

Open Data Governance

Dr. Arne Leißner, Jo van Nouhuys, Condat AG

arne.leissner@condat.de,
jo.van.nouhuys@condat.de

AK Umweltinformationssysteme
03.05. – 04.05.2012
in Dresden

Offene Daten lesbar für Mensch und Maschine. Das ist das Ziel. (Berlin Open Data)

- **Anbieter:**
 - Verwaltungen
- **Nachfrage/Nutzer:**
 - Verwaltung, Wirtschaft, Wissenschaft, Zivilgesellschaft
- **Nutzung:**
 - Information,
 - (Weiter-)verarbeitung der Daten in eigenen Nutzeranwendungen,
 - Bereitstellung von Applikationen durch Dritte
- **Zugang:**
 - One-Stop-Shop-Portale,
 - weiterführende Links (Linked Data)

- **deStatis GENESIS-Online Datenbank**

https://www-genesis.destatis.de/genesis/online/data;jsessionid=CCC673565D6A17CBBAEFE31E75528F71.tomcat_GO_2_2?operation=statistikenVerzeichnis

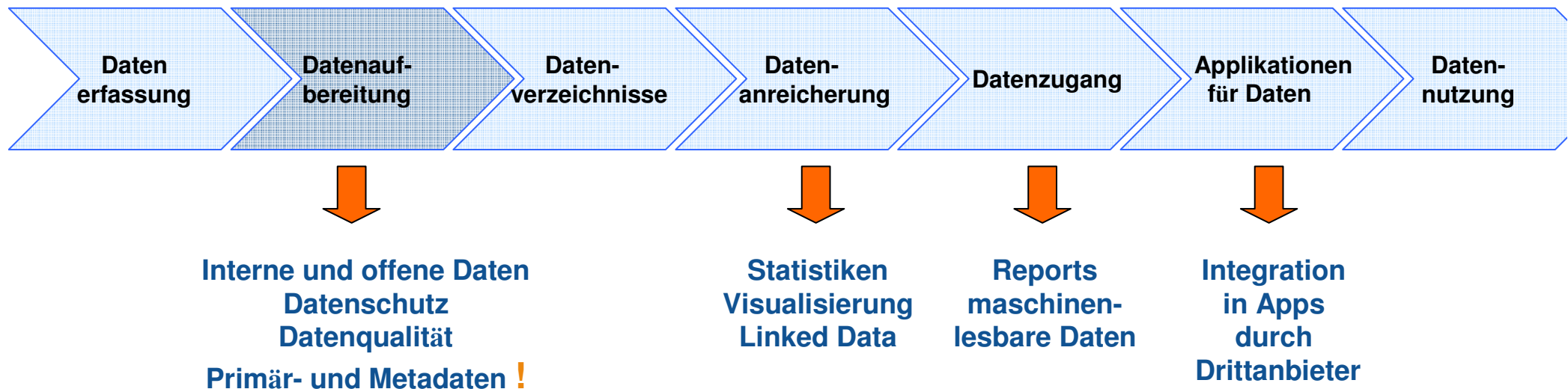
- **Initiative der Bundesregierung**

<http://www.daten-deutschland.de/>

- **Berliner Open Data Portal**

<http://daten.berlin.de/>

- **Open Government Data: „Wertschöpfungskette“**



- **Besonderheiten Umweltbehörden: Vielfalt heterogener Fachdaten und Informationssysteme / Fachanwendungen, ...**

