

Erzeugung der XML-Daten: Umsetzung in einem LIMS

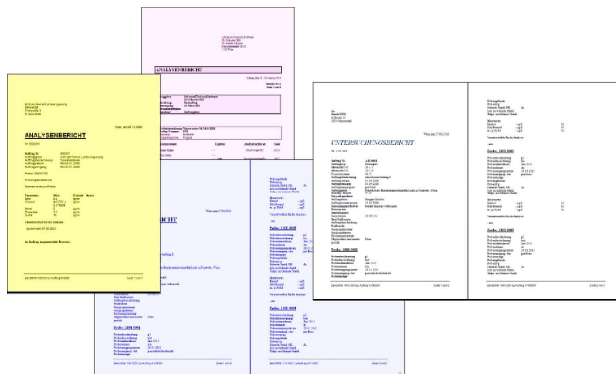
eGovernment im Umweltschutz, 17. November 2011

Gerald Schuster, Advanced Technical Software,
Zippererstraße 15, 1110 Wien
e-mail: office@ats-vienna.com Web: <http://www.ats-vienna.com>

Die Ausgangslage

Die Kerntätigkeit eines Untersuchungslabors besteht darin, Messungen (an bestimmten Proben oder direkt vor Ort) für einen Auftraggeber durchzuführen und die Messergebnisse an diesen oder Dritte (z.B. Behörden) zu übermitteln. Für Inhalt und Aufbau der Untersuchungsberichte existieren Mindestanforderungen in diversen Normen (z.B. ISO EN IEC 17025) bzw. Richtlinien. Die konkrete optische und inhaltliche Gestaltung bleibt aber dem Untersuchungslabor überlassen.

Vielfach ist das kein Problem, da der Auftraggeber nur die Untersuchungsberichte archiviert, um – falls nötig – jederzeit darauf zurückgreifen zu können. Häufig stellt sich auch nur die Frage, ob bestimmte Kriterien (z.B. gesetzliche Grenzwerte) erfüllt wurden oder nicht.

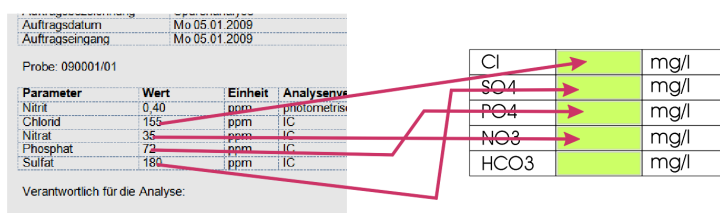


In der letzten Zeit kommt es aber immer häufiger vor, dass der Auftraggeber die Messergebnisse in exakt definierter, elektronischer Form benötigt, um sie entweder selbst weiter auszuwerten oder aber an Dritte (z.B. Gebietskörperschaften oder internationale Organisationen) zu übermitteln. Für diese Aufgabe stellen die „freien“ Berichtformate der verschiedenen Untersuchungslaboratorien natürlich ein großes Problem dar.

Konkrete Schwierigkeiten ergeben sich durch den unterschiedlichen Aufbau der Berichte sowie zwischen Sender und Empfänger nicht abgestimmte Bezeichner. Zum Beispiel könnte das Labor die Nitratkonzentration als „Nitrat“ im Bericht anführen, im Datensystem des Empfängers wird es aber als „NO3“ oder hingegen englisch „nitrate“ bezeichnet. Dasselbe gilt für Einheiten oder Textkonstanten (z.B. „n.n.“ vs. „N/D“ für „nicht nachweisbar“). Eine vollautomatische Übernahme der Daten ist in einem solchen Fall jedenfalls unmöglich.

