

Ontologiestützte Datenaustauschformate

Dargestellt am Beispiel der Musikindustrie

- Allgemeine Betrachtungen zum Thema Datenaustausch
- Ein Beispiel aus der Musikindustrie
- Praktische Aspekte des Einsatzes einer Ontologie

Steffen Lindek, Heidelberg

Herausforderungen des Datenaustauschs

- Unterschiedliche Datenmodelle
 - Bedingt durch unterschiedlichen Fokus oder Geschäftsprozesse
- Konsistente und effiziente Kommunikation
 - Herausforderung angesichts wachsender Transaktions- und Datenmengen
- Wichtigkeit von Klarheit und Standardisierung
 - Erleichterung der Implementierung
 - Verbesserung der Zukunftssicherheit

Datenschnittstellen

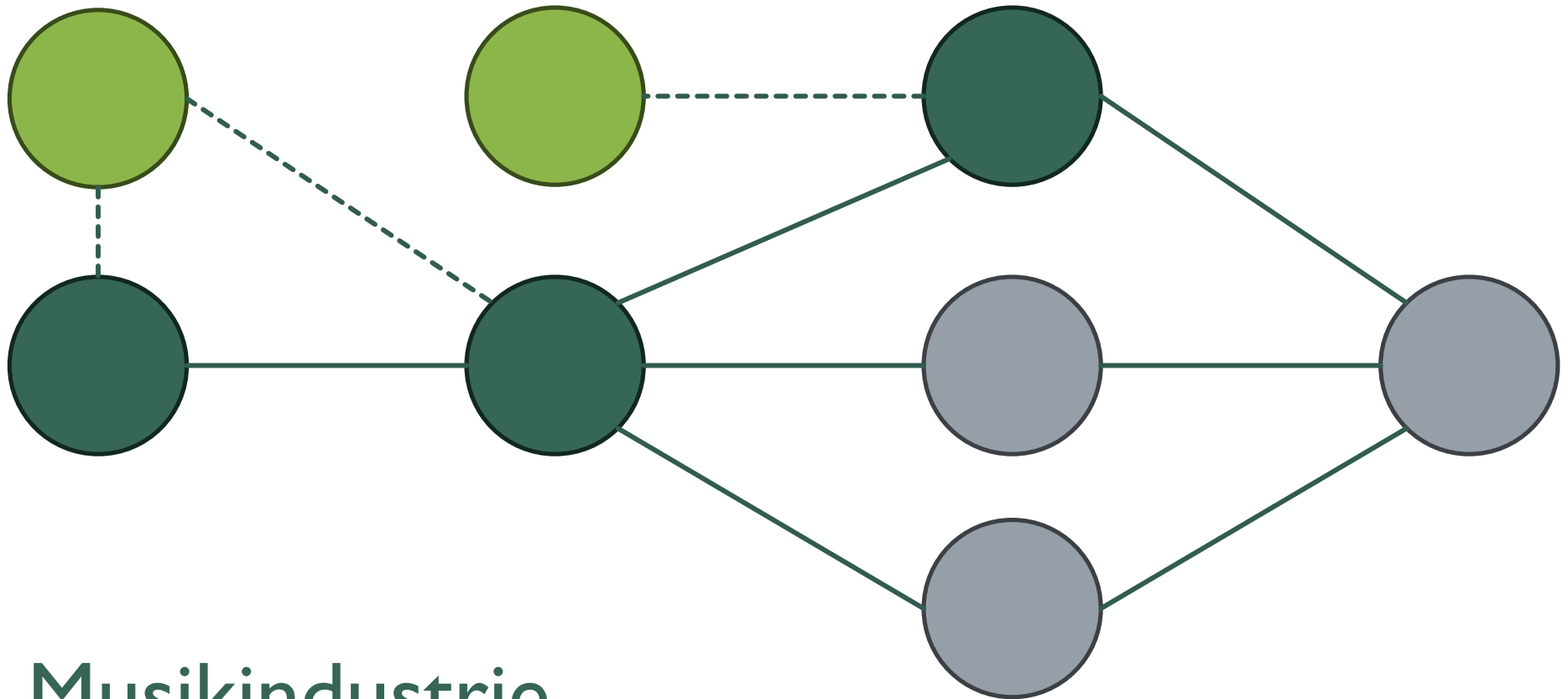
- Vorteil eines eigenen Datenmodells
 - Datenmodell sollte detaillierter als die Fremddatenmodelle sein
- Anwendungsbeispiele
 - Standardisierte Datenformate zum Datenaustausch zwischen Geschäftspartnern
 - Integration von diversen alten Datenbeständen
- Vorteile
 - Vereinfachung (je mehr proprietäre Formate, desto wichtiger)
 - Standardisierung
 - Skalierbarkeit

