

METHODEN UND ERFAHRUNGEN DER QUALITÄTSSICHERUNG VON UMWELTMONITORINGDATEN IN AL.VIS|TIMESERIES

Matthias Haase, Markus Hillmann, Mathias Fritz
WISUTEC Umwelttechnik GmbH, Jagdschänkenstraße 50, 09117 Chemnitz

Im Zuge der Digitalisierung werden stetig ansteigende Mengen an Geodaten und Messwerten kontinuierlich erfasst, dessen Management Aufgabe von Umweltinformationssystemen ist. Die Qualitätssicherung dieser Daten und Messwerte ist hierbei von entscheidender Bedeutung. Dabei ist es unverzichtbar, den Weg eines digitalen Messwertes von der Erfassung bis zum Endnutzer, über verschiedene Systeme und Komponenten hinweg, zu betrachten. Entlang dieses Pfades müssen Qualitätsinformationen an Messwerten transportiert, gespeichert und wenn möglich automatisiert erfasst werden. Diese Prinzipien und Komponenten werden am Beispiel der Software AL.VIS|Timeseries, die auf die Verwaltung von Zeitreihendaten spezialisiert ist, gezeigt.

1. Motivation

Die Nutzung von Umweltinformationssystemen (UIS) zur Verwaltung von räumlich verteilten, inhomogenen Umweltmonitoringdaten ist bei vielen Einrichtungen, wie Behörden, Unternehmen oder Ämtern, Standard. Der Fokus dieser Systeme liegt in der Verarbeitung von Sach- und Geodaten zur digitalen Abbildung der Umwelt mit der Zielsetzung zur Überwachung, Auswertung und Prognose. Im Zuge der Digitalisierung wird die Relevanz und Notwendigkeit zum Einsatz von UIS gesteigert. Dies liegt zum einen an der steigenden, zu verarbeitenden Datenmenge und zum anderen auch an der Vernetzung mit weiteren Systemen sowie der Verpflichtung oder dem Wunsch vieler Betreiber nach einer modernen Monitoringlösung für deren Anlagen bzw. Überwachungsgebiete.

Die Abdeckung der gewünschten Funktionen erfordert konzeptionelle und algorithmische Vorarbeit um zu gewährleisten, dass der spätere Anwender eine überschaubare, konsistente Sicht auf das System erhält. Dies gilt insbesondere für den Umgang mit Informationen zur Qualität der Umweltdaten, damit der Nutzer diese besser auswerten, bewerten und mit anderen Ergebnissen vergleichen kann.

In den folgenden Abschnitten werden Ansätze zur Qualitätssicherung (QS) am Beispiel der Software AL.VIS|Timeseries [1] vorgestellt. Die Software kann als Fachsystem zur Verwaltung von Zeitreihendaten in einem UIS eingesetzt werden.

2. Herausforderung

Bei der Nutzung und Bearbeitung gespeicherter Zeitreihendaten, können die Daten je nach Fragestellung in Klassen eingeteilt werden. Zum einen gibt es operative Daten, die für eine möglichst zeitnahe Auswertung verwendet werden und zu einem späteren Zeitpunkt an Wert verlieren. Anwendungsfälle dafür sind zum Beispiel Steuerungs- und Prozessinformationen oder Daten die zur Bewertung von Alarmzuständen herangezogen werden. Zum anderen gibt es strategische Daten, welche für eine langfristige Nutzung bestimmt sind. Typische Verwendungen finden sich hier in Berichten, Statistiken, Berechnung von Prognosen

oder Überwachung langfristiger Trends.

Ein Ziel der Weiterverarbeitung operativer Daten ist meist die Überführung der Messwerte hin zu langfristig nutzbaren und vergleichbaren Daten sowie der Sicherstellung der Qualität. Dazu werden Methoden der Zeitreihenanalyse angewendet und abgeleitete Zeitreihen gebildet sowie bestehende Werte validiert und korrigiert.

