

Flexible Datenanalyse und Workflow-Designer zur Identifizierung von Chemikalien im Wasserkreislauf - streamFind

Das Verbundprojekt streamFind des IUTA, des Forschungszentrums Informatik (FZI) und unterstützender Partner (siehe Abbildung 1) hat zum Ziel, eine Plattform mit modularen Workflows zur Datenverarbeitung für die Massenspektrometrie zu schaffen, die in Umweltstudien zum Wasserkreislauf eingesetzt werden kann. Die Software wird den Nachweis, die Priorisierung und die Identifizierung von Substanzen im Wasser ermöglichen, ohne dass deren Vorhandensein oder Existenz bekannt sein muss. Darüber hinaus wird die Software das Verständnis und die Anwendung fortgeschrittener Datenanalysen (z. B. „Non-Target-Screening“, statistische Analyse, usw.) in routinemäßigen Umweltstudien erleichtern, wobei der Schwerpunkt auf der Standardisierung der Datenstrukturen und der Verarbeitungs-Workflows liegt, sodass die retrospektive Datenauswertung einfacher wird und die Nutzbarkeit und Relevanz von Probenkampagnen in Studien zum Wasserkreislauf erhöht werden.

Die Datenverarbeitung innerhalb der streamFind-Plattform ist als unabhängiges R-Paket implementiert (Installation verfügbar über das GitHub-Repository <https://github.com/odea-project/StreamFind>), um die kontinuierliche Entwicklung von Datenverarbeitungsmodulen zu fördern (siehe Abbildung 2). Die Software richtet sich aufgrund der angestrebten umfassenden Dokumentation, der gut kategorisierten Reihe integrierter modularer Funktionen und der grafischen Bedienoberfläche an Wissenschaftlerinnen und Wissenschaftler, Akademikerinnen und Akademiker sowie Technikerinnen und Techniker aus dem Umfeld der Lebenswissenschaften.



Abbildung 1: Direkte Forschungspartner und assoziierte Netzwerkpartner für das Projekt streamFind.

Das R-Paket StreamFind bietet ein robustes Backend für die Verarbeitung von LC-HRMS-Daten, ergänzt durch ein flexibles Frontend, das interaktive Analysen und die automatisierte Erstellung von Berichten unterstützt. Eine wichtige Errungenschaft ist die Entwicklung nativer, leistungsstarker Algorithmen (d. h. Feature-Erkennung, Isotopen-Annotation, Matrix-Suppression-Korrektur, Fold-

Change-Analyse und Gap-Filling), die die analytische Zuverlässigkeit verbessern und die Abhängigkeit von proprietärer Software verringern. Zusätzliche Integrationen, darunter MetFrag für die In-silico-Fragmentierung und Tools zur Erzeugung von Transformationsprodukten, erweitern die Fähigkeit der Plattform, bekannte und unbekannte Chemikalien im Wasserkreislauf zu identifizieren.

Darüber hinaus konnte im Rahmen des Projekts demonstriert werden, wie eine Erweiterbarkeit der Plattform über den ursprünglichen Umfang hinaus möglich ist. Beispielsweise wurden zusätzliche Algorithmen für die Verarbeitung von LC-HRMS-, LC-UV- und Raman-spektroskopischen Daten implementiert und bei der Analyse großer Biomoleküle unter verschiedenen Workflow-Anwendungsfällen demonstriert. Dies unterstreicht die Vielseitigkeit des Systems und seinen Wert für verschiedene wissenschaftliche Fachdomänen. Insgesamt schafft streamFind eine nachhaltige Grundlage für eine moderne, FAIR-konforme Umweltdatenanalyse. Seine offene, modulare Architektur gewährleistet langfristige Wartbarkeit, fördert die Entwicklung der Community und ermöglicht Forschern und Praktikern die Durchführung fortschrittlicher, transparenter und reproduzierbarer NTS-Workflows. Das Projekt leistet somit einen bedeutenden Beitrag zur Umweltüberwachung, zur Bewertung der Wasserqualität und zur breiteren Einführung datengesteuerter Methoden in der analytischen Wissenschaft.

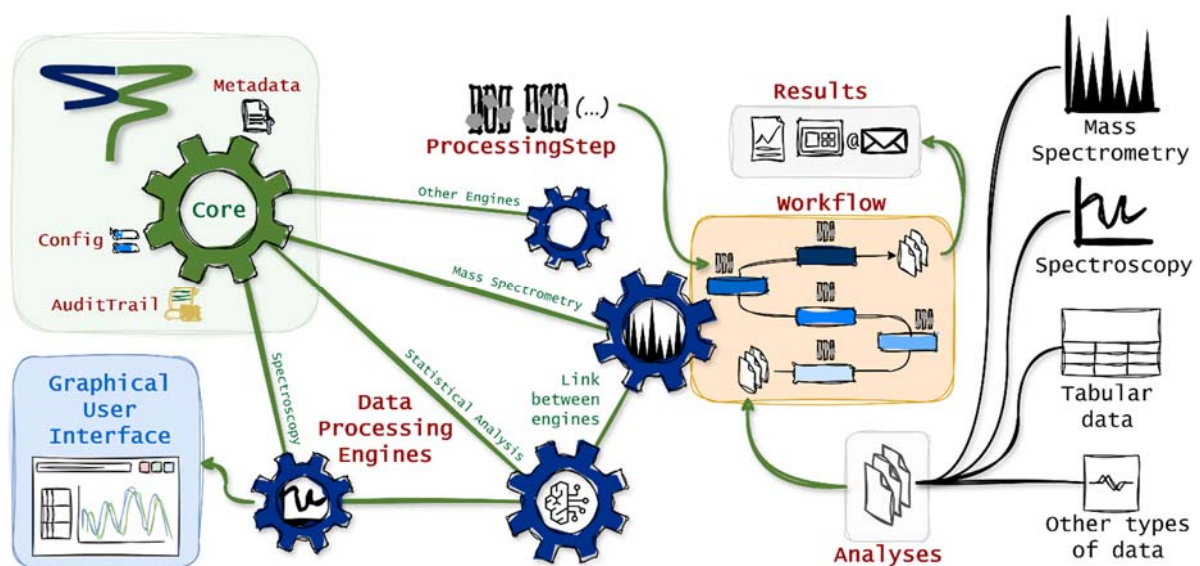


Abbildung 2: Implementierungskonzept des StreamFind R-Pakets mit Schwerpunkt auf den acht Hauptklassen, aus denen sich das Framework zusammensetzt.

Weiterführende Informationen können durch Scan des QR-Codes erhalten werden.



Bundesministerium
für Forschung, Technologie
und Raumfahrt