... Sicherstellung von

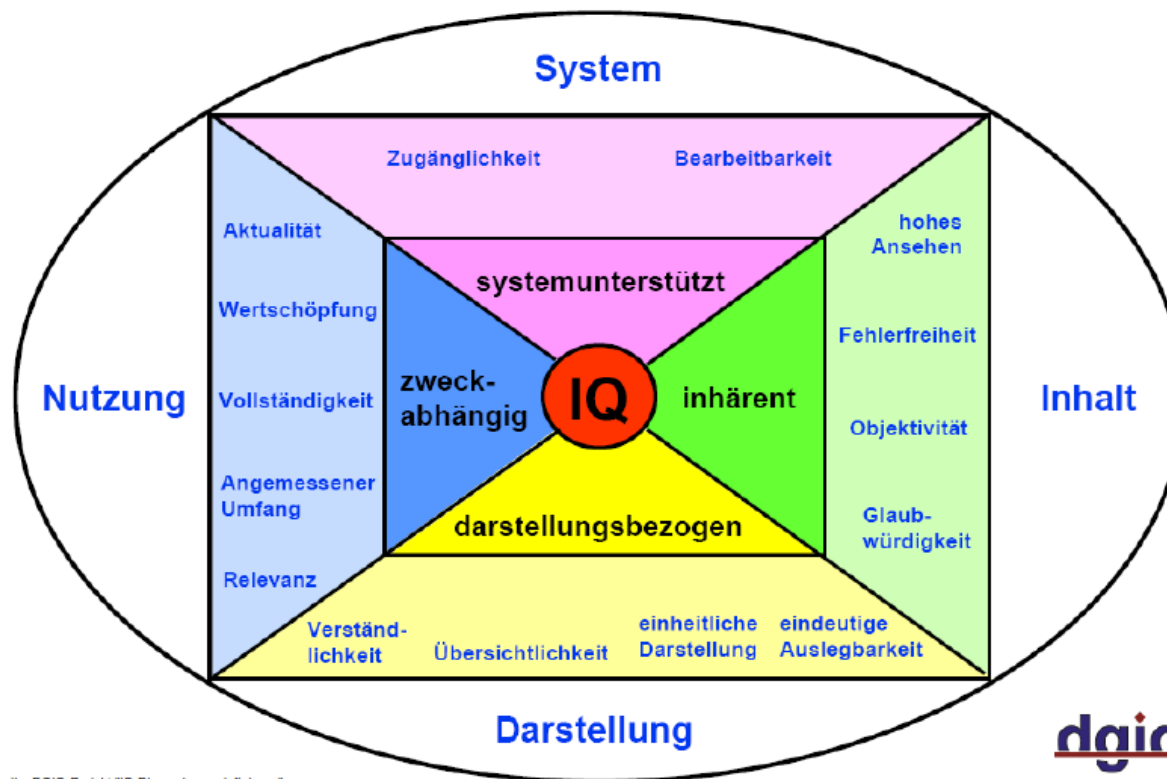
- Vertrauenswürdigkeit (Glaubwürdigkeit, Belastbarkeit)
- Eignung für den vorgesehenen Verwendungszweck (Relevanz, Angemessenheit)
- Integrationsfähigkeit (-> in Beziehung setzten, anreichern)
- Transformierbarkeit
- Beschreibbarkeit und Recherchierbarkeit (Zugangsstandards)
- Schutzwürdigkeit
- ...

Open | Data
Data | Governance
Open | Data | Governance

... in Einbeziehung und Beachtung von

- Sach- und Faktendaten <-> Metadaten <-> Master Data (Kataloge, Register),...
- Trends: Social Data, Unstructured Data, Big Data
- **Open Data Governance** -> Daten als wertvolles Gut behandeln („Datenschätze“)
- **Datenqualität** als Teildisziplin von Data Governance

