

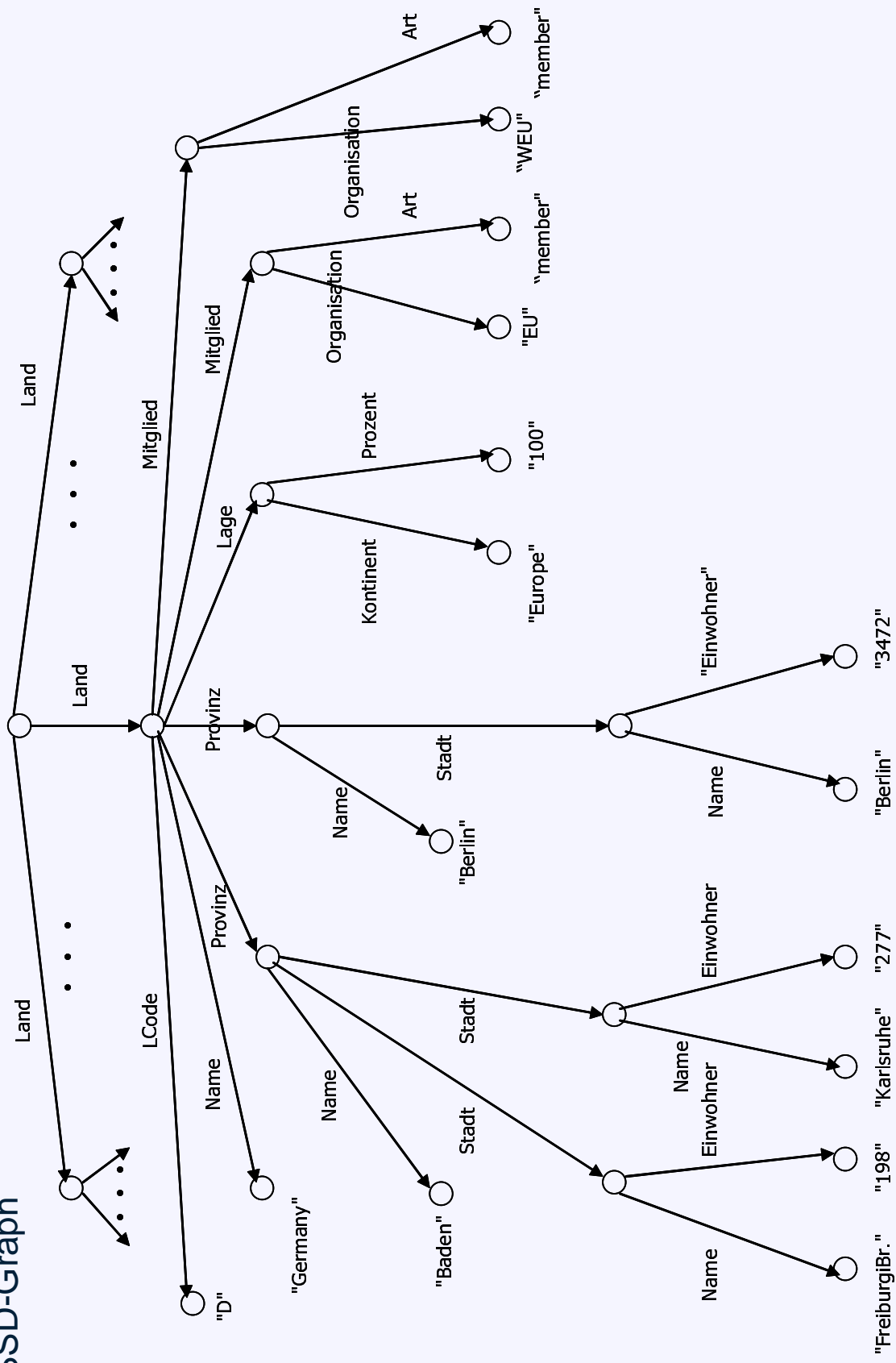
Semistrukturierte Daten (SSD)

