

Geodatenstandards

www.business-geomatics.com

PETER BAUMANN über die Entwicklung von Geodatenstandards.

„Die Zukunft gehört sprachbasierten Schnittstellen“

Die Interoperabilität von Geodaten ist eines der meist diskutierten und komplexesten Themen in der Geoinformationswirtschaft. Vernetzte Geschäftsprozesse, Geodateninfrastrukturen oder auch die intensivere Nutzung von Geodaten durch fachfremde Mitarbeiter erhöht den Bedarf. Grundlegende Voraussetzung für Interoperabilität sind funktionierende Normen und Standards. Maßgeblicher Treiber der weltweiten Standardisierung ist das internationale Open Geospatial Consortium (OGC). Dessen Aufgabe ist es, die notwendigen Rahmenbedingungen zu schaffen, um Datenaustausch und Datenzugang für bessere Informationsgewinnung zu optimieren. Business Geomatics sprach mit Professor Peter Baumann, der kürzlich vom OGC für seinen herausragenden Beitrag in der Standardisierung von raumbezogenen Daten ausgezeichnet worden ist.

Herr Baumann, wozu braucht man überhaupt Standards in der Geoinformationswirtschaft?

Tja, wozu überhaupt Standards? Manchmal empfindet man sie ja als Einschränkung – warum soll ich das genau so machen und nicht anders? Aber Standards umgeben uns überall und ermöglichen erst das Miteinander unserer vernetzten Gesellschaft. Stecker passen in Steckdosen, Musik in HiFi-Anlagen, Autos in Garagen, und Geodaten können von einem GIS ins andere fließen. Das ermöglicht den freien Fluss der Information und macht uns unabhängig von einzelnen Werkzeugen und Herstellern.

Sie sind Editor mehrerer OGC Standards. Was sind die Unterschiede der verschiedenen Standardisierungsgremien, was ihre jeweiligen Anforderungen betrifft?

ISO ist ja sozusagen die Mutter aller Standardisierungsgremien, dort werden alle möglichen Dinge international standardisiert, von IT über Qualitätsmanagement bis hin zu sozialer Verantwortung. OGC hingegen hat die Spezialexpertise im Geo-Bereich und arbeitet eng mit dem Gegenpart, der ISO TC 211 zusammen. Oft werden OGC-Spezifikationen von ISO übernommen. Das war so mit GML und WMS und findet derzeit mit dem OGC Coverage-Modell und dem OGC Web Coverage Service statt. Einen wichtigen Unterschied zwischen beiden Gremien sehe ich im praktischen Aussehen von Spezifikationen: Während OGC mehr und mehr das Existieren mindestens einer Implementierung, am besten Open Source, als Vorbedingung für einen neuen Standard macht, ist dies in ISO nicht gefordert. Auch ist in OGC das Vorhandensein von Konformanz-Tests obligatorisch, im Gegensatz zu ISO.

Besteht die Gefahr, dass es zu viele Standards für ein und dieselbe Sache gibt und die Standards nicht miteinander „kompatibel“ sind?

Aber ja, das ist eine ständige Herausforderung. Ein respektable Teil unserer Arbeiten geht in die sogenannte Harmonisierung von Spezifikationen. Bei neuen Standards versuchen wir Einfluss zu nehmen, um sie mit den existierenden Standards verträglich zu machen. Ein Beispiel ist der gerade entstehende ISO 19163 für Sa-



Über Peter Baumann

Dr. Peter Baumann ist Professor für Informatik an der Jacobs University Bremen sowie Erfinder und Chefarchitekt von rasdaman. Im Herbst 2014 wurde er vom Open Geospatial Consortium (OGC) mit dem Kenneth D. Gardels Award für seinen herausragenden Beitrag in der IT-Standardisierung ausgezeichnet. Der Computerwissenschaftler entwickelte in den vergangenen zehn Jahren mehr als zwölf Standards für raumzeitliche Daten. Baumann ist Editor der OGC Web Coverage Service (WCS) Suite und Gründer der Big-Data-Arbeitsgruppe (BigData.DWG) im OGC. Er konzeptionierte den neuen SQL ISO Standard MDA („Multi-Dimensional Arrays“). Der Award ist nach dem 1999 verstorbenen Wissenschaftler Kenneth D. Gardels benannt, der am Center für Environmental Design Research an der University of California in Berkeley arbeitete.

