

# **Dr. Christian Galinski: "Fragen der semantischen Interoperabilität brechen jetzt überall auf."**



**Terminologienormung ist eine Grundbedingung für internationalen Handel und birgt große Einsparungspotenziale in der industriellen Produktion. Tassilo Pellegrini sprach mit Dr. Christian Galinski, Leiter des Internationalen Informationszentrums für Terminologie (Infoterm), über Probleme der Mehrsprachigkeit, die Wechselwirkung zwischen Technik und Terminologie und wie durch semantische Interoperabilität die Schnittstelle Mensch-Maschine verbessert werden kann.**

## **Was ist der Hintergrund der ISO-Zertifizierung im Terminologiebereich?**

Die Tradition der Terminologienormung reicht zurück in das Jahr 1904/05. Damals wurde die internationale elektrotechnische Kommission (IEC) zu Zwecken der Hardwarenormierung, wie Kaliber und Gewinde, gegründet. Man erkannte schnell, dass auch Terminologienormung eine wichtige Rolle spielt, denn normen kann man nur dann, wenn man weiß, worüber man redet. So begann die Normung terminologischer Informationen wie Benennungen, Definitionen etc.

In den 1930-Jahren wurde offensichtlich, dass auch die Methodik genormt werden muss, da sonst in den einzelnen Standardisierungsinstitutionen Terminologien entstehen, die nicht miteinander kompatibel sind. Unter Federführung Österreichs wurde so der Bereich der Methodiknormung im Technischen Komitee 37 der ISO (ISO/TC 37) institutionalisiert und ist noch heute dort untergebracht.

Nach dem 2. Weltkrieg gab es einen Wandel in den Schwerpunkten der Normung, von der Hardware zur Software, wobei der Softwarebegriff damals auf Phänomene wie Management, Qualitätssicherung und frühe EDV Software Anwendung fand.

Und in den letzten Monaten verlagert sich der Schwerpunkt wieder zunehmend in Richtung der Methodiknormung von Industriedaten. Das Ziel ist, Terminologien und Klassifikationen sach- und sprachunabhängig festzulegen und so Produkte und Dienstleistungen in beliebig vielen Sachbereichen und Sprachen beschreiben zu können.

## **Mehrsprachigkeit ist ein Schlüssel zu internationalen Märkten, aber gleichzeitig auch eine Barriere. Wie schafft die Terminologienormung Abhilfe?**

Fragen der Mehrsprachigkeit und der internationalen Terminologie brechen jetzt überall auf. Die globalisierte Wirtschaft muss ihre Produkte vor Ort verkaufen und in einen lokalen Kontext einbetten. Amerikanische Produkte und Dienstleistungen können in China nicht auf

Englisch verkauft werden. Das ist nicht nur ein sprachliches Problem, sondern reicht bis in Sphären von Haftungs- und Ausbildungsfragen.

Hier sind Organisationen wie die ISO/TC37 gefragt. Die Regeln, die wir erstellen, gelten für alle technischen Komitees der ISO, aber auch für das Internationale Elektrotechnische Komitee (IEC) und für viele andere normenschaffende Organisationen, wie die internationale Luftfahrtorganisation (ICAO), die Internationale Telekommunikations Union (ITU) etc. In der ICAO geht es zum Beispiel darum, den Sprachgebrauch bis zur Abfolge phraseologischer Einheiten, z.B. in der Kommunikation zwischen zwei Flugzeugen, festzulegen.

In der ISO/TC 37 haben wir herausgefunden, dass es durchwegs verschiedene Formate und Datenmodelle geben kann, um Kommunikationsprozesse zu beschreiben und zu standardisieren. Die Kommunikationsprozesse funktionieren untereinander allerdings nur, wenn es ein Metamodell gibt. Und dieses Problem taucht jetzt überall in der Wirtschaft und Verwaltung auf. D.h. es ist notwendig Metadaten zu beschreiben, um das Problem der Interoperabilität in den Griff zu bekommen.

### **Was genau ist unter semantischer Interoperabilität zu verstehen?**

Der Begriff der semantischen Interoperabilität ist aufgetreten mit dem Semantic Web, einer Konzeption von Tim Berners-Lee, der sagt, das zunehmend die Computer miteinander über hochstandardisierte Sprachen, die wenig mit Natürlichsprachlichkeit zu tun haben, kommunizieren werden. Was er nicht sieht, ist dass rein technische Interoperabilität nicht ausreicht, um die semantische Interoperabilität herzustellen.