- Relationale Datenbanken sind *nicht semistrukturiert*, die Relationen ihrer Instanzen sind gemäß einem in einem Schema festgelegten Typ und haben eine vollkommen reguläre Struktur.
- Bilddaten sind *nicht semistrukturiert*; sie sind vom Typ BLOB und haben im Allgemeinen keine von ihrem Inhalt separate Struktur.
- HTML-Dokumente sind *semistrukturiert*; sie haben im Allgemeinen eine irreguläre, durch Verwendung von Tags erkenntlich gemachte Struktur, die gemäß einem Typ HTML ist.
- HTML-Dokumente mit Bilddaten sind *semistrukturiert*.
- XML-Dokumente sind *strukturiert* oder *semistrukturiert*; sie haben eine reguläre oder irreguläre Struktur, die durch Verwendung von anwendungsspezifischen Tags erkenntlich gemacht ist. Es existiert nicht notwendigerweise ein anwendungsspezifischer Typ.

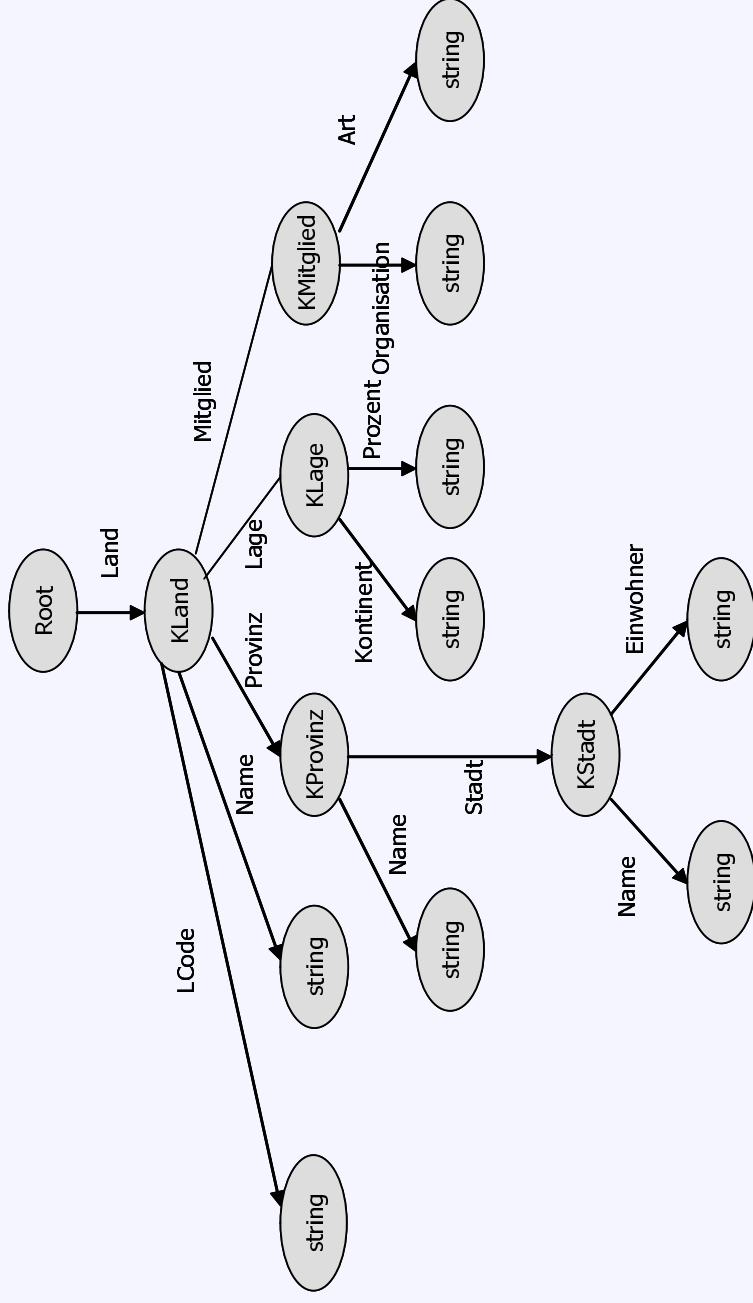
SSD-Dokument

```
{ Land:
  { LCode: "D", Name: "Germany",
    Provinz: { Name: "Baden",
      Stadt: { Name: "Freiburg i.Br.", Einwohner: "198" },
      Stadt: { Name: "Karlsruhe", Einwohner: "277" } },
    Provinz: { Name: "Berlin",
      Stadt: { Name: "Berlin", Einwohner: "3472" } },
    Lage: { Kontinent: "Europa", Prozent: "100" },
    Mitglied: { Organisation: "EU", Art: "member" },
    Mitglied: { Organisation: "WEU", Art: "member" }
  }
}
```