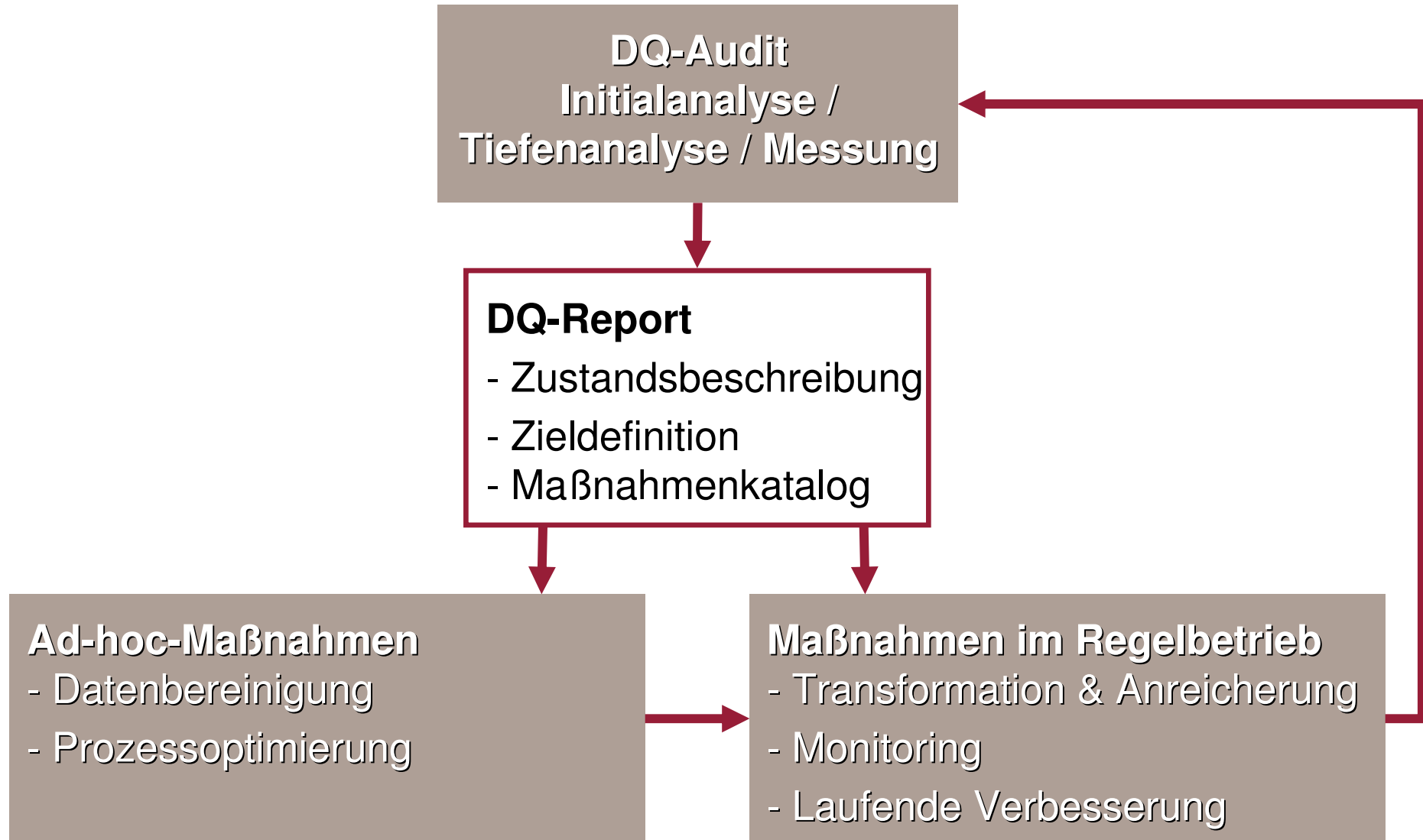
Informationsqualität: 15 Dimensionen, 4 Kategorien

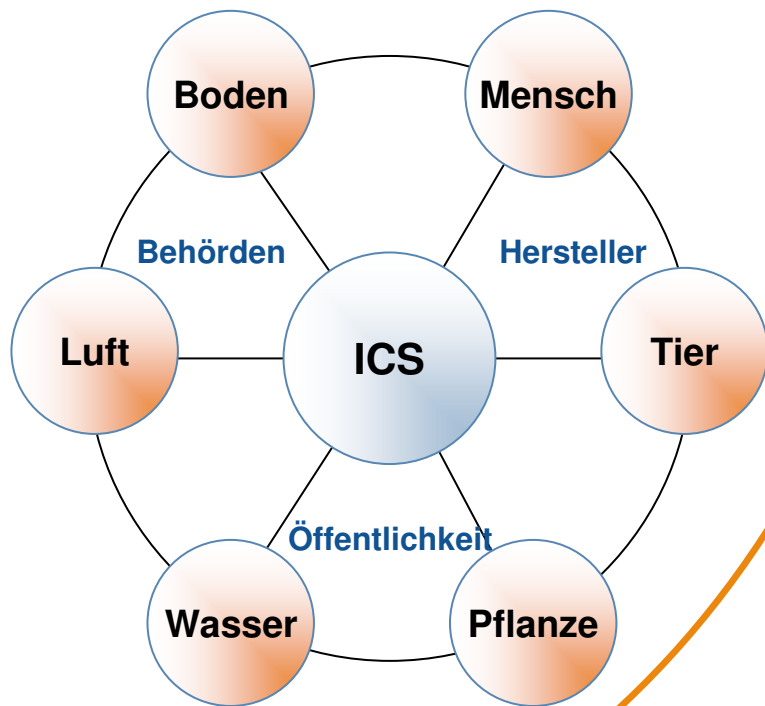


Quelle: DGIQ-Projekt "IQ-Dimensionen definieren"