Konsistenz durch Ontologien

- Strukturierter Rahmen für die Daten
- Modellübergreifende Konsistenz und Kompatibilität
- Vermeidung von Missverständnissen und Informationsverlust
- Verbesserte Qualität der Daten
- Leichtere und schnellere Implementierung
- Langfristige Stabilität
- Zeit- und Geldersparnis

Praktische Anwendungen

- Kommunikation entlang einer Wertbildungskette
 - z.B. Buchhandel
- Integration von alten oder Fremddatenbeständen
 - z.B. bei der Übernahme von Firmen
- Kommunikation unter Projektpartnern
 - z.B. Austausch von Metadaten für wissenschaftliche Materialien (wie Gewebeproben in der Medizin)



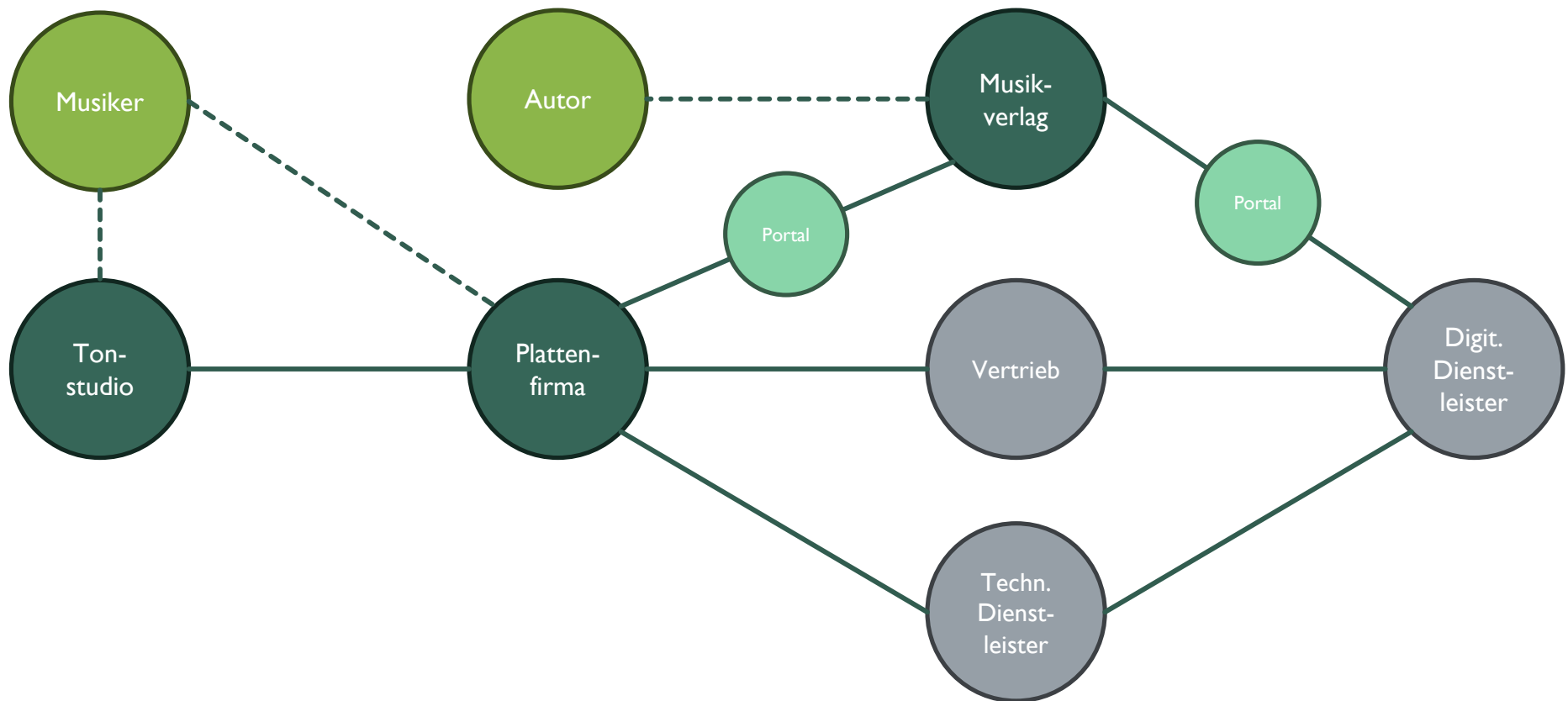
Musikindustrie

Beispiel für Datenaustauschformate

Beispiel: Musikindustrie

- Kommunikation entlang einer Wertbildungskette