Auftragsdatum	Mo 05.01.2009		
Auftragseingang	Mo 05.01.2009		
Probe: 090001/01			
Parameter	Wert	Einheit	Analysensystem
Nitrit	0.40	ppm	photometrisch
Chlorid	155	ppm	IC
Nitrat	35	ppm	IC
Phosphat	72	ppm	IC
Sulfat	180	ppm	IC
Verantwortlich für die Analyse:			
System vom 07.09.2006			

Cl	mg/l
SO4	mg/l
PO4	mg/l
NO3	mg/l
HCO3	mg/l



Lösungsansätze

Die einfachste Lösung des Problems wäre, dass auf Seiten des Berichtempfängers (Auftraggeber oder Behörde) ein Mitarbeiter die Werte aus dem Untersuchungsbericht in das eigene Datensystem eingibt. Naturgemäß ist aber die Bereitschaft gering, diesen erheblichen, zusätzlichen Aufwand zu tragen. Dazu gibt es am Markt derzeit eine Vielzahl konkurrierender Anbieter von Untersuchungsleistungen und es ist daher für den Auftraggeber nicht schwer, diesen Aufwand auf den Dienstleister zu überwälzen.

In der Praxis kann man daher eher von einem der beiden folgenden Szenarien ausgehen:

- a) das Untersuchungslabor gibt auf elektronischem Weg die Untersuchungsergebnisse in das Datensystem des Empfängers ein, beispielsweise über entsprechende Online-Formulare auf einer Internetseite des Auftraggebers
- b) Untersuchungslabor und Empfänger vereinbaren eine elektronische Schnittstelle zur automatisierten Datenübernahme. Aufgrund der Marktmacht des Auftraggebers kann es natürlich auch vorkommen, dass er ohne Mitwirkung des Labors eine bestimmte Schnittstelle vorschreibt

Häufig werden auch beide Varianten parallel angeboten und das Untersuchungslabor wählt die für sich günstigere aus. Falls das Labor über keine computergestützte Datenverwaltung (insbesondere ist damit ein Laborinformationssystem = LIMS gemeint) verfügt, könnte Variante a) sinnvoller sein, bei Laboratorien mit LIMS wird hingegen b) zum Einsatz kommen, um eine doppelte Erfassung aller Messergebnisse zu vermeiden.

Wir wollen uns im Folgenden nur mit Variante b) näher beschäftigen.

Was ist eine Schnittstelle?

Unter einer Schnittstelle in diesem Zusammenhang können wir uns einfach Definitionen bzw. Regeln vorstellen, die genau beschreiben, in welcher Form die zu übertragenden Daten erwartet werden. Vereinfacht gesprochen kann man sich ein Antragsformular einer Behörde vorstellen, in der die entsprechenden Daten (z.B. Familienname) in die passenden Felder einzutragen sind und bei manchen Angaben (z.B. Familienstand) nur die Auswahlmöglichkeit zwischen einigen, vordefinierten Möglichkeiten (ledig, verheiratet, geschieden, verwitwet) besteht.

Diese Daten können dann entweder als Datei gespeichert und an den Empfänger übermittelt (a) oder über direkten Weg elektronisch übertragen werden (b).

- a) Das Datensystem (z.B. LIMS) des Untersuchungslabors speichert die zu übertragenden Messdaten in eine Datei, die dann anschließend an den Empfänger geschickt wird. Das kann über Datenträger (z.B. CD-ROM oder USB-Stick) oder auch via e-mail geschehen. Praktischer ist es, wenn der Empfänger auf einer Seite seines Internetauftritts eine Möglichkeit zum Hochladen der Datei vorsieht, denn in diesem Fall kann nach Prüfung der Datei sogleich eine Empfangsbestätigung bzw. Fehlermeldung angezeigt werden.
- b) Das Datensystem des Labors stellt die Daten programmintern zusammen und sendet sie anschließend in elektronischer Form direkt an das Datensystem des Empfängers. Dieses prüft die Daten und antwortet direkt darauf, ob die Daten korrekt sind oder ob Fehler aufgetreten sind. Diese Variante ist natürlich für den Anwender mit Abstand die komfortabelste, die Übertragung kann mit einem einzigen Mausklick erfolgen. Technisch wird diese Art der direkten Datenverbindung in letzter Zeit häufig in Form von „WebServices“ realisiert, dazu unten mehr.