These:

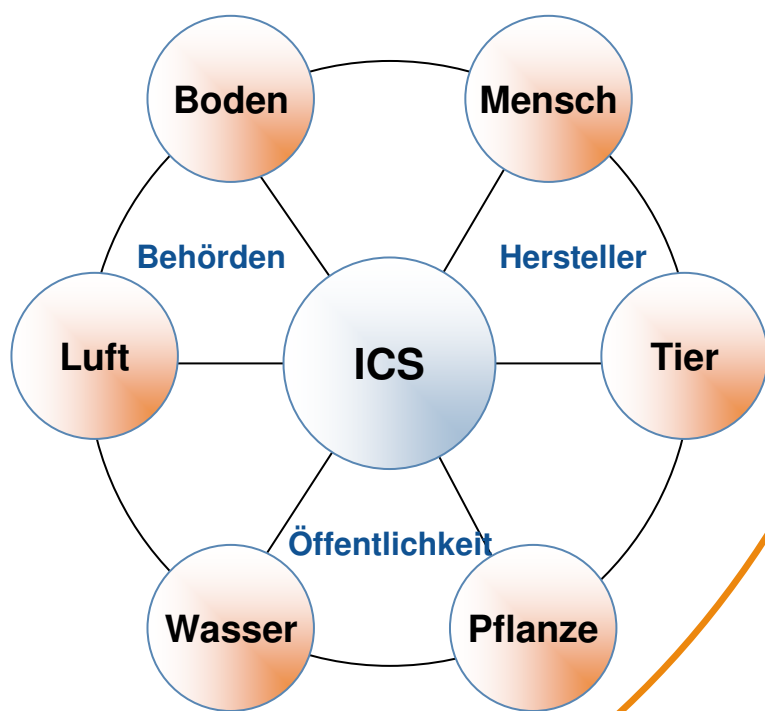
- Im Alltagsbusiness zumeist nur ausgewählte DQ-Aspekte relevant
- Im Kontext von Open Data eher der 360°-Anforderungsblick von Bedeutung





Informationssystem Chemikaliensicherheit

- Zentrale Vollzugsdatenbank
- Unterstützt Aufgaben der Stoffbewertung in unterschiedlichen Vollzügen
- Beinhaltet als Faktendaten strukturierte Daten aus Testberichten und Literaturquellen zu
 - Kennzeichnung und Verwendung
 - Physikalisch-Chemische Eigenschaften
 - Ökotoxikologie
 - Toxikologie
 - Umweltverhalten
- Manuelle Datenerfassung + diverse alte und neue Importschnittstellen
- Viele Textfelder
- Datenbestände in der Vergangenheit nicht über automatisierte Verfahren qualitätsgesichert
- Vermutung: Inhomogene Datenlage
- Open-Data-Relevanz: Hoch, Basis für Homogenität



