

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

## Teil I Kurzbericht

Förderkennzeichen: 03F0888A

Zuwendungsempfänger: Alfred-Wegener-Institut Helmholtz Zentrum für Polar- und Meeresforschung, Projektleitung: Dr. Sebastian Mieruch-Schnülle

Verbundprojekt: MOSAiC 2 - M-VRE: Online Services zur Erkundung, Analyse und Visualisierung von MOSAiC Daten

Vorhaben: Implementierung interdisziplinärer Analysen

Laufzeit: 01.07.2021 – 31.12.2024

Berichtszeitraum: 01.07.2021 – 31.12.2024

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 03F0888A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen / den Autoren.

### 1. Aufgabenstellung sowie wissenschaftlicher und technischer Stand

Das M-VRE-Projekt (MOSAiC Virtual Research Environment, <https://mosaic-vre.org/about>) hat das primäre Ziel, eine innovative webbasierte Plattform zu entwickeln, die dem MOSAiC-Konsortium, der globalen Klima-Community und der interessierten Öffentlichkeit moderne Online-Services zur Verfügung stellt. Diese Services sind darauf ausgelegt, die Analyse, Erforschung und Visualisierung der Daten der MOSAiC Expedition zu ermöglichen und bieten effizienten und benutzerfreundlichen Online-Zugriff, der es Wissenschaftler\*innen erlaubt, komplexe, heterogene und große Datensätze interdisziplinär auszuwerten. Die Plattform wird innerhalb der bestehenden Cloud-Infrastruktur des Alfred-Wegener-Instituts (AWI) implementiert und integriert fortschrittliche Methoden aus den Bereichen Online-Datenanalyse, Künstliche Intelligenz (KI) und Big Data. Ein wichtiger Aspekt ist die Aggregation und Harmonisierung der Daten in großen, einheitlichen Datenkollektionen für alle Kompartimente wie Ozean, Eis und Atmosphäre. Dies ermöglicht es den Wissenschaftler\*innen, arktische Parameter aus verschiedenen Sphären an denselben Orten und Zeiten zu vergleichen, was die interdisziplinäre Forschung erheblich erleichtert. M-VRE erhöht die Sichtbarkeit und Nutzbarkeit der MOSAiC-Daten und fördert die Zusammenarbeit zwischen verschiedenen wissenschaftlichen Disziplinen. Das M-VRE-Projekt sieht auch die Ausbildung des wissenschaftlichen Nachwuchses vor, indem es fünf PostDoc-Stellen / Wissenschaftler\*innen-Stellen bereitstellt (AWI, DKRZ, DLR). Die Stärkung der AWI-Rechenzentrumsinfrastruktur und die Teilnahme an nationalen sowie internationalen Konferenzen sind ebenfalls zentrale Aufgaben. M-VRE treibt die wissenschaftliche und gesellschaftliche Verwertung der Daten voran sowie die Verbreitung der MOSAiC-Daten. Durch Workshops, Video-Tutorials und eine eigene Website wird die Sichtbarkeit des Projekts gefördert. Insgesamt zielt M-VRE darauf ab, ein wichtiges Werkzeug für die Klima-Community zu schaffen, das die wissenschaftliche Datenauswertung in der Arktis unterstützt und die großen Herausforderungen des Klimawandels adressiert. Der wissenschaftliche und technische Stand des M-VRE-Projekts spiegelt die neuesten Entwicklungen in den Bereichen Cloud-Computing, Online-Services, Künstliche Intelligenz (KI) und Datenmanagement wider. Diese Technologien sind entscheidend für die effiziente Analyse und Verarbeitung großer, heterogener Datensätze, die im Rahmen des MOSAiC-Projekts