Alternativ ist es natürlich auch möglich, dass das Datensystem des Labors die Daten direkt in einem separaten Bereich des Datensystems des Empfängers ablegt, z.B. innerhalb der Datenbank. In der

Praxis kommt das aber nur innerhalb eines einzigen Unternehmens zum Einsatz, da ansonsten die Sicherheitsbedenken zu groß wären.

Datenformate

Als nächstes stellt sich die Frage, in welcher Form die Untersuchungsdaten innerhalb der zu übertragenden Datei (Variante a) codiert werden sollen. Hier gibt es verschiedene Ansätze, jeweils mit Vor- und Nachteilen:

- a) binäre Codierung: die Daten werden so ausgegeben, wie sie im Computer intern gespeichert werden. In der Praxis ist diese Form kaum gebräuchlich, da fehleranfällig und ohne geeignete Software zur Verarbeitung für Menschen unlesbar
- b) einfache Textdateien: die Daten werden nach einer vorgegebenen Struktur als Text ausgegeben, Reihenfolge und Trennzeichen der einzelnen Elemente sind vorgegeben. Mit jedem Texteditor zu lesen.
- c) XML (extensible markup language): XML-Dateien sind ebenfalls Textdateien, es gibt aber allgemeine, international einheitliche Regeln über den Grundaufbau, wie einzelne Werte gekennzeichnet und ausgegeben werden müssen. Zumindest bei geschickter Planung sind XML-Dateien sowohl für Menschen problemlos lesbar als auch vollautomatisch weiterverarbeitbar. Neue Schnittstellen werden überwiegend auf diese Weise definiert.

In unserem konkreten Anwendungsfall (Messergebnisse von Untersuchungslaboratorien, insbesondere in der Umweltanalytik) haben eigentlich nur die Varianten b) und c) praktische Bedeutung. Für die konkrete Wahl bzw. Definition einer Schnittstelle sind folgende Kriterien zu beachten:

- a) die Dateien sollten mit einem gängigen Texteditor für Menschen lesbar und verständlich sein
- b) die Korrektheit (formal und soweit wie möglich inhaltlich) einer Datei sollte für den Empfänger leicht zu überprüfen sein
- c) Hierarchien sollten einfach und übersichtlich abgebildet werden. Damit ist z.B. gemeint, dass ein Untersuchungsbericht mehrere Proben umfassen kann, von denen jede wiederum aus mehreren Teilproben bestehen könnte. Für jede Teilprobe werden dann untersuchten Parameter mit Messergebnissen und Einheiten angegeben. Schließlich könnte noch jedes Messergebnis als Mittelwert einer Mehrfachbestimmung mit den konkreten Einzelwerten angegeben sein.
- d) Möglichkeit zur „Übersetzung“ unterschiedlicher Begriffe zwischen Sender und Empfänger (z.B. „Na“ im Datensystem des Labors zu „Natrium“ beim Auftraggeber)

Alle diese Punkte lassen sich mit XML zweckmäßiger lösen, daher ist dieses Datenformat im Moment Stand der Technik.

Was ist XML?

XML ist die Abkürzung von extensible markup language. Unter „markup“ sind Bezeichner zu verstehen, die den „Sinn“ einer damit verbundenen Angabe erklären. Markups sind durch spitze Klammern gekennzeichnet und müssen sowohl den Beginn als auch das Ende definieren (z.B. <probe> ... </probe>). Mit „extensible“ ist gemeint, dass die erlaubten markups nicht vorgegeben sind, sondern zwischen Sender und Empfänger frei vereinbart werden können, solange die Grundstruktur von XML eingehalten wird.

XML hat schon vom Namen her Ähnlichkeiten mit der Seitenbeschreibungssprache für Webseiten „HTML“ (hypertext markup language). XML kann – etwas vereinfacht – als Verallgemeinerung von HTML mit strengeren Regeln verstanden werden. Während HTML definiert, welche Texte und Bilder auf einer Webseite in welcher Form (z.B. Größe, Schriftart, Formatierung) zu sehen sind, kann XML für beliebige Zwecke eingesetzt werden.