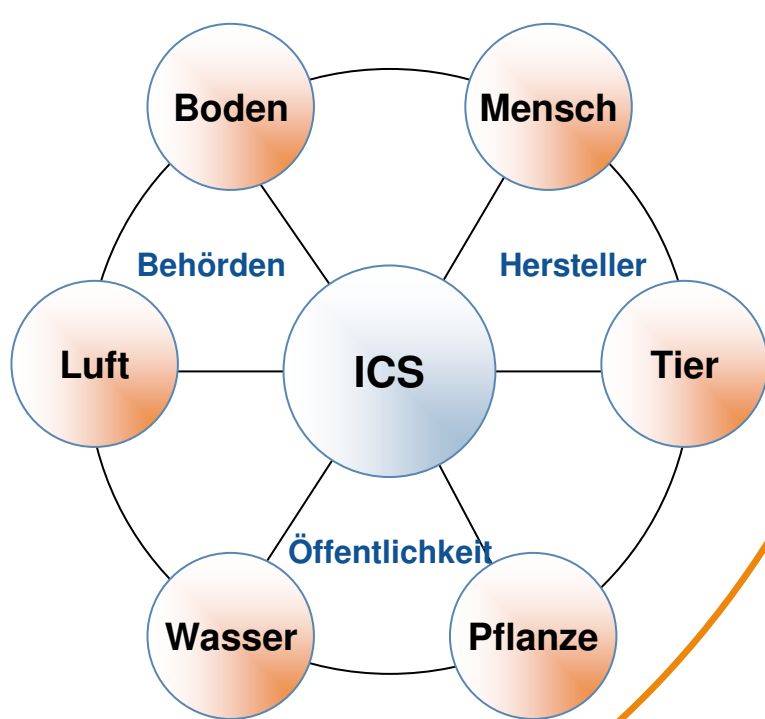
Hochqualitative Daten -> Effiziente Prozesse

DATA GOVERNANCE

- Data Quality
- Data Integration
- Master Data Management
- Metadata Management
- Enterprise Data Connectivity & Federation

- Methodik, die das Management von Unternehmensdaten als Asset steuert. Eine effektive Data Governance kümmert sich mit definierten Prozessen, Verantwortungen, Regeln und Ressourcen um die Verfügbarkeit, Integrität und Qualität der Unternehmensdaten.
- Kombiniert Personal, Prozesse und Technologie, um qualitative hochwertige Informationen zu gewinnen.
- Betrachtet die Organisation als Ganzes, so dass Daten gemeinsam genutzt und über Anwendungen, Systeme, Prozesse und Anwender hinweg wiederverwendet werden können

Beispiel UBA: ICS als zentrale Vollzugsdatenbank



Hochqualitative Daten -> Effiziente Prozesse

DATA QUALITY

- Erwartungsgerechtigkeit
- Vertrauenswürdigkeit
- Belastbarkeit
- Regelkonformität (Compliance)

Maßnahmen zur Messung und nachhaltigen Verbesserung der Datenqualität

- > akkurat und verlässlich,
- > vollständig und widerspruchsfrei,
- > integrationsfähig und migrationsfähig,
- > identifiziert eindeutig,
- > dublettenfrei,
- > standardisiert,
- > verifiziert

