Datenschutz durch maschinenlesbare Zertifizierung mittels XBRL

Michael Lang, Technische Universität München[172], NGCert sowie Christoph Pflügler, Maximilian Schreieck, Dr. Manuel Wiesche, Prof. Dr. Helmut Krcmar, Technische Universität München, ExCELL

Geschäftsprozesse in Unternehmen und im öffentlichen Sektor werden heute in immer komplexeren IT-Landschaften realisiert. Um einen bestimmten Geschäftsprozess umzusetzen, werden oft viele verschiedene Dienste kombiniert. Hierbei ist nicht gewährleistet, dass alle diese Dienste durch dieselbe Organisation bereitgestellt werden – im Zuge von Outsourcing-Maßnahmen bietet es sich an, Dienste externer Anbieter in Anspruch zu nehmen, um Teile ‚eines Geschäftsprozesses zu realisieren. Da hierbei gegebenenfalls sensible Daten weitergegeben werden, sollten in Frage kommende externe Anbieter besonders den Datenschutz analysieren und sicherstellen. Zertifizierungen durch dritte Parteien, wie beispielsweise Regulierungsbehörden, können die Auswahl vertrauenswürdiger Anbieter erheblich erleichtern und außerdem als rechtliche Grundlage für eine Auswahl dienen.[173]

Besonders in komplexen Dienstlandschaften mit vielen externen Anbietern ist jedoch eine manuelle Kontrolle aller Zertifikate nur mit hohem Zeitaufwand möglich. Eine Automatisierung dieses Prozesses könnte durch maschinenlesbare Zertifikate realisiert werden. Diese Zertifikate, kontrolliert durch staatlich akkreditierte Zertifizierungsstellen (vgl. Art. 43 DSGVO), können datenschutztechnische Aspekte, wie den Speicherstandort der Daten oder die verwendete Verschlüsselung, garantieren. Durch automatisierte Routinen können die Zertifikate vor jeder Inanspruchnahme eines externen Dienstes auf ihre Eignung für den jeweiligen Geschäftsprozess überprüft werden. Während eine technische Realisierung solcher Zertifikate bereits möglich ist, existiert bis heute noch kein standardisiertes einsatzfähiges Konzept zur Übermittlung der Zertifikatsinhalte. Dieser Beitrag soll die Grundzüge eines solchen Konzepts mit Hilfe des Datenübertragungsformats XBRL (eXtensible Business Reporting Language) aufzeigen. Das Konzept entstand in einer Zusammenarbeit zwischen dem Projekt ExCELL des Smart-Data-Förderprogramms und dem Projekt NGCert des Förderprogramms „Sicheres Cloud Computing“ des Bundesministeriums für Bildung und Forschung (BMBF).

Als Illustrationsbeispiel wird der physische Speicherstandort verarbeiteter Daten verwendet. Datenschutztechnisch ist dies ein relevantes Thema – so ist es beispielsweise im Zuge des sogenannten Patriot Acts der US-amerikanischen Regierung möglich, in den USA gespeicherte kundenbezogene Daten einzusehen, selbst wenn die Kunden keine US-amerikanischen Staatsbürger sind und sich auch nicht auf US-amerikanischem Territorium aufhalten. Aus diesem Grunde legt ein großer Teil deutscher Unternehmen Wert darauf, dass sensible Daten in Deutschland oder zumindest einem Land der europäischen Union gespeichert werden. Natürlich kann der Datenspeicherstandort allein keine Einhaltung aller unternehmensspezifischen Datenschutzrichtlinien garantieren. Er kann jedoch durchaus die Auswahl eines externen Diensteanbieters beeinflussen – weitere Aspekte, wie beispielsweise die verwendete Verschlüsselung oder Zugriffskontrollen, können analog zu dem illustrierten Konzept umgesetzt werden.