generiert werden. M-VRE zielt darauf ab, die Prinzipien von FAIR Data (Findable, Accessible, Interoperable, Reusable) zu implementieren, um den Zugang zu und die Wiederverwendbarkeit von Daten zu verbessern. Die Plattform ermöglicht es Wissenschaftler\*innen, komplexe Datenanalysen durchzuführen und interdisziplinär zu arbeiten, was für die Beantwortung großer wissenschaftlicher Fragen im Zusammenhang mit dem Klimawandel und der arktischen Meereisbedeckung von entscheidender Bedeutung ist. Die Integration von KI-Methoden zur automatisierten Qualitätskontrolle von Ozeanmessdaten lotet neue Chancen und Möglichkeiten aus. Technologisch nutzt M-VRE eine robuste Cloud-Infrastruktur, die auf den bestehenden Ressourcen des AWI basiert. Zu den implementierten Technologien gehören Docker-Container, Data Cubes, und Jupyter-Notebooks, die eine flexible und skalierbare Datenverarbeitung ermöglichen. Die Plattform verwendet auch moderne Web-Frameworks und Programmiersprachen wie Laravel, Node.js, jQuery, Vue und Angular, um benutzerfreundliche Online-Tools für die Datenanalyse, Visualisierung und Exploration bereitzustellen. Darüber hinaus nimmt M-VRE aktiv an internationalen Konferenzen teil, um state-of-the-art Entwicklungen voranzutreiben und die Möglichkeiten von KI-Methoden weiter zu erkunden. Die technische Infrastruktur wird so gestaltet, dass sie den Anforderungen der globalen Wissenschafts-Community gerecht wird und die Reproduzierbarkeit sowie die kollaborative Nutzung von Daten fördert. Die M-VRE Homepage unter <https://mosaic-vre.org> repräsentiert die Anlaufstelle für Wissenschaftler\*innen aber auch für die interessierte Öffentlichkeit. Ein kurzes Video präsentiert M-VRE und informiert über die wichtigsten Funktionen. Über die einfache Navigationsleiste erreichen Nutzer\*innen leicht die gewünschten Informationen, z.B. Zugang zu den Web-Services, der Dokumentation oder den Video-Tutorials.

## **2. Ablauf des Vorhabens**

Das M-VRE Projekt verfolgt das Ziel, eine webbasierte Plattform für die Analyse, Erforschung und Visualisierung von MOSAiC-Daten zu entwickeln und der wissenschaftlichen Community und der Öffentlichkeit zur Verfügung zu stellen. Der Ablauf des Projektes gliedert sich in mehrere Phasen, die auf die technische Entwicklung, Nutzerinteraktion und Verwertung abzielen. Die Basis des M-VRE Projektes bildet der MOSAiC Datensatz, welcher im Langzeitarchiv PANGAEA abgelegt wird. Dieser Datensatz muss für die Verwendung mit Tools zur Online-Analyse (z.B. webODV) in das passende Format übertragen werden, was der Schwerpunkt von Arbeitspaket 2 ist. Entscheidend ist hier die Entwicklung von automatisierten Skripten, welche die Übertragung von Datensätzen schnell und effizient durchführen. Gleichzeitig wird die Integration der Web-Services angegangen (Arbeitspaket 1). Dazu gehören webODV (Arbeitspaket 3), Jupyter Notebooks mit dem Tool DIVA, SalaciaML und Data Cubes (Arbeitspaket 4). Diese Tools vereinfachen die Datenanalyse und -visualisierung, wobei hier auch moderne KI-Technologien zur Datenverarbeitung integriert werden. Regelmäßige Workshops werden organisiert, um Nutzer\*innen in der Anwendung der Plattform zu schulen und den Austausch von Erfahrungen zu fördern. Besonders der Austausch mit dem MOSAiC Konsortium ist hier ein Schwerpunkt. Die M-VRE wird auf nationalen und internationalen Konferenzen präsentiert, um die Sichtbarkeit des Projekts zu erhöhen und neue Nutzer\*innen zu gewinnen. Die Teilnahme an Konferenzen ermöglicht es, die wissenschaftliche Community über die Fortschritte und Möglichkeiten der M-VRE zu informieren.

## **3. Wesentliche Ergebnisse**

Kernziel von M-VRE ist nutzerfreundliche, effiziente und mächtige Tools zur online Analyse der MOSAiC Daten stabil und operationell anzubieten. Entscheidend dafür ist die Vorbereitung bzw. Harmonisierung der Daten. Die MOSAiC Daten, wurden in Arbeitspaket 2 aggregiert, harmonisiert und kategorisiert (Atmosphäre, Biogeochemie, ...). Die jeweiligen Kategorien beinhalten eine Vielzahl von Datensätzen. Dennoch bleibt die Anzahl der Datensätze in einem handhabbaren Rahmen, da wir viele Datensätze effizient aggregieren konnten. Somit haben Nutzer\*innen hier einen extrem effizienten und zentralisierten Zugriff auf die MOSAiC Daten. Auf dieser Datenbasis setzt das Tool webODV auf und bietet eine mächtige Plattform für wissenschaftliche Analysen, aber auch für