Mit diesem Hintergrund sind die Entwicklungen in der Methodik der Umweltüberwachung zu beachten. Die Entwicklung neuer Messverfahren, die Verfügbarkeit neuer, preiswerter Sensoren, deren Vernetzung und ihre Einbindung in Online-Übertragungssysteme haben zu einem deutlichen Anstieg verfügbarer Messwerte in immer kürzeren Zeitintervallen geführt. Aus diesen Gründen müssen Importe und Verarbeitungsschritte von Zeitreihen möglichst automatisiert ausgeführt werden und die sich ständig ändernden Datenlagen einbeziehen. Das gilt für die Basiszeitreihen und im Besonderen für abgeleitete Zeitreihen, als auch für darauf aufbauende Auswertungen. So sollten Aktualisierungen von Messwerten an Basiszeitreihen auch in abgeleitete Zeitreihen überführt werden.

Bei steigender Anzahl von Zeitreihen wird jedoch auch weiterhin erwartet, dass, die Daten mit einem gleichen Qualitätslevel bereitgestellt werden. Somit müssen Strategien entwickelt werden, welche die Qualitätssicherung möglichst dynamisch und automatisiert auf sich ändernden Daten ermöglicht.

Eine Lösung zur Sicherstellung der Qualität von Daten kann nicht durch eine einzelne Funktion umgesetzt werden, sondern resultiert aus dem Zusammenspiel der Systemkomponenten, welche jeweils ihre Aufgabe unter Berücksichtigung der QS erfüllen. In den nachfolgenden Abschnitten wird dieser Ansatz komponentenweise erläutert.

3. Datenhaltung

Die Grundlage für die Implementierung durchgängiger QS-Routinen beginnt mit der Speicherung der Daten. Diese müssen sicher, dauerhaft und konsistent abgelegt werden sowie zuverlässig und schnell abrufbar sein. Mit den derzeitigen Technologien werden diese Anforderungen datenbankgestützt realisiert. Eine lediglich dateibasierte Ablage

kann den Anforderungen nicht gerecht werden.

Strukturell werden zu den Zeitreihendaten, welche primär aus Tupeln von Messzeitpunkten mit dazugehörigen Messwerten bestehen, zusätzliche Felder für die QS vorgesehen und in Tabelle 1 aufgeführt.

Die genannten Felder müssen mit dem eigentlichen Messwert eine Einheit bilden und von allen Komponenten der Datenverarbeitung ausgewertet, bearbeitet und interpretiert werden können. Nur so kann sichergestellt werden, dass die Qualitätsinformation durchgängig zur Verfügung stehen und für den Anwender ersichtlich sind. Die Einheit von Messwert und den Qualitätsinformationen bedeuten zudem, dass Änderungen an den Flags auch als Aktualisierung des Messwerts interpretiert werden und entsprechende Logik auf diese Aktualisierung reagieren kann.

Feld	Kennzeichnung für:	Beispiel
<i>Status</i>	Verfügbarkeit	Unverändert Gelöscht Korrigiert
	Qualität	Suspekt Grenzwert Ungeprüft
<i>Import</i>	Validierungsverfahren beim Import	mit autom. Validierung Händisch
<i>Bemerkung</i>	textuelle Informationen	

Tabelle 1: Qualitätsinformationen an Messwerten

Zur zusätzlichen Sicherheit und Nachvollziehbarkeit werden Änderungen an den Originalwerten einer Zeitreihe als Korrekturwerte zu dem Originalwert gespeichert. Demzufolge bleibt der ursprüngliche Rohwert auch bei nachträglichen Änderungen gespeichert und kann bei Bedarf wiederhergestellt werden.

4. Datenimport

Der automatisierte oder händische Import von Daten bildet die Voraussetzung zur Schaffung der Datenbasis. Davon ausgehend können andere Komponenten, beispielsweise Export- und Berichtsfunktionen, Auswertungsmodule oder Validierungsroutinen, auf die Daten zugreifen und diese prozessieren.

Aus den praktischen Anforderungen der Anwender der AL.VIS Software wurden für den Import in die zentrale Datenbank diverse Module entwickelt (siehe Bild 1). Dabei wird deutlich, dass es heterogene Quellen, wie zum Beispiel Textdateien, Sensoren, Handeingaben, Datenbanken oder Dienste, gibt und deren Daten durch eigenständige Module in die zentrale Datenbasis importiert werden. Die Module unterscheiden sich je nach Technologie, Automatisierungsgrad oder implementierter Logik und können in Abhängigkeit des Einsatzgebiets und der

Nutzeranforderungen eingesetzt werden.