- Sehr viele Geschäftspartner
 - Künstler (Komponisten, Texter, Musiker, Sänger, aber auch DJs, Tontechniker etc.)
 - Rechteverwertungsgesellschaften
 - Musikverlage, Tonstudios, Plattenfirmen
 - Vertriebspartner
 - Digitale Dienstleister (z.B. Streamingdienste)
 - Technische Dienstleister (z.B. *Music Recognition Technology*)

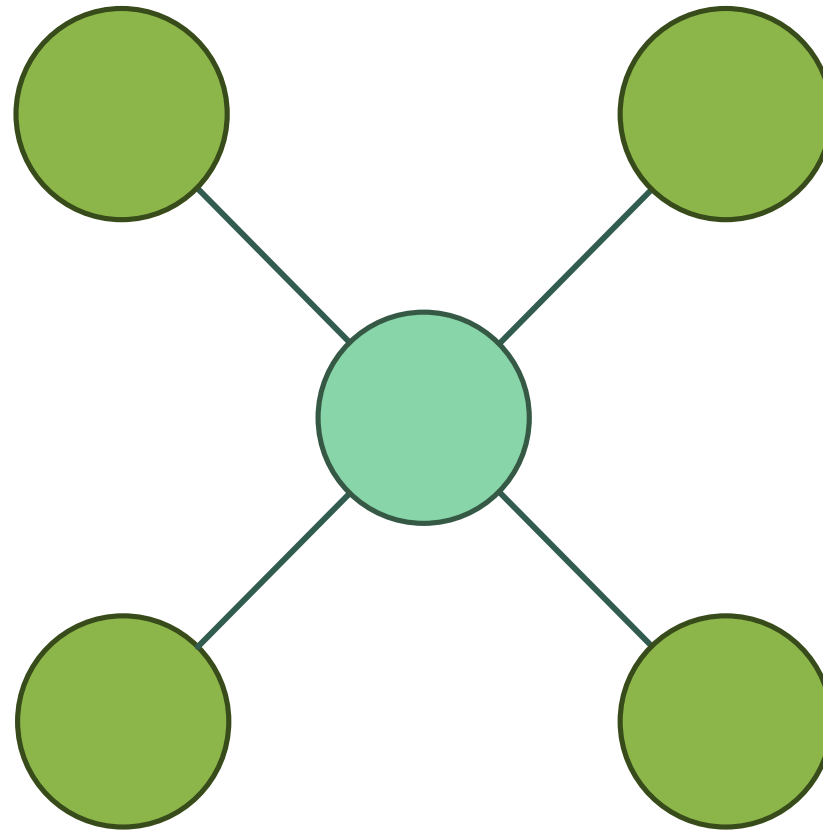
Kommunikation in der Musikindustrie



Typische Fragen

- Zu welchen Konditionen wird ein Release angeboten?
- Wer hat was zu welchem Preis verkauft?
- Wer hat einen Song geschrieben?
- Wer hat die Rechte an einem Song?
- Wer kann eine Nutzungslizenz erteilen?
- Welche Werke sind in einem Song enthalten?
- Welche Musik wird in einem Restaurant gespielt?
- Welche Songs wurden bei welchem Konzert gespielt?