Auf Basis einer Analyse, die im Rahmen des EU Interoperability Framework angestellt wurde, ist man zu dem Ergebnis gekommen, dass Interoperabilität auf den Ebenen Technik, Semantik und Organisation gegeben sein muss. Die Vernachlässigung der organisatorischen Interoperabilität hat z.B. dazu geführt, dass zwei Organisationen zwar auf Basis der selben Rahmenbedingungen und Standards semantische Einheiten schaffen, die zugrunde liegenden Datenmodelle aber nicht miteinander kompatibel sind.

In einem weiteren Schritt hat man die semantische Interoperabilität unterteilt in eine syntaktische, pragmatische und begriffliche Ebene. Informatiker beschränken sich meist auf die syntaktische Ebene. Aber Daten müssen auch auf ihrer begrifflichen Ebene und der Ebene von Datenkategorien so aufbereitet sein, dass sie zwischen technischen Plattformen ausgetauscht werden können, und die Weiterverwertbarkeit gesichert ist. Dazu ist die pragmatische Interoperabilität wichtig, d.h. die Repräsentation der Begriffe durch sprachliche oder nichtsprachliche Symbole, die zum Teil sehr stark kulturabhängig ist. Damit meine ich nicht nur Kultur von Menschen sondern auch Firmenkulturen.

Um Ihnen ein Beispiel zu geben: in der Softwareproduktion gibt es das Cultural Diversity Management, wo es um Datumsschreibweise und Zeichensatzkodierung geht. Hier gibt es enge Anknüpfungspunkte zu der Frage, wie internationale Produkt beschaffen sein müssen, um in lokalen Märkten angeboten und vermarktet werden zu können.

### **Was ist der Status Quo der Standardisierungsbestrebungen?**

Wir werden mit Metamodellen Leben müssen. Möglicherweise mit ganzen Hierarchien von Metamodellen. In der ISO/TC 37 wurde eine sehr schöne Modelllösung gefunden, die so neutral und generisch ist, dass sie z.B. auf unterschiedlichste Produktdaten und – klassifikationen anwendbar ist. Wir haben uns im Rahmen eines Memorandum of

Understanding, wo es um Normungsaktivitäten auf dem Gebiet des elektronischen Handels geht, mit der ISO, dem IEC, der ITU und der Europäischen Wirtschaftskommission der Vereinten Nationen auf dieses Metamodell geeinigt.

Die Standardisierung im Bereich der Methodik ist sehr weit gediegen. Es gibt über 300 Datenkategorien, die auftreten können, aber nicht auftreten müssen. Diese sind mit einer sehr hohen Detailtiefe beschrieben. Es gibt ein Metamodell mit dem Titel „Terminology Markup Framework“. Mit diesem können heterogene Datenmodelle messbar und interoperabel gemacht werden. Messbar heißt, dass man ziemlich genau identifizieren kann, wo bei den einzelnen Datenmodellen menschlich nachgebessert werden muss, um eine 100-prozentige Interoperabilität herzustellen.

### **Welchen Nutzen zieht man daraus?**

Am Beispiel der Stahlindustrie könnte das Szenario folgendermaßen aussehen. Es gibt hunderte Firmen mit hunderten Abteilungen, die weltweit Stahl erzeugen und verarbeiten. Innerhalb der Firmen gibt es Entwicklungsabteilungen, die benutzen eine wissenschaftlich-technische Terminologie mit meist sehr langen Benennungen. Diese Benennungen sind im Bereich der Werkstattsprache nicht praktikabel, da erstens nicht verlangt werden kann, dass die Arbeiter diese Terminologie verstehen, und zweitens diese „unhandliche“ Terminologie im Produktionskontext aufgrund von Missverständnissen eine Gefahr für Leib und Leben darstellen kann.

Man sieht, dass schon innerhalb einer Firma für die gleichen Dinge berechtigterweise unterschiedliche Benennungen verwendet, die sich je nach Kontext - ob in technischen Beschreibungen, in Marketingunterlagen, in Angeboten, in Verträgen, in Bauplänen, in Ausbildungsaktivitäten - erheblich unterscheiden.

Hier summieren sich ganz gewaltige Verluste, die man mit bisherigen Modellen gar nicht greifbar machen kann. In einer Schätzung von J.D. Edwards wurde berechnet, dass durch die Umsetzung von Content-Management-Prinzipien in der schriftlichen und mündlichen Kommunikation einer Firma bei fünf Sprachen, die im Verkauf verwendet werden, Ersparnisse pro Begriffseintrag von 100 Dollar pro Jahr realisiert werden können. D.h. bei 1000 Begriffen ergibt sich eine Einsparung von 100.000 Dollar pro Jahr.

Rationalisierungseffekte sind z.B. in der Lagerverwaltung, dem Einkauf, dem Verkauf zu erzielen. Ein wichtiges Anwendungsfeld ist aber auch die Mobilkommunikation und das mobile Computing und darauf abgestimmte Inhalte – der mContent. Die hier stattfindende vierte Welle in der Entwicklung der globalen Kommunikations- und Informationstechnologien geht einher mit viel größeren Anforderungen an die semantische Interoperabilität als wir sie bisher kannten.