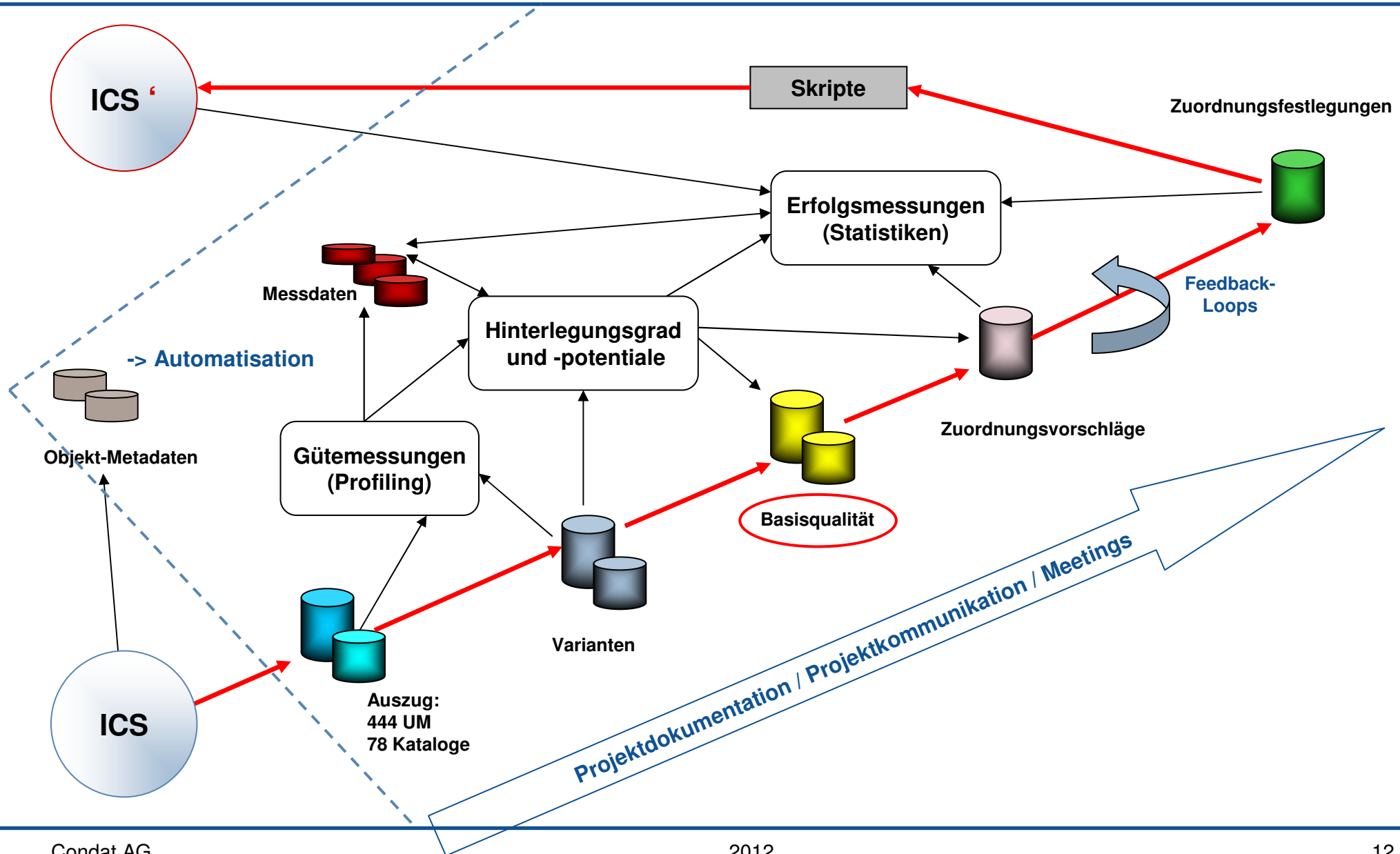
Erarbeitung von Methoden der Qualitätssicherung für das Informationssystem Chemikaliensicherheit ICS und deren Anwendung auf den Datenbestand des ICSalt

- Auslöser:
Migration zu ICSneu
- Thema:
 - Konsolidierung
 - Datenabgleich
 - Master Data Mapping
- Ziel:
 - Erhöhung der Akzeptanz
 - Beseitigung inhomogener Datenlagen (Bereinigung)