XBRL als Datenübertragungsformat

XBRL ist eine frei verfügbare, XML-basierte Sprache, die einen automatisierten Austausch von Daten mit Hilfe von standardisierten elektronischen Dokumenten ermöglicht. In der Praxis wird XBRL vor allem im Rahmen der Finanzberichterstattung eingesetzt. Die Grundlage aller XBRL-Dokumente bildet die XBRL-Spezifikation.[174] Sie beschreibt die Regeln und die Syntax zur Erstellung XBRL-basierter Artefakte, sogenannter Instanzen und Taxonomien. XBRL-Spezifikationen sind Erweiterungen der XML-Spezifikation, wobei jedes XBRL-Dokument auch ein valides XML-Dokument darstellt. Ein XBRL-Instanzdokument stellt eine Sammlung verschiedener Sachverhalte dar, im Beispiel dieses Artikels also den konkreten Datenspeicherstandort oder

die verwendete Verschlüsselung. Um die Vergleichbarkeit verschiedener Instanzdokumente zu ermöglichen, werden Metadaten zu den verwendeten Sachverhalten benötigt. Diese Metadaten werden in XBRL-Taxonomien gespeichert – hier wird festgelegt, welche Bedeutung den Sachverhalten zukommt, welche konkreten Werte diese annehmen können oder welche Beziehungen zwischen verschiedenen Sachverhalten bestehen. Die Deklaration eines Sachverhalts in einer Taxonomie wird auch als Konzept bezeichnet. Aus technischer Sicht ist eine XBRL-Taxonomie gleichwertig mit einem XML-Schema.

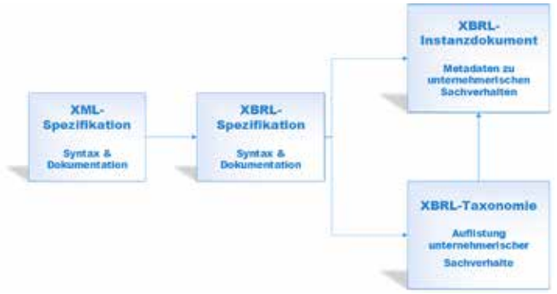

Abbildung 4: XBRL-Aufbau

Code 1, „Beispielhafte XBRL-Taxonomie“, beschreibt eine Taxonomie für den aufgezeigten Anwendungsfall. Aus Gründen der Lesbarkeit wird im Folgenden in allen Code-Beispielen auf die Deklaration der verwendeten Namespaces verzichtet.

Das Element, das den Datenspeicherstandort repräsentiert, trägt den Namen „RechenzentrumStandort“ und wird als eine Sequenz von Subelementen definiert, die die Adresse genauer beschreiben. Durch die Verwendung des Werts „stringItemType“ für das Attribut „type“ wird definiert, dass all diese Elemente durch Zeichenketten repräsentiert werden. Standardmäßig bietet XBRL zudem Unterstützung für Geldeinheiten („monetaryItemType“) sowie Zahlen- und Prozentwerte („decimalItemType“) an. Die Spezifikation erlaubt es Entwicklern außerdem, eigene Datentypen

```
<?xml version=“1.0“ encoding=“US-ASCII“?>
<xs:schema targetNamespace=
„http://www.example.com/ServiceCert“>
<!-- Aus Gründen der Lesbarkeit verzichten wir hier auf die
Deklaration der verwendeten Namespaces -->
 <xs:element name=“RechenzentrumStandort“
  id=“ServiceCert_RechenzentrumStandort“
  substitutionGroup=“xbrli:tuple“
  abstract=“false“>
   <xs:complexType>
    <xs:sequence>
     <xs:element ref=
      „ServiceCert:Street“/>
     <xs:element ref=
      „ServiceCert:BuildingNumber“/>
     <xs:element ref=
      „ServiceCert:PostalCode“/>
     <xs:element ref=
      „ServiceCert:TownCity“/>
     <xs:element ref=
      „ServiceCert:Country“/>
    </xs:sequence>
   </xs:complexType>
 </xs:element>
 <xs:element name=“Street“ id=
  „ServiceCert_Street“
  substitutionGroup=“xbrli:item“
  type=“xbrli:stringItemType“
  xbrli:periodType=“instant“/>
 <xs:element name=“BuildingNumber“ id=
  „ServiceCert_BuildingNumber“
  substitutionGroup=“xbrli:item“
  type=“xbrli:stringItemType“
  xbrli:periodType=“instant“/>
 <xs:element name=“PostalCode“ id=
  „ServiceCert_PostalCode“
  substitutionGroup=“xbrli:item“
  type=“xbrli:stringItemType“
  xbrli:periodType=“instant“/>
 <xs:element name=“TownCity“ id=
  „ServiceCert_TownCity“
  substitutionGroup=“xbrli:item“
  type=“xbrli:stringItemType“
  xbrli:periodType=“instant“/>
 <xs:element name=“Country“ id=
  „ServiceCert_Country“
  substitutionGroup=“xbrli:item“
  type=“xbrli:stringItemType“
  xbrli:periodType=“instant“/>
</xs:schema>
```

Code 1, „Beispielhafte XBRL-Taxonomie“

zu definieren. Das Attribut „periodType“ gibt den zeitlichen Kontext an, für den der Sachverhalt gültig ist – hier kann es sich um einen spezifischen Zeitpunkt („instant“) oder um eine Zeitspanne („duration“) handeln. Die konkreten Zeitdaten werden im Instanzdokument festgelegt.