SSD-Graph



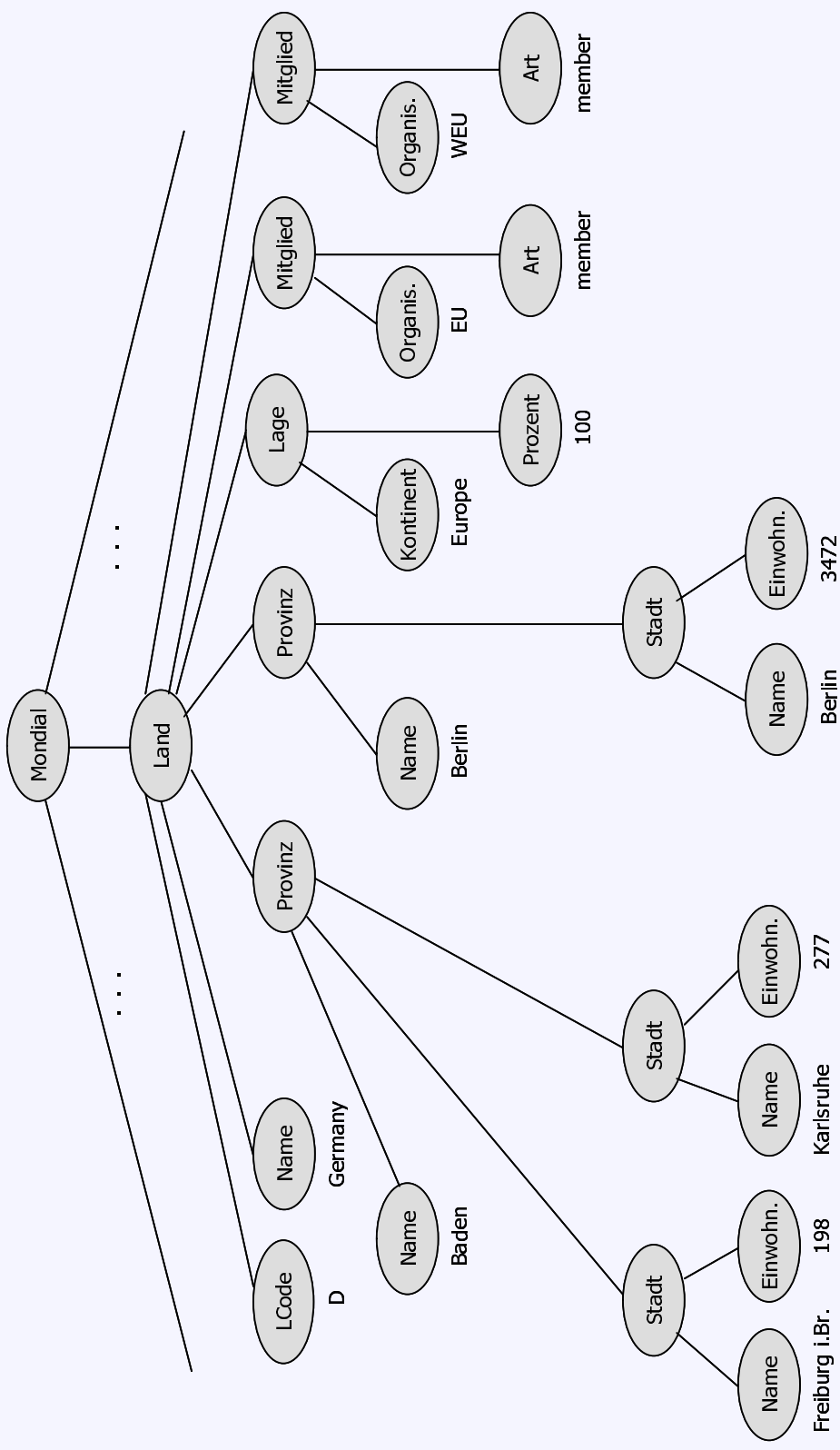
SSD-Schema



XML-Dokument: Variante 1 (lokale Definitionen)

```
<Mondial>
  <Land>
    <LCode>D</LCode>
    <Name>Germany</Name>
    <Provinz>
      <Name>Baden</Name>
      <Stadt>
        <Name>Freiburg i.Br.</Name>
        <Einwohner>198</Einwohner>
      </Stadt>
      <Stadt>
        <Name>Karlsruhe</Name>
        <Einwohner>277</Einwohner>
      </Stadt>
    </Provinz>
    <Provinz>
      <Name>Berlin</Name>
      <Stadt>
        <Name>Berlin</Name>
        <Einwohner>3472</Einwohner>
      </Stadt>
    </Provinz>
  . . .
</Land>
<Kontinent>Europe</Kontinent>
<Prozent>100</Prozent>
</Lage>
<Mitglied>
  <Organisation>EU</Organisation>
  <Art>member</Art>
</Mitglied>
<Mitglied>
  <Organisation>WEU</Organisation>
  <Art>member</Art>
</Mitglied>
</Land>
</Mondial>
```

XML-Baum: Variante 1



DTD: Variante 1

```
<!DOCTYPE Mondial[
  <!ELEMENT Mondial (Land)*>
  <!ELEMENT Land (LCode, Name, Provinz*, Lage*, Mitglied*)>
  <!ELEMENT Provinz (Name, Stadt*)>
  <!ELEMENT Stadt (Name, Einwohner)>
  <!ELEMENT Lage (Kontinent, Prozent)>
  <!ELEMENT Mitglied (Organisation, Art)>
]>
```