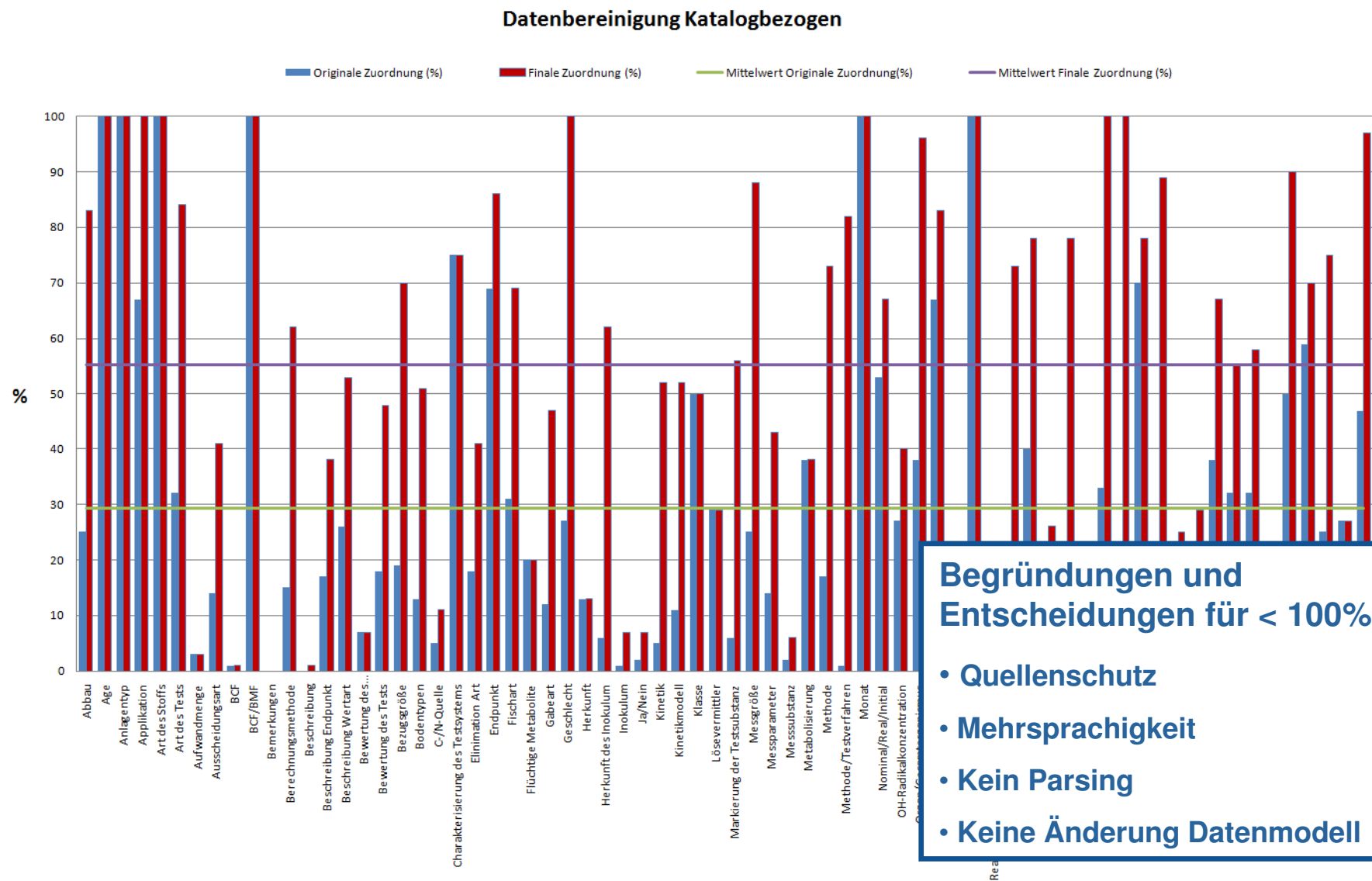
DQ-Erwartungen an die ICS-Daten

- ...
- ...
- (alle) Textwerte aus Untermerkmalen (Attributen) mit Katalogbezug sind im jeweiligen Katalog als Eintrag hinterlegt
- ...
- ...
- (alle) Angaben sind dem richtigen Merkmal zugeordnet
- ...
- ...

Beispiel UBA: Überblick zum Vorgehen



Beispiel UBA: Messbarkeit und erzielte Verbesserungen



Open Data Governance: Best Practice

- Klarheit zur Motivation (Pflichterfüllung versus Bedürfnis)
- In Strategie (Data Governance Initiative) einbetten und Sponsoren finden
- Start small and focused („Klasse statt Masse“)
- Überprüfbare Erwartungen an die Daten zu formulieren (auch wenn sie noch so trivial erscheinen) => Aufstellen von einfachen bis komplexen Regeln
- Messbarkeit sicherstellen-> Erfüllungs- bzw. Abweichungsgrad pro Regel (-> DQ-Index über die Gesamtheit)
- Immer mit dem Data Profiling beginnen
- Zunächst stets Basisqualität und Vergleichbarkeit herstellen bzw. simulieren
- Grenzen von Excel und SQL akzeptieren
- Tools für hochwertigen unscharfe Verfahren (Parsing, Standardisierung, Matching) einsetzen
- Weiterhin auch visuelle / manuelle (Prüf-)Tätigkeiten vorsehen
- In der Lage und gewillt sein, Entscheidungen zu treffen
- Zusammenarbeit IT, Fachabteilungen, Data Stewards organisationsübergreifend organisieren
- Erfüllung von Erwartungshaltungen bestätigen / Defizite identifizieren, quantifizieren und bewerten -> realistische Verbesserungsziele und –maßnahmen -> erneut messen