Auf Basis dieser Taxonomie kann anschließend ein Instanzdokument erstellt werden. Das vollständige Ergebnis ist dem Code 2, „Beispielhaftes XBRL-Instanzdokument“, zu entnehmen. Auch an dieser Stelle wird wieder auf die Angabe der verwendeten Namespaces verzichtet.

```
<?xml version=“1.0“ encoding=“UTF-8“?>
<!-- Aus Gründen der Lesbarkeit verzichten wir hier auf die
Deklaration der verwendeten Namespaces -->
<xbrli:xbrl>
 <link:schemaRef xlink:type=“simple“
  xlink:href=“ServiceCert.xsd“/>
 <ServiceCert:RechenzentrumStandort>
  <ServiceCert:Street contextRef=
   „Anbieter1“>Cloudstrasse</ServiceCert:Street>
  <ServiceCert:BuildingNumber contextRef=
   „Anbieter1“>1</ServiceCert:BuildingNumber>
  <ServiceCert:PostalCode contextRef=
   „Anbieter1“>80331</ServiceCert:PostalCode>
  <ServiceCert:TownCity contextRef=
   „Anbieter1“>Muenchen</ServiceCert:TownCity>
  <ServiceCert:Country contextRef=
   „Anbieter1“>Germany</ServiceCert:Country>
 </ServiceCert:RechenzentrumStandort>
 <xbrli:context id=“Anbieter1“>
  <xbrli:entity>
   <xbrli:identifier scheme=
    „http://www.services.com“>
    Service Provider 1
   </xbrli:identifier>
  </xbrli:entity>
  <xbrli:period>
   <xbrli:instant>2016-01-01</xbrli:instant>
  </xbrli:period>
 </xbrli:context>
</xbrli:xbrl>
```

Code 2, „Beispielhaftes XBRL-Instanzdokument“

Zunächst wird auf die zuvor definierte Taxonomie referenziert und das dort bereitgestellte Konzept mitsamt seinen Unterkonzepten instanziiert. Über XBRL-spezifische Validierungssoftware kann jederzeit sichergestellt werden, dass das Instanzdokument den Regeln der Taxonomie folgt und beispielsweise die korrekten Datentypen verwendet werden. Im Instanzdokument wird zudem ein sogenannter Context definiert, der mit dem RechenzentrumStandort-Element verknüpft wird, um weitere Metadaten anzugeben. Hierzu gehört eine Entity, eine Organisation oder ein Individuum, auf die sich das Element bezieht und die in diesem Beispiel den Diensteanbieter darstellt. Außerdem kann der bereits erwähnte Zeitraum angegeben werden, für den der berichtete Sachverhalt gültig ist. Wie bereits in der Taxonomie definiert, handelt es sich hier um einen spezifischen Zeitpunkt, nämlich den 1. Januar 2016.

Um weitere Informationen zu unseren bereits definierten Konzepten bereitzustellen oder um Beziehungen zwischen verschiedenen Konzepten zu definieren, bietet XBRL sogenannte Linkbases an. Hierbei handelt es sich um zusätzliche XML-Dateien, die der XLink-Spezifikation folgen und in fünf Kategorien unterteilt werden: Label, Definition, Presentation, Reference und Calculation. Über die Label-Linkbase können menschenlesbare Zeichenketten als Bezeichner für bestimmte Konzepte definiert werden. Diese Bezeichner können dann in grafischen Oberflächen angezeigt werden – hiermit lassen sich Konzepte auch internationalisieren, da je Sprache verschiedene Bezeichner gewählt werden können. Mit Hilfe der Definition-Linkbase können verschiedene Beziehungen zwischen jeweils zwei Konzepten erstellt werden. Hierzu gehören hierarchische Strukturen (Parent-Child-Beziehungen) oder auch Spezialisierungen bzw. Generalisierungen. In der zuvor dargestellten Taxonomie wird beispielsweise das generische Konzept „Postal Code“ verwendet. Eine mögliche Spezialisierung für den geografischen Standort Deutschland könnte ein neues Element „Postleitzahl“ darstellen, für das wiederum besondere Validierungsregeln gelten könnten. Über die Reference-Linkbase lassen sich Verweise auf relevante Gesetzestexte oder Kommentare in externen Dokumenten wie Internetseiten oder Gesetzbüchern hinterlegen. Presentation-Linkbases beziehen sich ähnlich wie Labels auf die grafische Darstellung der Elemente. Hier können hierarchische Strukturen für die verwendeten Elemente definiert werden, die dann in grafischen Oberflächen für die Darstellung der Instanzdokumente verwendet werden können. Calculation-Linkbases zielen auf Anforderungen der Finanzbranche ab. Sie definieren vereinfacht gesagt Rechenregeln zwischen verschiedenen monetären Elementen und werden hier nicht genauer erläutert.