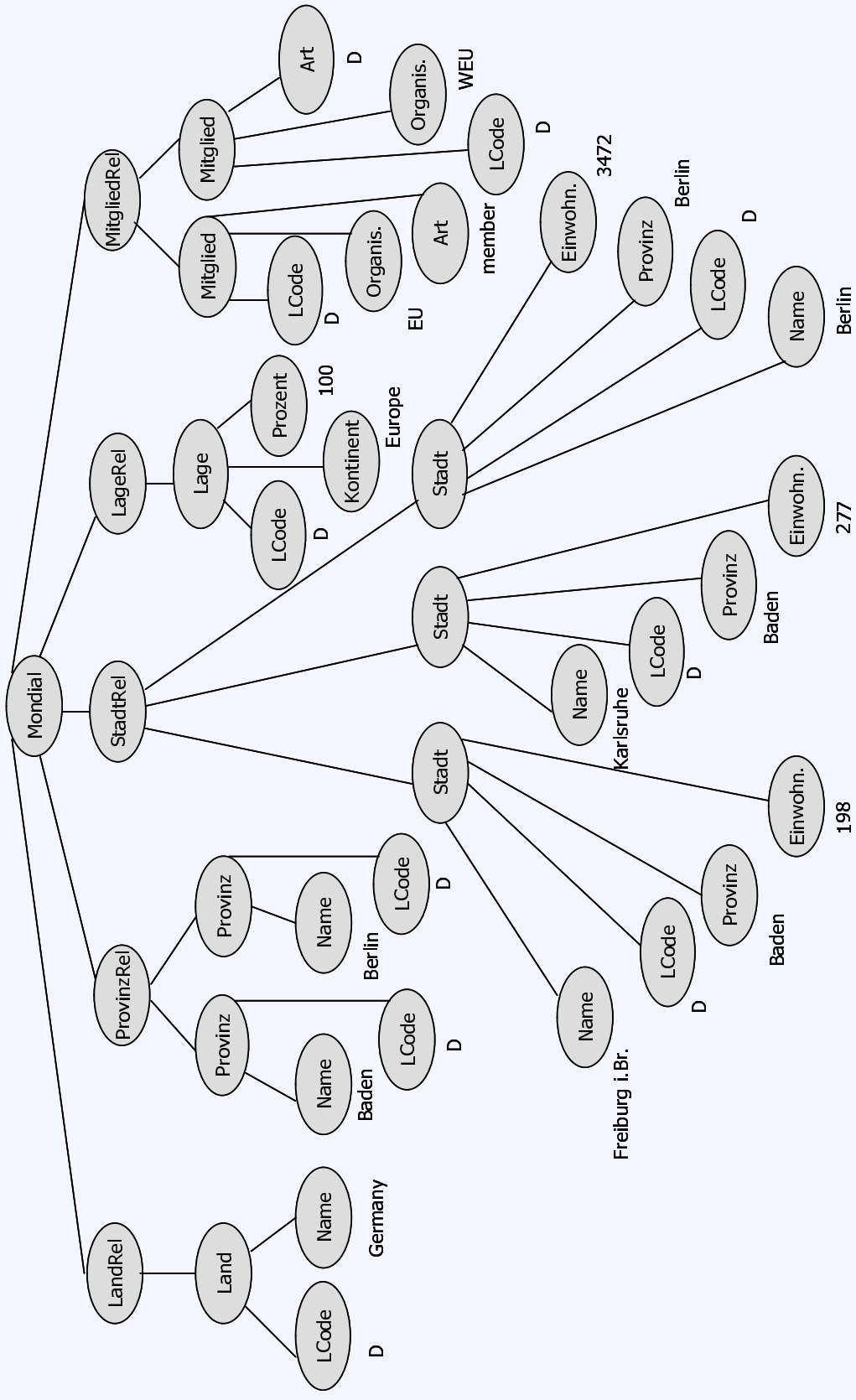
XML-Schema: Variante 1

```
<?xml version = "1.0" encoding = "UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" >
<xs:element name = "Mondial">
<xs:complexType>
<xs:sequence>
<xs:element ref = "Land" minOccurs = "0"
maxOccurs = "unbounded" />
</xs:sequence>
</xs:complexType>
<xs:key name = "PrimaryKeyForLand">
<xs:selector xpath = "Land"/>
<xs:field xpath = "LCode"/>
</xs:key>
</xs:element>
<xs:element name = "Land">
<xs:complexType>
<xs:sequence>
<xs:element name = "LCode" type = "xs:string"/>
<xs:element name = "Name" type = "xs:string"/>
<xs:element ref = "Provinz" minOccurs = "0"
maxOccurs = "unbounded" />
<xs:element ref = "Lage" minOccurs = "0"
maxOccurs = "unbounded" />
<xs:element ref = "Mitglied" minOccurs = "0"
maxOccurs = "unbounded" />
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:element>
<xs:element name = "Provinz">
<xs:complexType>
<xs:sequence>
<xs:element name = "Name" type = "xs:string"/>
<xs:element name = "Stadt" minOccurs = "0"
maxOccurs = "unbounded" />
</xs:sequence>
</xs:complexType>
</xs:element>
<xs:element name = "Stadt">
<xs:complexType>
<xs:sequence>
<xs:element name = "Name" type = "xs:string"/>
<xs:element name = "Einwohner" type = "xs:string"/>
</xs:sequence>
</xs:complexType>
</xs:element>
<xs:element name = "Lage">
<xs:complexType>
<xs:sequence>
<xs:element name = "Kontinent" type = "xs:string"/>
<xs:element name = "Prozent" type = "xs:string"/>
</xs:sequence>
</xs:complexType>
</xs:element>
<xs:element name = "Mitglied">
<xs:complexType>
<xs:sequence>
<xs:element name = "Organisation" type = "xs:string"/>
<xs:element name = "Art" type = "xs:string"/>
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:schema>
```