tellitendaten; hier finden intensive Gespräche statt, um Kompatibilität mit den einschlägigen existierenden OGC-Standards zu erreichen. Manchmal müssen auch bestehende Standards miteinander abgeglichen werden, wenn sich aufgrund erweiterter Einsatzgebiete plötzlich Überlappungen ergeben. Ein Beispiel ist der OGC-Standard WaterML; dort werden bisher Zeitreihen von einzelnen Messdaten verwaltet, mit WaterML 2.0 sind jedoch auch Zeitreihen von Bildern plötzlich im Fokus, allerdings sind die bisherigen 1D-Methoden in 3D nicht effizient. Hier ist noch Abstimmungsarbeit mit dem Coverage-Modell des OGC zu leisten, welches bereits seit 2010 raumzeitliche Datenwürfel wirkungsvoll darstellen kann.

Die Anforderungen verändern sich im Laufe der Zeit. Welche Ihrer Standardspezifikationen halten Sie für besonders wegweisend oder zukunftssträftig und warum?

Die Zukunft bei Webdiensten gehört eindeutig sprachbasierten Schnittstellen. Nicht natürlich sprachlichen, sondern kompakten, von Datenrepräsentationen unabhängigen Abfragesprachen.

Warum ist das von Vorteil?

Dienste mit einer Menge von einzelnen Funktionsaufrufen lassen sich nicht zu einem einzigen mehrschrittigen Aufruf kombinieren – das ist ineffizient. Noch weniger können solche Dienste sich automatisch die Arbeit teilen, also „orchestrieren“. Das muss nicht so sein. Im administrativ-betriebswirtschaftlichen Bereich hat uns SQL vorgemacht, wie leistungsstarke, schnelle und dabei flexible Dienste aussehen können. Im OGC

finden sich bereits die Geo-Pendants: GeoSPARQL für semantische Suche, Filter Encoding für die Suche auf Features sowie Web Coverage Processing Service (WCPS) für Prozessierung und Suche auf raumzeitlichen Rasterdaten. Unsere WCPS-Implementierung rasdaman beispielsweise läuft bei der ESA auf einer mehr als 130 Terabyte

noch wünschenswert. Zwei Grenzen sehe ich: Erstens, wo alles gut läuft und somit kein Bedarf besteht, sollte auch keine Regel aufgestellt werden. Zweitens, Standards sollten Schnittstellen festlegen, damit Kommunikation und Austausch stattfinden kann – sie sollten jedoch keinesfalls bestimmte Technologien und Algorithmen zementieren, denn schon morgen kann eine neue Idee das gleiche Problem viel besser lösen.

Sie heben immer wieder die großen Herausforderungen hervor, die Big Data mit sich bringt. Eine davon ist „Variety“, die Verschiedenartigkeit der Daten. Können Standards hier helfen, heterogenes Datenmaterial kombinierbar zu machen oder den Umgang mit Big Data in der täglichen Arbeitsroutine zu erleichtern?