Ein sehr einfaches XML-Dokument für unsere Aufgabe (Messdatenübertragung) könnte wie folgt aussehen:

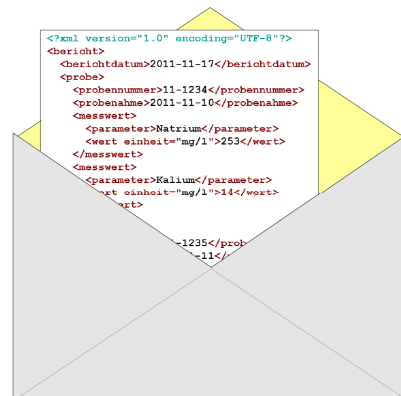
```
<?xml version="1.0" encoding="UTF-8"?>
<bericht>
  <berichtsdatum>2011-11-17</berichtsdatum>
  <probe>
    <probenummer>11-1234</probenummer>
    <probenahme>2011-11-10</probenahme>
    <messwert>
      <parameter>Natrium</parameter>
      <wert einheit="mg/l">253</wert>
    </messwert>
    <messwert>
      <parameter>Kalium</parameter>
      <wert einheit="mg/l">14</wert>
    </messwert>
  </probe>
  <probe>
    <probenummer>11-1235</probenummer>
    <probenahme>2011-11-11</probenahme>
    <messwert>
      <parameter>Natrium</parameter>
      <wert einheit="mg/l">195</wert>
    </messwert>
  </probe>
</bericht>
```

Bereits in der ersten Zeile wird definiert, dass es sich um eine XML-Datei handelt und dass als Zeichensatz UTF-8 verwendet wird. Gegenüber anderen Textdateien sind dadurch schon viele mögliche Fehlerquellen ausgeschaltet. Die einzelnen Elemente können für bessere Lesbarkeit eingerückt sein. Hierarchien wie im Beispiel lassen sich damit effizient und übersichtlich abbilden.

Ein enormer Vorteil von XML ist aber auch, dass fast jede Programmiersprache - sogar viele Makrosprachen – Funktionen zur vollautomatischen Prüfung auf formale Korrektheit haben. Gegenüber allgemeinen Textdateien spart dies viel Entwicklungszeit und vermeidet Fehlerquellen. Durch sogenannte XML-Schemata kann man sogar inhaltliche Richtlinien festlegen, z.B. welche Elemente vorkommen müssen oder welche Werte für bestimmte Angaben zulässig sind.

Die weiter oben erwähnte – noch wesentlich komfortablere - Variante der Datenübertragung über Webservice anstelle der Speicherung als XML-Datei verwendet ebenfalls XML-Daten, die gewissermaßen in eine „envelope“ verpackt und über Internet zwischen Sender und Empfänger ausgetauscht werden.

Dabei werden die Daten vollautomatisch vom Labor an das Datensystem des Empfängers geschickt. Dort werden sie auf formale und – soweit möglich – inhaltliche Korrektheit geprüft und das Ergebnis an das Labor zurückgesandt. Für den Anwender ist die Datenübertragung damit also mit einem einzigen Mausklick erledigt.



Umsetzung in einem LIMS für den Wasserdatenverbund NÖ

Für den Wasserdatenverbund Niederösterreich ist eine Schnittstelle auf XML-Basis vorgesehen (derzeit noch nicht im Echtbetrieb). Der aktuelle Stand ist aber bereits im LIMS „uniLIME“ implementiert.

Da – wie bereits oben erwähnt – die „Übersetzung“ von Parameternamen, Einheiten etc. vom LIMS in die die vom Empfänger erwarteten Begriffe notwendig ist, müssen diese Informationen zunächst im LIMS konfiguriert werden. Darüber hinaus werden vom Empfänger auch noch Listen zulässiger Werte für bestimmte Angaben übermittelt (z.B. Geruch einer Probe), die von Zeit zu Zeit erweitert werden und dann auch ins LIMS übernommen werden können.

Sobald diese Definitionen vollständig erfasst sind, kann man im LIMS mit wenigen Mausklicks eine geeignete XML-Datei erzeugen und an den Auftraggeber übermitteln.