das einfache und schnelle Erkunden der Daten. Bis November 2023 wurden etwa 4121 MOSAiC-Datensätze in PANGAEA veröffentlicht, von denen ca. 2786 für die Integration in das webODV-System geeignet sind. Der Großteil dieser Datensätze wurde aggregiert und zu 178 Daten-Kollektionen zusammengeführt, die nun in M-VRE, bzw. über die webODV und Jupyter Notebook Tools verfügbar sind. Spezielle Konvertierungsskripte wurden entwickelt, um die unterschiedlichen Datenformate in ein einheitliches ODV-Format zu überführen. Diese Automatisierung ist für die Effizienz der Datenkonvertierung essentiell. Wichtig zu erwähnen ist auch die Erstellung von ausführlichem Informationsmaterial, um Nutzer\*innen den Einstieg in die Datenanalyse mit M-VRE zu erleichtern, welche über die Homepage zugänglich sind. Für webODV wurden 8 ausführliche Video-Tutorials produziert, die die wichtigsten Funktionen und Features zeigen.

Für den Jupyter-Notebook Service haben wir gemeinsam mit dem DKRZ, Vorlagen bzw. Templates entwickelt, die den Zugriff auf die Daten für Nutzer\*innen so einfach wie möglich gestalten und damit den Zugriff sicherstellen.

Abschließend schauen wir noch auf die Nutzerstatistik für webODV von Februar 2022 bis Februar 2025. Abbildung 1 zeigt die Zugriffe pro Monat. Und Abbildung 2 zeigt die globale Verteilung der Nutzer\*innen. Während Hochphasen der MOSAiC Datenauswertung, besonders der Konferenzen wurden mehr als 100 Besucher pro Monat registriert. Und obwohl die MOSAiC Expedition nun schon fünf Jahre zurückliegt wird webODV immer noch benutzt. Deshalb ist es das Ziel, webODV noch viele Jahre im Produktivmodus zu betreiben, um diesen einmaligen „Datenschatz“ möglichst effizient auszuwerten.

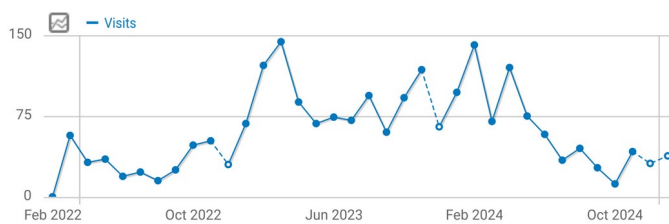


Abbildung 1: webODV Nutzerzahlen pro Monat.

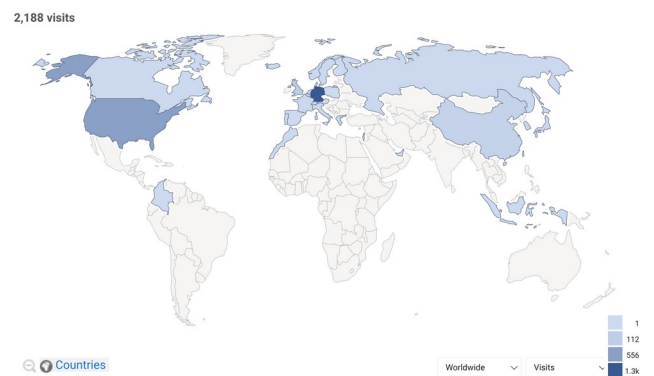
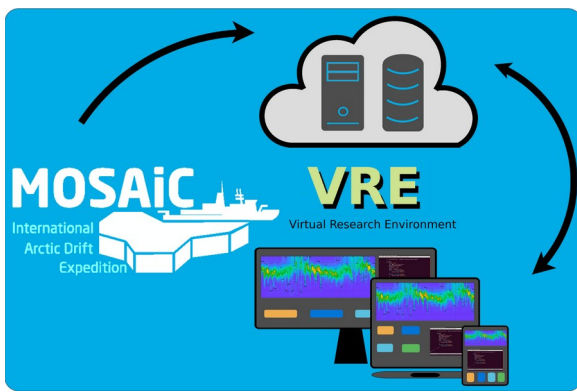


Abbildung 2: Globale Nutzung von webODV.