- Digital Data Exchange (DDEX) - Standards für den Austausch von Daten in der Musikindustrie
- Formate für XML-Dateien, aber auch CSV und JSON
- Downstream: Musikmetadaten
 - Idealerweise vom Tonstudio bis zum Streaming-Dienst
- Upstream: Nutzungs- und Finanzdaten
 - Rückmeldung von Verkaufszahlen und Umsätzen zur Vergütung der Künstler
- Zeitnaher standardisierter Datenaustausch vermeidet Fehler und Ineffizienz



Ontologie

Praktische Aspekte des Einsatzes einer Ontologie

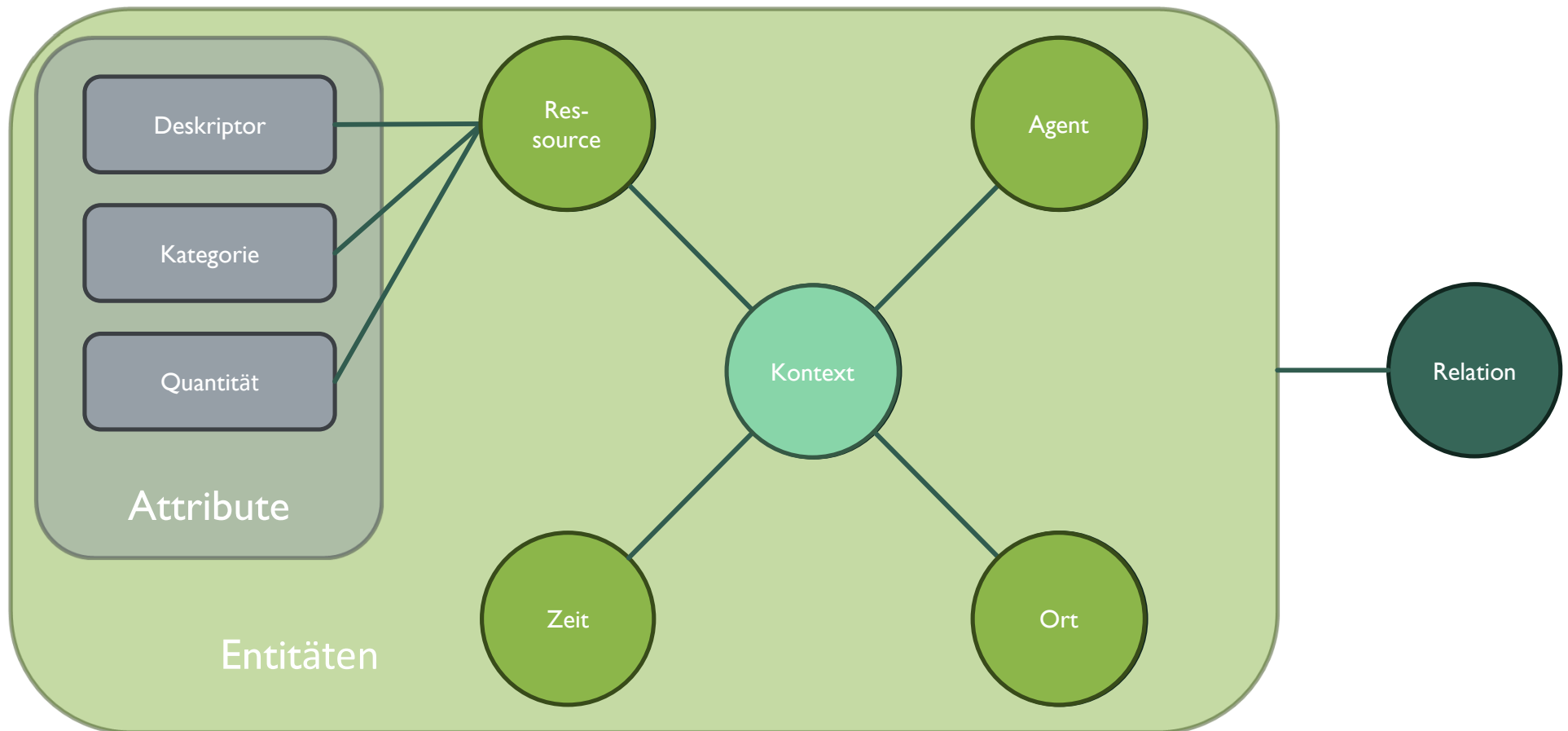
Ontologie zur Unterstützung der Datenformate

