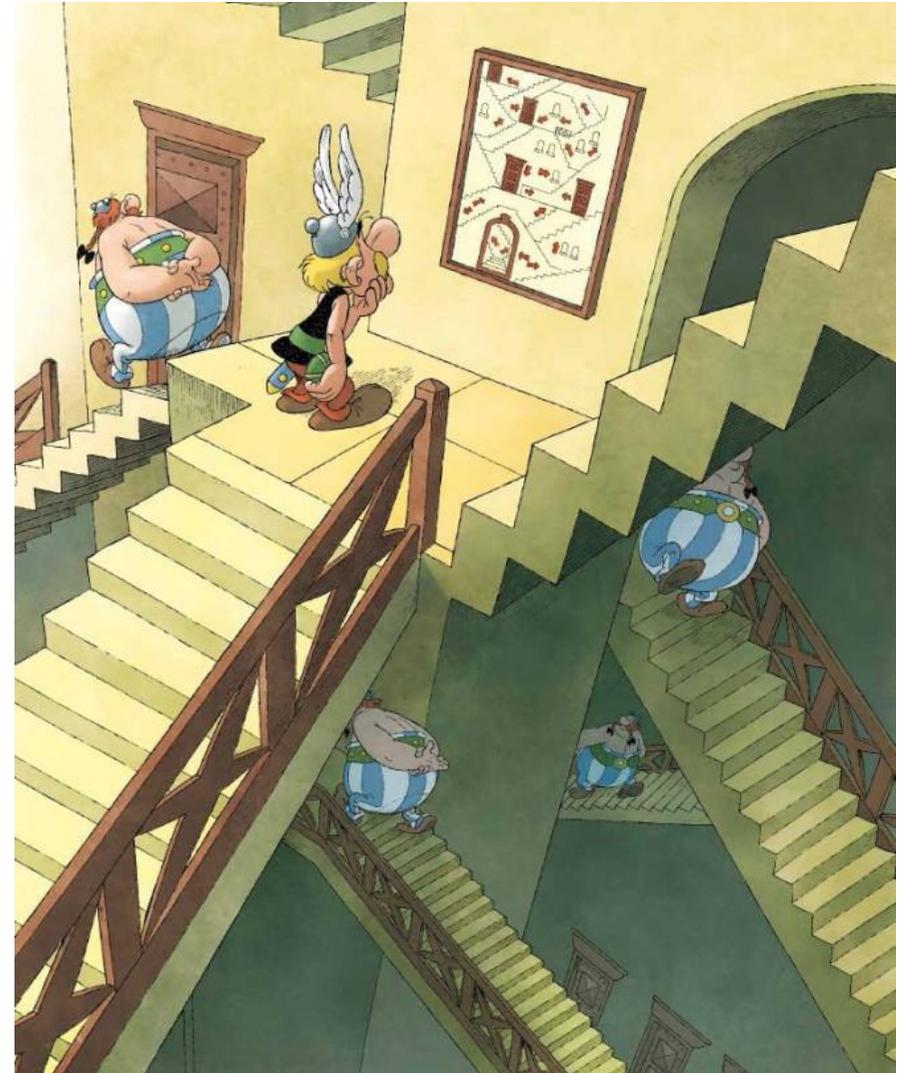
- Definition; Pflege; Übernahme; Implementierung
 - Metrologie-Institute (**PTB**, BIPM)
 - Forschungsdaten-Initiativen (**ePIC**, DataCite, RDA, GO FAIR ...)
 - Standardisierungsorgane (IEC, ISO, DIN, VDI/VDE, CODATA)
 - Datenrepositorien; Suchmaschinen; Infrastrukturanbieter

- Anwendungsfälle; gute Praxis; Feedback
 - Wissenschaftskonsorzen (**NFDI**, EOSC)
 - Fachgemeinschaften
 - Industrie

- Numerische Daten gezielt mit Metadaten erfassbar
- Standard-Metadatenstrukturen erforderlich
- Terminologischer Ansatz hilfreich

Vision: so nicht...

- Verzettelung
 - nichtkompatible Standards
 - voneinander unabhängige Aufgaben



Vision: sondern...