## Teil II Eingehende Darstellung

Förderkennzeichen: 03F0888A

Zuwendungsempfänger: Alfred-Wegener-Institut Helmholtz Zentrum für Polar- und Meeresforschung, Projektleitung: Dr. Sebastian Mieruch-Schnülle

Verbundprojekt: MOSAiC 2 - M-VRE: Online Services zur Erkundung, Analyse und Visualisierung von MOSAiC Daten

Vorhaben: Implementierung interdisziplinärer Analysen

Laufzeit: 01.07.2021 – 31.12.2024

Berichtszeitraum: 01.07.2021 – 31.12.2024

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 03F0888A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen / den Autoren.

### 1. Erzielte Ergebnisse und Gegenüberstellung der vorgegebenen Ziele

Ziel des AWI Vorhabens ist große ODV-Kollektionen aus den MOSAiC-Daten zu erstellen, um einen effizienten, einheitlichen und interdisziplinären Zugriff auf diese Daten zu ermöglichen. Dies ist besonders wichtig für die Analyse und Visualisierung der Daten, die aus verschiedenen wissenschaftlichen Disziplinen stammen.

Das Arbeitspaket WP2 (Datenmanagement und Konvertierung) ist zentral für die Erzeugung und Verwaltung der ODV-Kollektionen. Hier werden die im PANGAEA-Datenarchiv vorhandenen Datensätze, die im .tab-Format vorliegen, in ein ODV-lesbares .txt-Format umgewandelt. Ein Konvertierungsskript, das von Dr. Roland Koppe (AWI) entwickelt wurde, wird als Basis benutzt. Besonders wichtig zu erwähnen sind die Metadaten, die in die ODV Kollektionen integriert werden und z.B. für die Nachverfolgbarkeit wichtig sind. Die Metadaten enthalten unter anderem auch Links zu den Originaldaten und Datenerzeuger\*innen auf PANGAEA. Es wird eine vorläufige Datenstruktur festgelegt, die auf den verschiedenen MOSAiC-Teams basiert (z.B. Ozean, Atmosphäre, Meereis). Diese Struktur ermöglicht eine sinnvolle Organisation der Daten und sorgt dafür, dass die Daten-Kollektionen übersichtlich bleiben. Abbildung 1 zeigt die Datenstruktur in webODV.

Hervorzuheben sind die Herausforderungen, die sich aus der Vielfalt der Datensätze ergeben, insbesondere hinsichtlich der Bereitstellung der Daten für unterschiedliche Kollektionstypen (z.B. Profile oder Trajektorien). Daher sind Anpassungen des Konvertierungsskripts (wir nennen es PANGAEA2ODV) notwendig, um den spezifischen Anforderungen der verschiedenen Datensätze gerecht zu werden. PANGAEA2ODV ist in der Programmiersprache Python implementiert und als Teil der automatisierten Datenpipeline im AWI Cloud-System integriert.

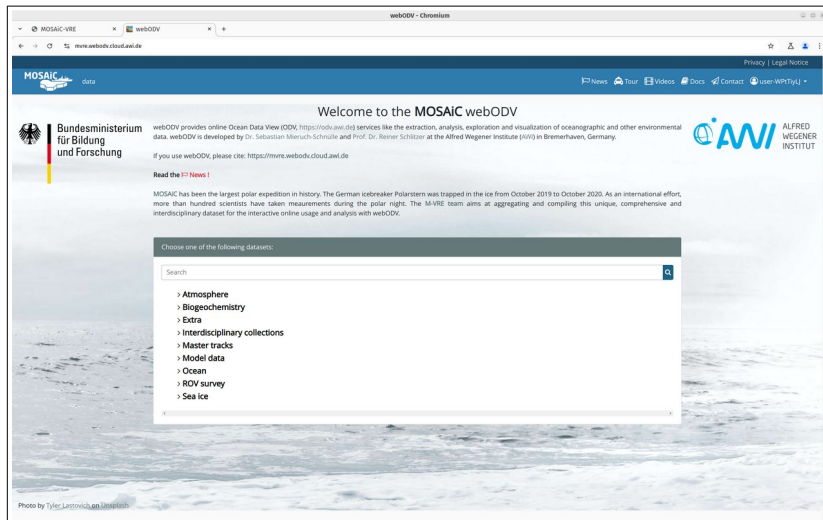


Abbildung 1: MOSAIC Daten wurden in M-VRE in Kategorien geordnet.