- Nennt sich *Data Dictionary*, ist aber die anwenderorientierte Oberfläche einer Ontologie
 - Namen, Definitionen, Annotationen
 - Hierarchie (Klassen, Instanzen)
 - Querverweise (durch Relationen zwischen Begriffen und Links in Definitionen)
- Begriffe werden als Elemente der Datenformate oder als kontrolliertes Vokabular verwendet
- Bausteine der Datenformate werden definiert
- Erlaubt die Generierung von Schemata für die Standards

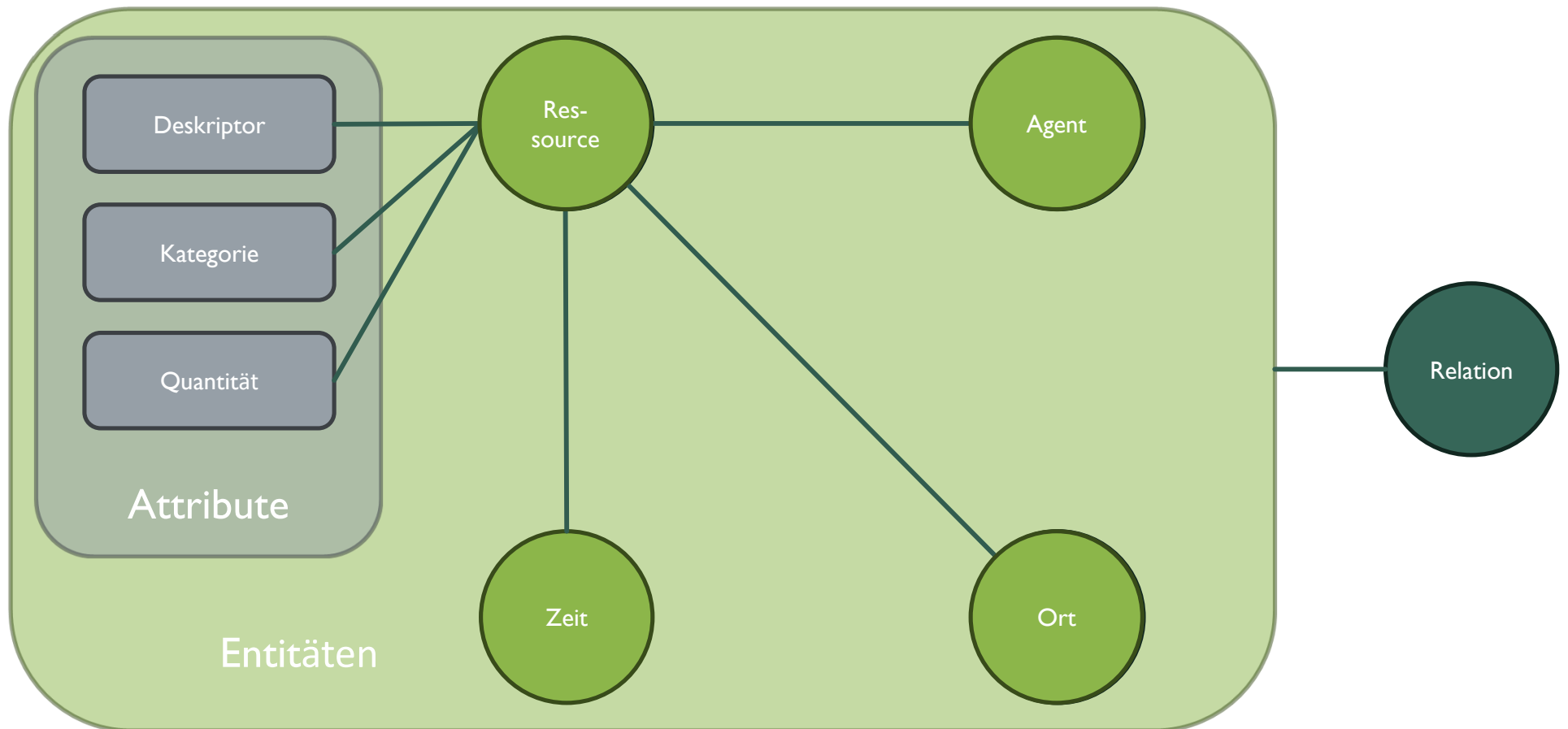
Ontologisches Grundmodell

- Kontext-zentrisches Datenmodell
 - Gegensatz zu objekt-zentrischem Datenmodell
 - Hauptentitäten: Kontext mit Ressourcen, Agenten, Zeit und Ort
- Noch wichtiger: Modell für die Attribute
 - Deskriptoren (Freitext: ID, Name, Annotationen)
 - Kategorien (kontrolliertes Vokabular)
 - Quantitäten (numerische Werte)
 - Zeitangaben
- Relationen zwischen den Entitäten