Das Instanzdokument kann anschließend zusammen mit der verwendeten Taxonomie sowie den gegebenenfalls benötigten Linkbase-Dateien einer staatlich akkreditierten Zertifizierungsstelle signiert und vom dazugehörigen Diensteanbieter bereitgestellt werden. Eine automatisierte Prüfung der Informationen oder ein Vergleich mit anderen Anbietern kann durch spezialisierte Software erfolgen. Auch eine manuelle Einsicht oder Kontrolle der übertragenen Daten wird durch entsprechende Software ermöglicht – da die grafische Repräsentation allein durch die standardisierten Label- und Presentation-Linkbases definiert wird, lässt sich insbesondere für diesen Anwendungsfall bereits existierende XBRL-Software verwenden.

Fazit und Ausblick

Die Eigenschaften und Konzepte der Sprache XBRL eignen sich hervorragend als Grundlage für die Übertragung maschinenlesbarer Zertifikate zum automatischen Abgleich servicerelevanter Anforderungen. Taxonomien, die von unabhängigen, vertrauenswürdigen Instanzen wie Regulierungsbehörden erstellt werden können, sorgen für Konsistenz und durchgehende Validierbarkeit der übertragenen Zertifikatsdaten und stellen somit auch die Grundlage für maschinelle Auswertung und Weiterverarbeitung dar. Auch wenn eine solche Standardisierung heute noch aussteht, würde sie die Entwicklung von Analyse- und Vergleichssoftware begünstigen. Da Zertifikate verschiedener Anbieter denselben Taxonomien folgen, können diese einfach und automatisiert miteinander verglichen werden. Bestehende Taxonomien können außerdem für Spezialfälle um zusätzliche Konzepte erweitert werden.

Durch die weite Verbreitung von XBRL im Bereich der Finanzberichterstattung existieren bereits viele Programme zur Verarbeitung von XBRL-Dokumenten. Gerade im Bereich der grafischen Darstellung gibt es hier ein großes Potenzial für eine Wiederverwendung. Aber auch bei Neuentwicklungen kann von der Popularität des XBRL durch die Verwendung von bereits existierenden und praxiserprobten Programmierschnittstellen profitiert werden. Für die Konzeption und die Entwicklung eines Bereitstellungsmechanismus für signierte Instanzdokumente und Taxonomien lohnt sich ein Blick auf das Online-Portal EDGAR Online.[175] Über EDGAR Online lassen sich Finanzberichte vieler verschiedener Unternehmen im XBRL-Format suchen und herunterladen. Eine ähnliche Plattform wäre für unsere XBRL-basierten Zertifikate denkbar – Benutzer hätten damit eine zentrale Anlaufstelle zur Verfügung, die sie für die Suche nach einem geeigneten und zertifizierten Diensteanbieter verwenden könnten.

Eine der größten Herausforderungen auf organisatorischer Ebene stellen sicherlich die Entwicklung und Einigung auf einen XBRL-basierten Datenübertragungsstandard dar, der einheitliche Taxonomien und Linkbase-Definitionen einschließt. Hier wird es wichtig sein, Diensteanbieter und Dienstekonsumenten, aber auch den Gesetzesgeber, aktiv in Entscheidungen und Diskussionen einzubeziehen, um einen übersichtlichen und nachhaltigen Datenübertragungsstandard zu entwickeln, der die Anforderungen und Bedürfnisse aller beteiligten Nutzer abdeckt.