Probleme und Lösungen für den Import von Bodendaten in das Bodeninformationssystem BORIS

SCHREIER Ingrid[1], TULIPAN Monika[1], MIKSITS Herbert[1]

1.1. Abstract

Für den Betrieb des österreichweiten Bodeninformationssystems BORIS (**BO**-den – **R**echnergestütztes **I**nformations-**S**ystem) ist die Zusammenführung und Bereitstellung großer Mengen bodenkundlicher Daten von zentraler Bedeutung. Diese Daten werden für den Import in die BORIS-Datenbank (ORACLE 8i) in ein definiertes Textformat (Schnittstelle) gebracht.

In der vorliegenden Arbeit werden die, bei der Bearbeitung dieser Schnittstellendateien erkannten Schwierigkeiten erklärt und in weiterer Folge entsprechende Lösungen vorgestellt.

Folgende wesentliche Problembereiche sind während der Entwicklung der Anwendung BORIS INTERN, über die u.a. der Datenimport durchgeführt wird, aufgetreten:

1. Länderspezifisches Zahlenformat
   Da die Darstellung von Zahlen aufgrund unterschiedlicher Dezimaltrennzeichen und Zifferngruppierungssymbole in verschiedenen Ländereinstellungen am PC (englische, deutsche,...) nicht eindeutig ist kann es zur falschen Interpretation einer Zahl kommen, z. B. 1.001 kann Tausendeins oder Einskommanullnulleins bedeuten.
2. Konvertierung von Daten (möglicherweise unbemerkt) durch externe Software wie z. B. beim Öffnen einer Datei (CSV-Extention) mit MS EXCEL, das üblicherweise zur Datenbearbeitung verwendet wird. Dies kann zu Fehlinterpretationen von Daten führen, z.B Umwandlung von Zahlen in Datumsangaben
3. Vorzeitiger Abbruch des Datenimports
4. Fehlerhafte und nicht konsistente Datensätze
5. Plausibilität der Daten

[1] Umweltbundesamt GmbH, Spittelauer Lände 5, A.-1090 Wien

Lösungen durch die Anwendung BORIS INTERN:

1. Mit der Anwendung BORIS INTERN wird eine Datenanalyse der Zahlen und der Versuch einer Interpretation der Werte unter der Voraussetzung, dass eine Datei nur ein Zahlenformat enthält, durchgeführt. Kann kein eindeutiges Zahlenformat erkannt werden, wird ein Hinweis „Bitte die Datei auf eindeutige Trennzeichen prüfen“ gegeben und die Daten werden nicht importiert.
2. Der Export von Datensätzen über BORIS INTERN erfolgt zwar im CSV-Format allerdings mit der Dateierweiterung „.txt“, was einen benutzerdefinierten Transfer in externe Bearbeitungsprogramme ermöglicht und softwarebedingten Datenverfälschungen weitestgehend entgegenwirkt.
3. Bis zum Abbruch werden die bereits importierten Datensätze in einer Protokolldatei dokumentiert.
4. Daten werden über das eindeutige Schnittstellenfomat importiert und während des Imports auf Konsistenz, Vollständigkeit und Eindeutigkeit geprüft. Fehlerhafte Datensätze werden nicht importiert. In diesem Fall wird sowohl ein Fehlerprotokoll als auch eine dem Schnittstellenformat entsprechende Datei mit allen fehlerhaften Datensätzen zur Nachbearbeitung erstellt.
5. Die importierten Zahlenwerte werden gegen vordefinierte Ober- und Untergrenzen und Ausgabe von nicht plausiblem Werten zur Kontrolle und eventuellen Nachbearbeitung geprüft.

Weiters werden die vorgestellten Probleme und Lösungen mit Beispielen aus der bodenkundlich-fachlichen Arbeitspraxis bei der Datenbearbeitung und -aufbereitung verbunden und ein Ausblick auf die weiteren Arbeitsschritte gegeben.