### **Technik und Terminologie stehen also in einem engen Wechselverhältnis?**

Ja natürlich! Mobile Content wird beeinflusst durch die Vorgaben der Nutzerschnittstellen. Die Tastaturen werden immer kleiner, zunehmend wird die menschliche Sprache als Eingabemittel genutzt. Der Computer muss diese Signale in geschriebene Sprache umsetzen, ohne Zweideutigkeiten zu erzeugen. Schließlich müssen nachfolgende Aktionen wie Bestellungen und Bezahlung fehlerfrei verlaufen.

Die Relevanz dieser Entwicklungen für den Nutzer ergibt sich dort, wo es einmal möglich sein wird z.B. über das Handy einzukaufen. D.h. man ruft im Supermarkt an und bestellt

Butter, Nudeln, Mehl, Obst, Eier in definierten Mengen. Über Rückfragesysteme wird festgestellt, welche der 10 verschiedenen Buttersorten der Kunde haben will. So kann interaktiv in gesprochener Sprache eine Bestellung aufgegeben werden, die Abbuchung findet automatisch statt, der Parkschein ist aufgrund des kurzen Abholvorganges nicht notwendig.

Die Voraussetzungen dieses Szenarios sind, dass die Butter durch einen Strichcode kodiert ist, sich hinter dem Strichcode ein weltweites System von Produktkodierungen verbirgt, und man ist jetzt schon dabei das einzelne Butterpaket zu kodieren, um so die Lagerhaltungskosten zu optimieren. Hier sieht man, wie sehr industrielle Vorgänge mit der Benutzerebene verzahnt sind.

## **Wie kommt man zu eindeutigen Produktdefinitionen?**

Zur Verdeutlichung ein Beispiel aus Deutschland. Auf der Erkenntnis, dass viele Produkte und Dienstleistungen mit gleichen Merkmalen bedacht werden, die zum Teil branchenübergreifend, zum Teil branchenspezifisch sind, wurde vom Deutschen Institut für Normung (DIN) ein Merkmalsserver erstellt. Wenn Produktmerkmale nicht gleich definiert sind, kann der Anschein entstehen, dass es sich um andere Produkte handelt. Dies würde zum Beispiel im E-Government die elektronische Ausschreibung stark behindern und erhebliche Nachteile für Firmen erzeugen, die nicht auf harmonisierte Daten zurückgreifen können.

Beim DIN wurden unter Rückgriff auf den alten Normungsansatz der sog. Merkmalsleisten Produkte und deren Merkmalsausprägungen genau analysiert und abgeleitet, wie man Ausprägungen eindeutig erfassen und branchenübergreifend den Firmen zur Verfügung stellen kann.

Der DIN-Merkmalsserver ist damit ein Modell, wie man auf der Ebene der kleinsten Contenteinheiten die semantische Interoperabilität von digitalisierten Fachinformationen (ausgeklammert ist hier der digitale Content im Sinne der Contentindustrien wie Video, Musik, Computerspiele etc. – Anm. d. A.) sicherstellen kann. Es existieren derzeit mehrere hundertmillionen Produkte. Jedes Produkt kann durch seine Merkmale beschrieben werden, und jedes Merkmal als Begriff numerisch, alphanumerisch, oder grafisch kodiert werden. Aufgrund der gewaltigen Mengen von Contenteinheiten, die zwischen den PCs kommuniziert werden, sind Repositorien, wie der Merkmalsserver, notwendig, wo einzelne Gruppierungen in der Wirtschaft für die Qualität der Daten verantwortlich zeichnen, bestimmte Typen von Content warten und der gesamten Wirtschaft zur Verfügung stellen.

Diese Modell geht in Richtung open content, d.h. allgemein zugängliche, nicht proprietäre Information, wobei es sich - wie bei open source - nicht zwangsläufig um kostenlose Information handelt. In diese Richtung scheint semantische Interoperabilität verwirklichtbar zu sein.

## **Zur Person**

Der Japanologe, Dr. Christian Galinski, ist seit 1986 Leiter des Internationalen Informationszentrums für Terminologie (Infoterm), eine Non-Profit-Organisation zu Zwecken der internationalen Terminologienormierung. Er hat eine Reihe von Initiativen und Netzwerken in Ostasien und Europa aufgebaut, unter anderem auch Termnet, ein internationales Terminologienetzwerk, das weltweit als Koordinationsstelle für Firmen agiert, die auf dem Gebiet des mehrsprachigen Contents aktiv sind.