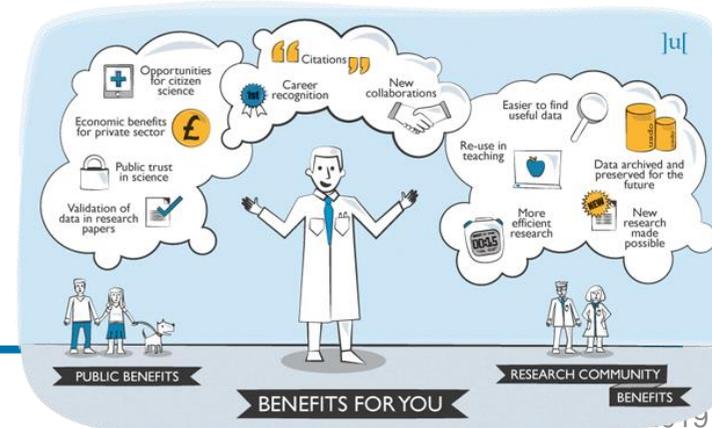
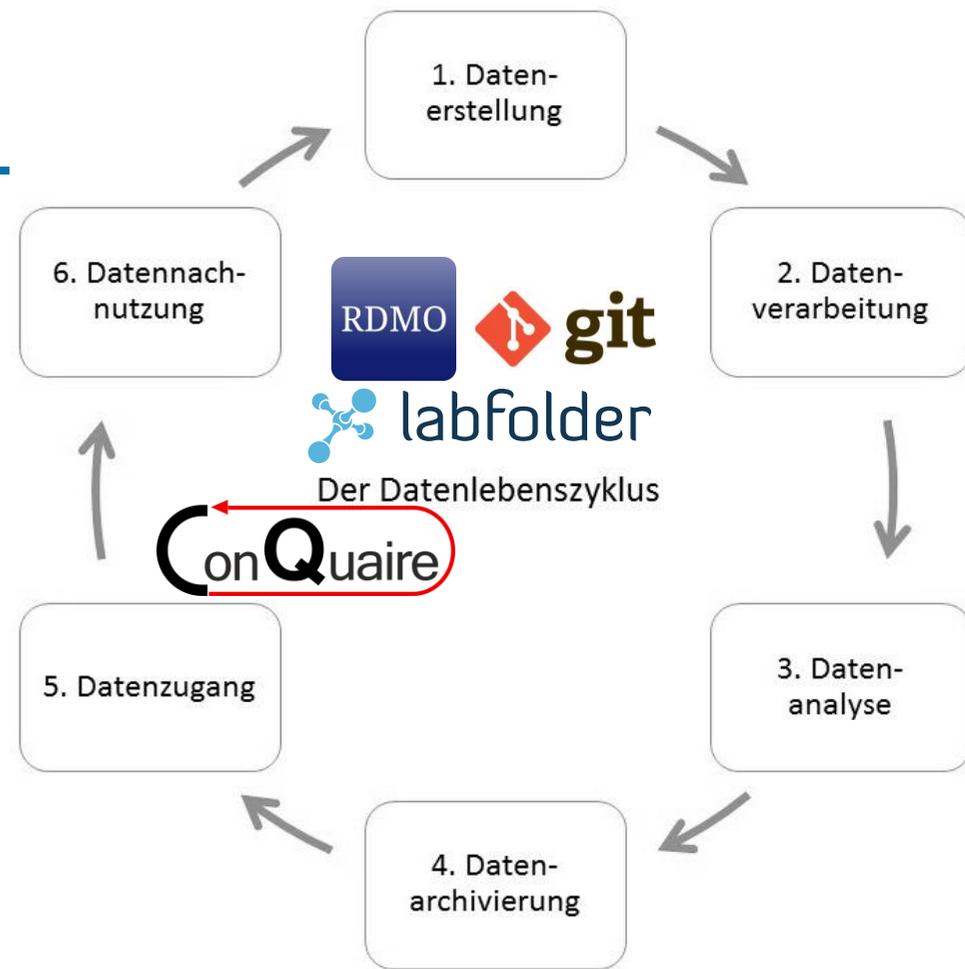
- Automatisierte Messprozesse

- verzahnte Werkzeuge
- nahtlose Übertragung von Daten und Metadaten entlang der Prozesskette

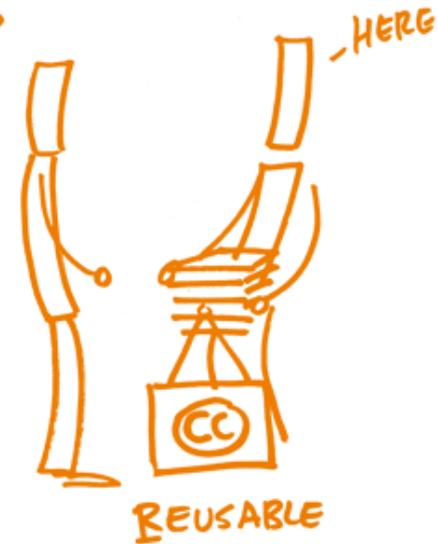
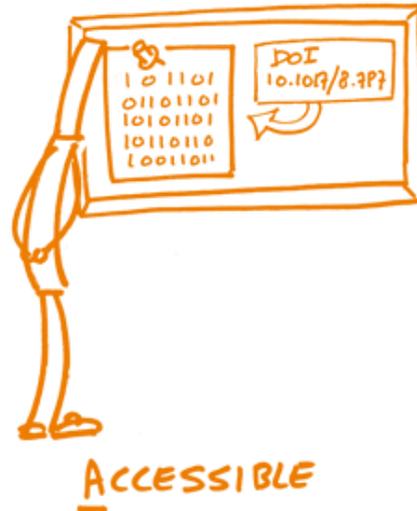
(Miteinbeziehung der Geräte- und Softwarehersteller)

- Anerkennung

- Kennzahlen
- Ressourcen



FAIR DATA PRINCIPLES



Vielen Dank für die Aufmerksamkeit!



Dr. Giacomo Lanza, giacomo.lanza@ptb.de

Dr. Joachim Meier, joachim.meier@ptb.de

Referat Q.11 – Wissenschaftliche Bibliotheken

Vielen Dank an:

U. Schwardmann (GWDG – ePIC)

T. Wiedenhöfer, D. Hutzschenreuter (PTB – SmartCom)

S. Eichstädt, R. Becker, J. Neumann (PTB – Digitalisierung)