Sie sprechen hier die gängigen vier Vs an, welche zur Definition von „Big Data“ herangezogen werden: Volume, Velocity, Variety und Veracity. Man kann etwas locker sagen, dass der technologische Fortschritt entlang dieser Liste verläuft. Um auf Ihre Frage einzugehen: Oftmals ist tatsächlich nicht das schiere Volumen, sondern die Vielfalt und Verschiedenartigkeit der Daten das größere Problem. Das haben mir die Explorationsindustrie und auch die INSPIRE-Experten immer wieder bestätigt. Standards können hier beitragen, indem sie übergreifende, allgemeingültige Konzepte bereitstellen, so dass jeder Spezialfall weiß, wie er sich eingliedern muss. Idealerweise – das ist die Vision – sieht der Anwender irgendwann gar nicht mehr, welches Datenformat und welchen Dienst er benutzt. Beispielsweise ist das Coverage-Konzept im OGC abstrakt definiert und konkrete Coverage-Objekte können in prinzipiell jedem geeigneten Austauschformat codiert werden. Es muss lediglich für jedes dieser Formate einmal festgelegt werden, wie die einzelnen Bestandteile – etwa Georeferenzierung und „Pixel“-Datentyp – abzubilden sind. Bereits mehrfach haben wir darin unterstützt, Werkzeuge OGC-konform zu machen. Implementierer haben dem Coverage-Modell regelmäßig eine besonders klare und einfache Umsetzbarkeit attestiert.

Welche Disziplinen der Geodatenwelt profitieren besonders davon?

Vor allem diejenigen, in denen massiver Austausch zwischen den Welten stattfinden muss. Das Paradebeispiel ist Katastrophenmanagement, wo Satellitendaten, Katasterdaten, InSitu-Messungen, Crowdsourcing, Wetterdaten et cetera schnell und zuverlässig zusammengeführt werden müssen. Allerdings ist in sämtlichen Bereichen eine zunehmende Vernetzung zu beobachten, etwa bei den bisher eher isolierten marinen Daten. Insofern wird Standardisierung sukzessive für immer mehr Disziplinen ein wichtiges Thema.

Sie sind Chefarchitekt des Big Data Servers rasdaman. Inwieweit fließen die Ergebnisse Ihrer Standardisierungsarbeit in diese Software, die ja Open Source ist, ein?

Eigentlich ist das ein Wechselspiel. Die „Big Data“-Standards für raumzeitliche Datenwürfel basieren auf den Konzepten und dem internationalen praktischen Einsatz von rasdaman. Andererseits implementieren wir neue Konzepte in rasdaman, bevor sie als Standard fest gefahren werden; damit können wir auch sicher sein, dass der neue Standard umsetzbar und praxistauglich sein wird. Ein „Schreiben ohne Implementieren“ wäre mir sehr suspekt.

Wenn Open Source- und Lizenzprodukte durch Standardisierung irgendwann mühelos miteinander vereinbar wären – welche Auswirkungen könnte dies auf die zukünftige Entwicklung von Geosoftwares insgesamt haben?

Das ist eine wunderbar visionäre Frage! Allzuoft beobachte ich heute einen tiefen Graben – kommerzielle Anbieter gehen auf Distanz zu quelloffenen Werkzeugen, und Open Source-Gemeinden stehen einer Kommerzialisierung misstrauisch gegenüber. Eigentlich sollten wir aber nicht nur diese beiden „reinen“ Modelle, sondern auch beliebig gemischte Geschäftsmodelle in Betracht ziehen, beispielsweise duale Lizenzmodelle. Im Endeffekt gibt das den Nutzern maximale Freiheit, für ihren speziellen Fall das beste passende Modell zu finden. Wenn wir erst die Ressentiments auf beiden Seiten überkommen, ergeben sich hier spannende Perspektiven.

www.rasdaman.de

arte

FERNSEHTIP FÜR GEO-EXPERTEN

Am 6. Februar um 21.45 Uhr strahlt der Sender Arte die Dokumentation „Die digitalisierte Erde“, aus. Bei der Sendung wirkte Peter Baumann als wissenschaftlicher Leiter mit. Die einstündige Dokumentation

thematisiert Chancen und Herausforderungen, die im Umgang mit den großen Datenmengen von Fernerkundung und Erdbeobachtung liegen und zeigt auf, wie die Daten verarbeitet und aus ihnen sinnvolle Informa-

tionen abgeleitet werden können. Wissenschaftler wie Olaf Trieschmann, der mit Satelliten die Meere überwacht, oder Elisabeth Schöpfer, die die Entwicklung von Flüchtlingslagern analysiert, kommen dabei zu Wort.