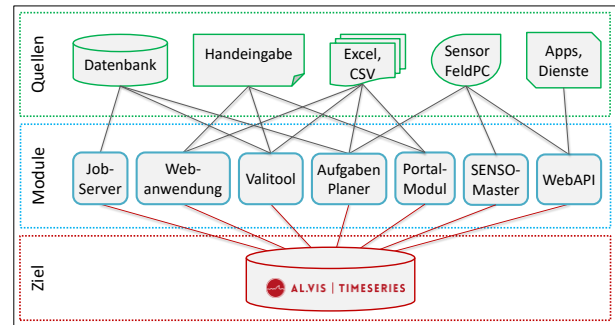


Bild 1: Übersicht zu möglichen Importmodulen und Datenquellen in AL.VIS|Timeseries

Eine wichtige Gemeinsamkeit zwischen den einzelnen Importmodulen ist der Umgang mit den Qualitätsinformationen der Messwerte (vgl. Tabelle 1). Bei allen Arbeitsschritten werden diese Informationen mitgeführt und in Abhängigkeit des Moduls automatisiert erfasst oder, wenn möglich, von der Quelle übernommen. Zum Beispiel erhalten durch einen Nutzer erfasste Messwerte eine entsprechende Import-Markierung, es werden Flags zur Qualität durch Validierungsroutinen gesetzt oder es werden eventuell vorhandene Qualitätsinformationen zu den Messwerten aus externen Datenbanken übersetzt und übernommen.

5. Methoden der Qualitätssicherung

Die Implementierung konkreter Methoden der Qualitätssicherung hängt stark von Einsatzfeld, den Datenquellen und Nutzeranforderungen ab. Daher werden nachfolgend die wichtigsten Komponenten und Grundsätze aufgezeigt.

Eine automatische Validierungskomponente für zu importierende Werte ist wichtig, damit die Messwerte auf Plausibilität mit An-/ Abstiegsgrenzen und Grenzwerte geprüft werden können. Die Ergebnisse der Validierung werden direkt während des Importprozesses am Messwert als Kennzeichner gespeichert und sind nachfolgend für weitere Komponenten sichtbar. So können zum Beispiel Messwerte, die bestimmte Grenzwerte verletzen, automatisiert als suspekt markiert und für eine spätere Nutzung in Berichten herausgefiltert werden. Eine weitere Möglichkeit zur automatischen Bewertung eintreffender Messwerte sind Software-Detektoren [2] mit dem Fokus auf der Alarmwerkerkennung. Die Detektoren bewerten stets den zuletzt eingetroffenen Messwert und sollen natürliche Schwankungen der Zeitreihen durch eine dynamische Anpassung der Entscheidungsparameter berücksichtigen, sodass nur auffällige Messwerte automatisiert gekennzeichnet werden. Die Kombination mehrerer Detektoren kann zur Bildung eines Alarmindex genutzt werden, bei dessen Überschreitung ein Alarmzustand ausgelöst wird. Ein Nachteil der Softwaredetektoren ist deren teils aufwändige zeitreihenbezogene Konfiguration bzw. Parametrisierung.

Zusätzlich zu der automatischen Validierung

unterstützt ein Tool zur händischen Validierung den Spezialanwender bei der nachträglichen Validierung. Dieser Vorgang ist zwar zeitaufwändig, jedoch spiegeln manuell validierte Werte für viele Nutzer die verlässlichste Datenbasis wieder. Messwerte mit dieser Kennzeichnung können in späteren Verarbeitungsphasen gesondert behandelt werden, indem beispielsweise manuell validierte Werte von automatischen Bewertungen unverändert bleiben oder bei Konflikten bevorzugt verwendet werden.