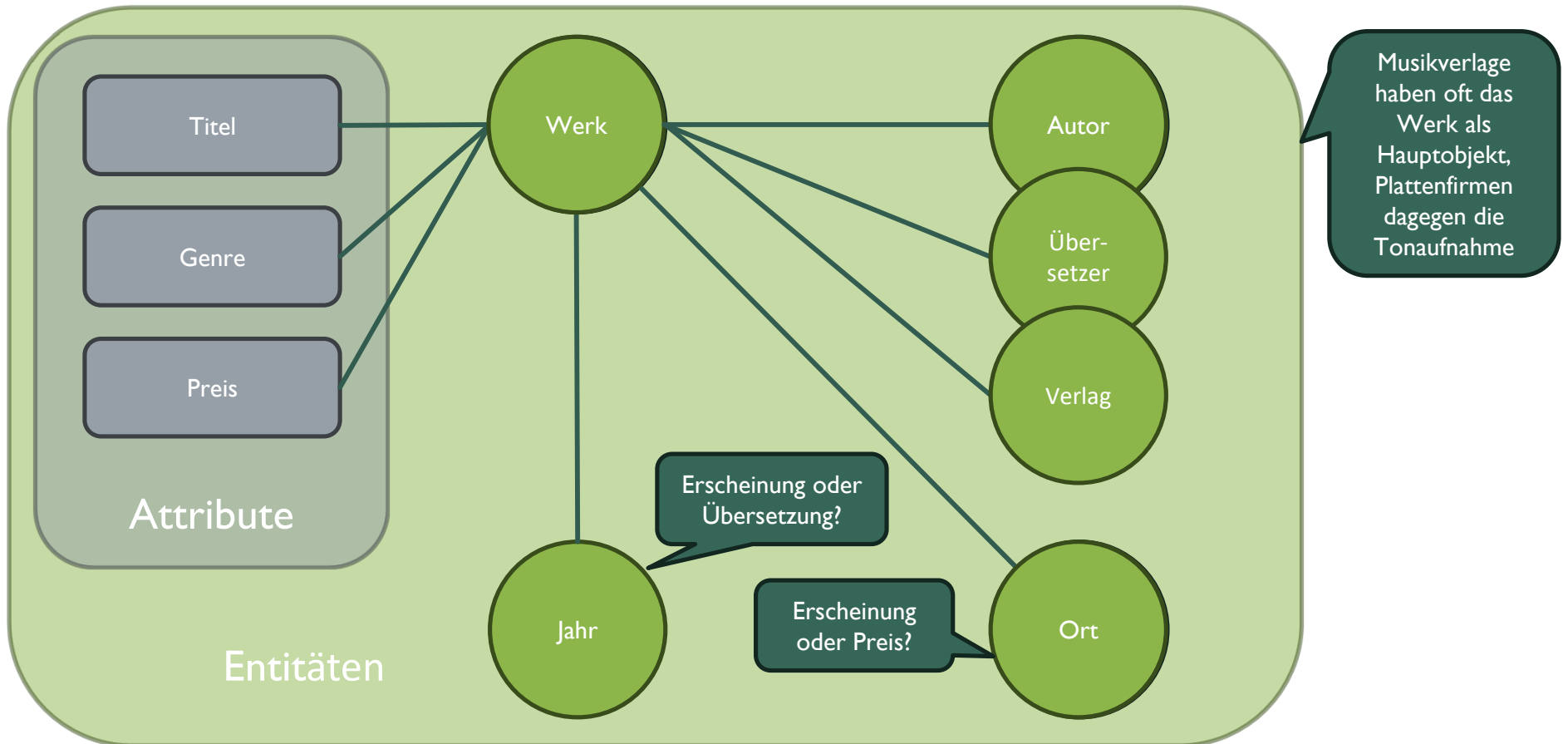
Ontologisches Grundmodell



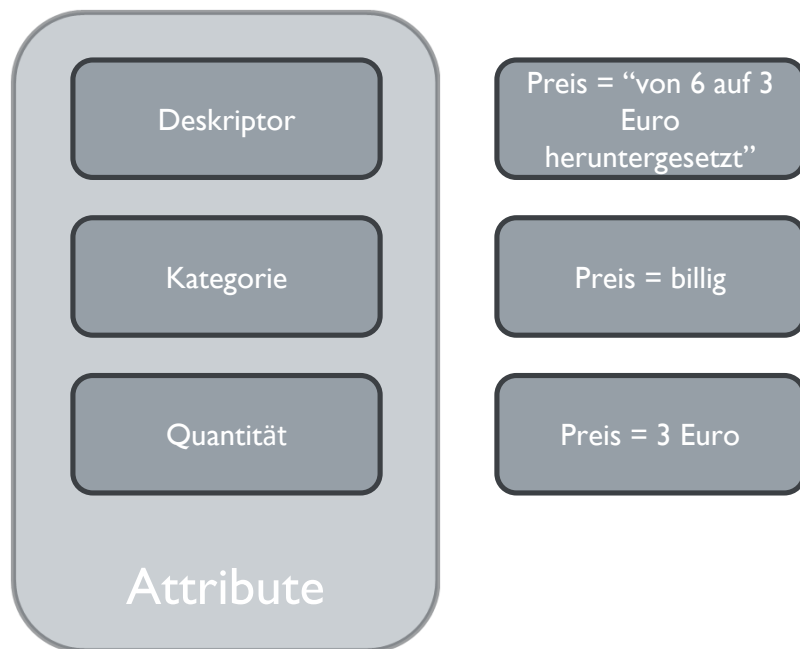
Objekt-zentrisches Modell



Objekt-zentrisches Modell – ein Beispiel



Modell für die Attribute



- Was ist der Datentyp?
- Wenn Kategorie, dann welche Werte?
- Wenn Quantität, dann welche Einheit?
- Quantitäten sind komplex: Einheit, Wertebereiche, Qualifikationen (z.b. circa)

Grundlegende Fragen, Teil I

- Was braucht man wirklich?
 - Unterscheidung von Klassen und Instanzen?
 - Keine sichtbaren Unterschiede für Nutzer
 - Unterscheidung von Verben (“Streamen”) und Kontext (“Streaming”)?
 - Unterschiede für Nutzer, aber nicht im Modell
 - Unterscheidung von Qualitäten (“Digital”) und Objekten, die diese Qualitäten haben (“DigitalesObjekt”)?
 - Unterschiede für Nutzer, aber nicht im Modell

Grundlegende Fragen, Teil 2

- Was braucht man wirklich?
 - Unterscheidung von ähnlichen Begriffen?
 - Beispiel: Rumba ist ein Tanz, ein Genre, ein Rhythmus
 - Synonyme für Nutzer, aber man braucht differenzierte Definitionen
 - Saubere Trennung der Begriffe von Anfang an 😞