XML-Dokument: Variante 2 (globale Definitionen)

```
<Mondial>
  <LandRel>
    <Land>
      <LCODE>D</LCODE>
      <Name>Germany</Name>
    </Land>
  </LandRel>
  <ProvRel>
    <Provinz>
      <Name>Baden</Name>
      <LCODE>D</LCODE>
    </Provinz>
    <Provinz>
      <Name>Berlin</Name>
      <LCODE>D</LCODE>
    </Provinz>
  </ProvRel>
  <StadtRel>
    <Stadt>
      <Name>Freiburg i.Br.</Name>
      <LCODE>D</LCODE>
      <Provinz>Baden</Provinz>
      <Einwohner>198</Einwohner>
    </Stadt>
    <Stadt>
      . . .
    </Stadt>
  </StadtRel>
  <Stadt>
    <Name>Berlin</Name>
    <LCODE>D</LCODE>
    <Provinz>Berlin</Provinz>
    <Einwohner>3472</Einwohner>
  </Stadt>
  </StadtRel>
  <LageRel>
    <Lage>
      <Kontinent>Europe</Kontinent>
      <LCODE>D</LCODE>
      <Prozent>100</Prozent>
    </Lage>
  </LageRel>
  <MitgliedRel>
    <Mitglied>
      <Organisation>EU</Organisation>
      <LCODE>D</LCODE>
      <Art>member</Art>
    </Mitglied>
    <Mitglied>
      <Organisation>WEU</Organisation>
      <LCODE>D</LCODE>
      <Art>member</Art>
    </Mitglied>
    <MitgliedRel>
      </Mondial>
    </MitgliedRel>
  </MitgliedRel>
  </Mondial>
```

XML-Baum: Variante 2



DTD: Variante 2

```
<!DOCTYPE Mondial[
  <!ELEMENT Mondial (LandRel, ProvRel, StadtRel, LageRel, MitgliedRel)>
  <!ELEMENT LandRel (Land (Name))*>
  <!ELEMENT ProvRel (Provinz (Name))*>
  <!ELEMENT StadtRel (Stadt (Name, Einwohner))*>
  <!ELEMENT LageRel (Lage (LCode, Kontinent, Prozent))*>
  <!ELEMENT MitgliedRel (Mitglied (Organisation, Art))*>
  <!ATTLIST Land LCode ID #REQUIRED>
  <!ATTLIST Provinz PCode ID #REQUIRED LCode IDREF #REQUIRED>
  <!ATTLIST Stadt PCode IDREF #REQUIRED>
  <!ATTLIST Lage LCode IDREF #REQUIRED>
  <!ATTLIST Mitglied LCode IDREF #REQUIRED>
]>
```

⇒ keine referentielle Integrität!

typische Eigenschaften semistrukturierter Dokumente:

- Die Struktur ist in die Daten eingebettet (*selbstbeschreibende Daten*).
- Die Struktur ist irregulär.
- Die Verwendung von Schemata, bzw. Typen, ist nicht zwingend.
- Die Struktur ist tief geschachtelt, so dass sich zu ihrer Darstellung gerichtete Graphen anbieten.
- Der Zugriff auf die einzelnen Teile eines Dokumentes erfolgt pfadorientiert.