Für den späteren Anwender steht mit der Webanwendung ein Recherchewerkzeug mit Funktionen der QS bereit. Eine umfangreiche, interaktive Diagrammkomponente (siehe Bild 2) ermöglicht die visuelle Überprüfung der Zeitreihendaten sowie einer Vielzahl an Filtern, Berechnungen und Darstellungsformen. In einem Kommentarsystem können von allen Nutzern Hinweise direkt an den Zeitreihen für die Fachadministratoren erfasst werden. Zusätzlich sind die Qualitätsinformationen an Messwerten (vgl. Tabelle 1) durchgängig für den Nutzer recherchierbar.

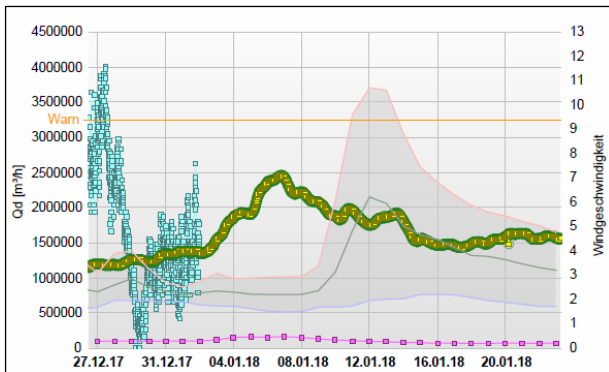


Bild 2: Beispiel eines interaktiven Diagrammes mit Zeitreihen mehrerer Quellen sowie berechneter Zeitreihen

Neben den genannten Komponenten sind auch Vorgaben und Prinzipien beim Umgang mit aktualisierten Daten notwendig. Damit wird zum Beispiel festgelegt, wie mit geänderten Qualitätsinformationen umgegangen wird oder wie sich Änderungen einer Zeitreihe auf eine abgeleitete Zeitreihe auswirken. Die Erfahrungen haben diesbezüglich gezeigt, dass die implementierten Auswertefunktionen immer auf den aktuellen Datenbestand zurückgreifen müssen und keine Kopien angelegt werden dürfen, damit Korrekturen oder neue Messwerte auch für alle Funktionen bereitstehen. Exemplarisch dafür ist die Diagrammkomponente, bei der nur die Konfiguration fertiger Diagrammvorlagen gespeichert wird und das Diagramm erst zum Zeitpunkt des Aufrufs mit den aktuellen Daten geladen wird.

Ein weiteres Prinzip betrifft abgeleitete Zeitreihen, z.B. Tagessummen- oder Tagesmittelwertzeitreihen die aus Zeitreihen mit einem 15min Intervall gebildet werden. Dabei wird das Ziel verfolgt, dass jeder Workflow bzw. jede Berechnungsvorschrift zur Erstellung einer oder mehrere abgeleiteter Zeitreihen, unabhängig und getrennt von anderen Workflows abgearbeitet werden kann. Das stellt vor

allen bei hochauflösenden, automatisch importierten Messwerten sicher, dass Berechnungsreihenfolgen und Zustände fest definiert sind und sich keine Workflows gegenseitig blockieren oder wechselseitig überschreiben.

6. Fazit

Die gestiegenen Datenmengen und die damit verbundenen Anforderungen an eine Zeitreihenverwaltung kann durch entsprechende Softwarekomponenten erfüllt werden. Essentiell ist die durchgehende Berücksichtigung der genannten Prinzipien, sodass Messwerte von der Erfassung bis hin zum Endnutzer mit den Informationen zur Qualität angereichert werden. Dies ermöglicht für den Nutzer eine bessere Einschätzung und Nachverfolgung der Werte.

Insgesamt ist die Tendenz erkennbar, dass Funktionen für eine flexible Reaktion der Auswertung sowie zur automatischen QS für eine wachsende Datenbasis an Bedeutung gewinnen. Dabei können zwar nicht alle Anforderungen der QS voll automatisiert erfüllt werden, aber die Softwarekomponenten können den Nutzer zur Erfüllung seiner Aufgaben bestmöglich unterstützen.

Literaturverzeichnis

- [1] AL.VIS, <http://alvis.software>
- [2] M. Haase, M. Fritz, F. Neubert, R. Kahnt, A. Kutzke, Scientific Reports, Journal of the University of Applied Sciences Mittweida, Nr. 1, 2014, 9-19.