# Anpassung und Weiterentwicklung des Konzeptes ProvDecks für CTD-Messungen am IOW

**Hintergrund** Die Sicherstellung der Nachvollziehbarkeit und Reproduzierbarkeit wissenschaftlicher Daten und Ergebnisse ist eine zentrale Voraussetzung für qualitätsgesicherte Forschung. Insbesondere in datenintensiven Disziplinen wie den Meereswissenschaften gewinnt die strukturierte Dokumentation von *Provenance*-Informationen, d.h. Informationen über Herkunft der Daten, Verarbeitungsschritte, beteiligte Akteure und Werkzeuge, zunehmend an Bedeutung. Am **Leibniz-Institut für Ostseeforschung Warnemünde (IOW)** werden seit vielen Jahrzehnten ozeanographische Messungen durchgeführt, die ein wertvolles Archiv zur Langzeitbeobachtung der Ostsee darstellen. Diese Daten sind jedoch aufgrund unterschiedlicher Erhebungszeiträume, Gerätegener-ationen und Formate stark heterogen.

Zur Messung der elektrischen Leitfähigkeit (*Conductivity*), der Temperatur (*Temperature*) und der Tiefe (*Depth*) nutzt das IOW sogenannte *CTD-Sonden*. Zusätzlichen Sensoren messen weitere Parameter, wie beipsielsweise den Sauerstoff oder die Wassertrübung. Jedes vertikale Messprofil (*Cast*) erzeugt komplexe Daten- und Metadatensätze, die in standardisierten Dateien abgelegt werden. Für die effiziente Nachnutzung dieser Daten sowie die Nachvollziehbarkeit ihrer Auswertungen ist eine strukturierte Provenance-Dokumentation erforderlich.

Zur automatischen Erhebung und Auswertung von Provenance-Daten haben wir das Konzept der ProvDecks entwickelt. Dieses Konzept basiert auf drei miteinander verbundenen Strukturen:

- **ProvCards**, die als maschinenlesbare Metadateneinheiten einzelne Entitäten, Aktivitäten oder Agenten beschreiben;
- **ProvDecks**, die Sammlungen von Provenance-Informationen (inklusive ProvCards, Metadaten und Graphstrukturen) für Versionierung und Austausch bereitstellen;
- **ProvGraphs**, die automatisch erzeugte Visualisierungen der Abhängigkeiten zwischen Daten, Prozessen und Akteuren ermöglichen.

**Aufgabenstellung** Ziel der Arbeit ist die Anpassung und Weiterentwicklung dieses Konzeptes für die spezifischen Anforderungen der CTD-Messungen am IOW. Dazu gehört die präzise Definition geeigneter Metadatenfelder für die ProvCards, die formale Spezifikation von ProvDecks im Kontext typischer CTD-Datenworkflows sowie die Erzeugung von ProvGraphs, die eine nachvollziehbare Abbildung der Beziehungen zwischen Daten, Prozessen und Ergebnissen ermöglichen. Die Arbeit verfolgt somit drei Ziele:

(1) die Entwicklung eines konzeptionellen Modells zur Provenance-Erfassung im CTD-Kontext,

(2) die prototypische Umsetzung dieses Modells in Form von ProvCards, ProvDecks und ProvGraphs, sowie

(3) die Validierung der Ansätze anhand realer CTD-Datensätze aus der Forschungsarbeit am IOW.

Durch die Integration dieser Konzepte wird eine verbesserte Nachvollziehbarkeit und Transparenz erreicht, die sowohl die interne Datenverwaltung unterstützt als auch die langfristige Nachnutzung der Daten in internationalen und interdisziplinären Zusammenhängen erleichtert.

Die Ergebnisse dieser Bachelorarbeit leisten einen Beitrag zur Standardisierung und nachhaltigen Nutzung von marinen Forschungsdaten und tragen dazu bei, die Provenance-Frameworks für heterogene wissenschaftliche Datenbestände weiterzuentwickeln.

### Teil-Aufgaben/Forschungsfragen

- Analyse und Strukturierung bestehender CTD-Datenworkflows am IOW in Bezug auf Provenance-Aspekte.
- Entwicklung und formale Spezifikation (inklusive Implemeniterung) geeigneter ProvCard-Typen für CTD-Messungen (Entitäten, Aktivitäten, Agenten).

- Aufbau von ProvDecks, die typische Verarbeitungsschritte und Datensammlungen im CTD-Kontext repräsentieren.
- Implementierung von ProvGraphs zur Visualisierung und Abfrage von Abhängigkeiten innerhalb der Datenflüsse.
- Validierung des entwickelten Ansatzes anhand ausgewählter CTD-Datensätze.

## Formalia

- Ansprechpartnerinnen:
  - Tanja Auge (tanja.auge@ur.de)
  - Susanne Jürgensmann (susanne.juergensmann@io-warnemuende.de)
  - Susanne Feistel (susanne.feistel@io-warnemuende.de)
- Voraussetzungen:
  - Interesse an der Entwicklung von Provenance-Konzepten für real world-Szenarien
  - Motivation zur eigenständigen, kreativen Lösung abstrakter Modellierungsprobleme
  - Fähigkeit zur eigenständigen Recherche und Auswertung wissenschaftlicher Literatur
  - Bereitschaft zur vertieften Einarbeitung in die Themen CTD-Messungen am IOW sowie Provenance
  - Erste Erfahrungen mit Provenance, PROV-Standard, alternativ Bereitschaft zur vertieften Einarbeitung
  - Grundkenntnisse in Programmierung (Python)

## Literatur


- T. Auge, S. Feistel, F. J. Ekaputra, M. Klettke, S. Jürgensmann, E. Michels, L. Waltersdorfer: Towards an Integrated Provenance Framework: A Scenario for Marine Data. In: *EuroS&P Workshops*. IEEE, 2024.
- T. Auge, F. J. Ekaputra, S. Feistel, S. Jürgensmann, M. Klettke, L. Waltersdorfer: Challenges of Tracking Provenance in Marine Data. In: *International Conference on Marine Data and Information Systems*, 2024.
- T. Auge, G. Bali, C. Kindler, M. Klettke, D. Knüttel, W. Söldner, Wolfgang, T. Wettig: Provenance for Lattice QCD Workflows – An Update. In: *ProvenanceWeek*, 2025.
- Git Repository: provQCD. https://phygit.ur.de/kic04594/provqcd
- T. Auge, F. Ekaputra: ProvDecks in Action: Bridging Workflow and Data Provenance in Natural Science Data Pipelines. (Ausschnitte aus dem Projektantrag)
- F. Z. Khan, et al.: Sharing Interoperable Workflow Provenance – A Review of Best Practices and their Practical Application in CWLProv. In: *GigaScience*, 2019.
- M. Mitchell, et al.: Model Cards for Model Reporting. In: *FAT*. ACM, 2019.
- L. Moreau, et al.: W3C PROV-DM – The PROV Data Model. https://www.w3.org/TR/prov-dm/
- L. Moreau, et al.: PROV-JSON – A JSON Serialization for PROV. https://www.w3.org/submissions/prov-json/
- L. Moreau, et al.: PROV-Template – A Quick Start. https://lucmoreau.wordpress.com/2017/03/30/prov-template-a-quick-start/