- Pragmatischer Ansatz
 - In der Praxis reicht meistens eine Lösung ...
- Wichtig sind ein sauberes Datenmodell und eine konsequente Anwendung



Schlussbetrachtungen

Interoperabilität und Entwicklung von Standards

Der Vorteil der Interoperabilität

- Eindeutigkeit der Begriffe (unterschiedliche Bedeutung in den Industriezweigen)
 - Vermeidung von Fehlinterpretationen
- Einheitlichkeit des Vokabulars durch zentrale Entwicklung
 - Erleichterung des Verständnisses
- Konsistenz der Datenformate durch Wiederverwendung von Bausteinen
 - Erleichterung der Implementierung
- Flexibilität und Skalierbarkeit durch generische Strukturen
 - Wichtig wegen der Dynamik des Marktes

Entwicklung von Standards

- Unterschiedliche Geschäftspartner
 - Unterschiedliche Daten, Formate, Interessen, Terminologie
- Wichtigkeit von Arbeitsgruppen für Konsens
 - Wer braucht welche Daten?
 - Wer kann diese liefern?
- Ontologie gewährleistet Konsistenz bei Aufteilung der Entwicklung
 - Wichtig bei Arbeit in Gruppen und bei langfristigen Projekten

Vielen Dank

Steffen Lindek, steffen@lindek.de