Das AWI Cloud-System basiert auf VM-Ware. Hier können wir über ein Web-Interface (Abbildung 2) Ressourcen (Virtuelle Maschinen, VM's) ausrollen bzw. aktivieren. Über verschiedene Konfigurationen kann festgelegt werden, ob die VM Zugriff auf das interne AWI Netzwerk haben soll, oder in einem abgesicherten (DMZ) Bereich außerhalb des internen Netzwerkes eingerichtet werden soll.

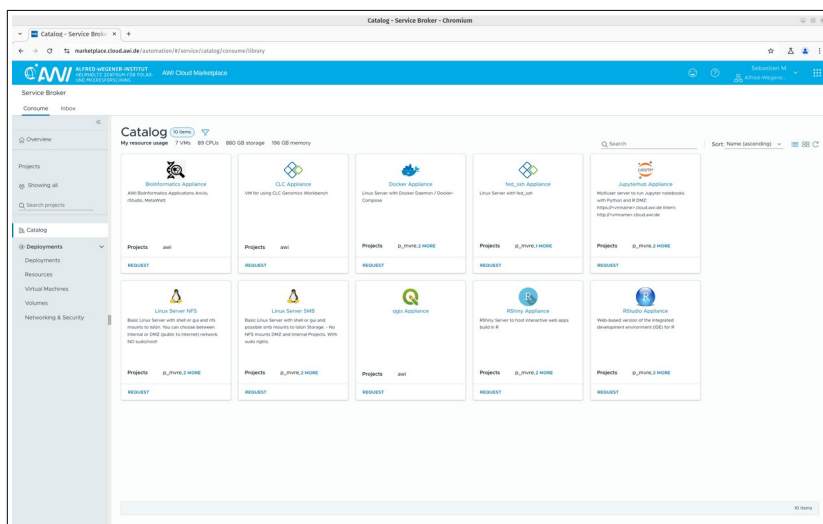


Abbildung 2: AWI's Cloud-System: Marketplace.

Im AWI Cloud-System legen wir vorkonfigurierte Jupyter-Hub Systeme an, um die Datenpipeline PANGAEA2ODV zu betreiben. Abbildung 3 zeigt einen Teil des dafür zuständigen Notebooks. Abbildung 4 zeigt schematisch die Vorgehensweise der Datenpipeline.

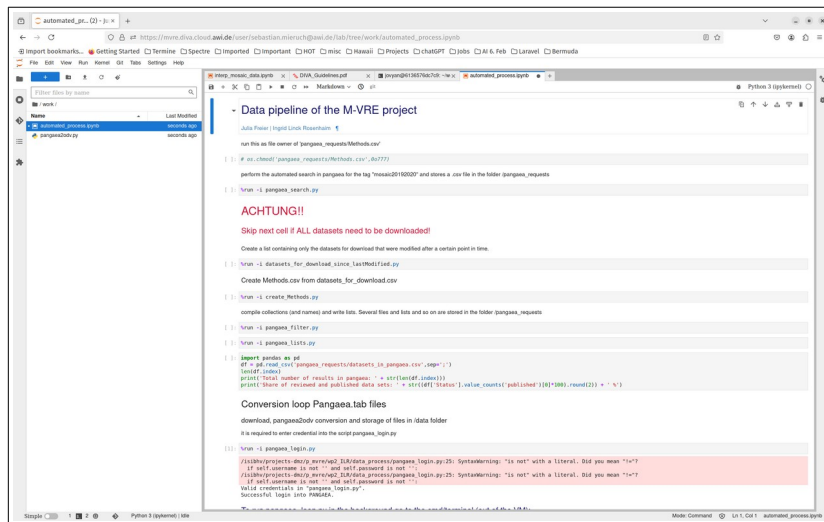


Abbildung 3: Ausschnitt der M-VRE Datenpipeline, implementiert als Jupyter Notebook in Python.

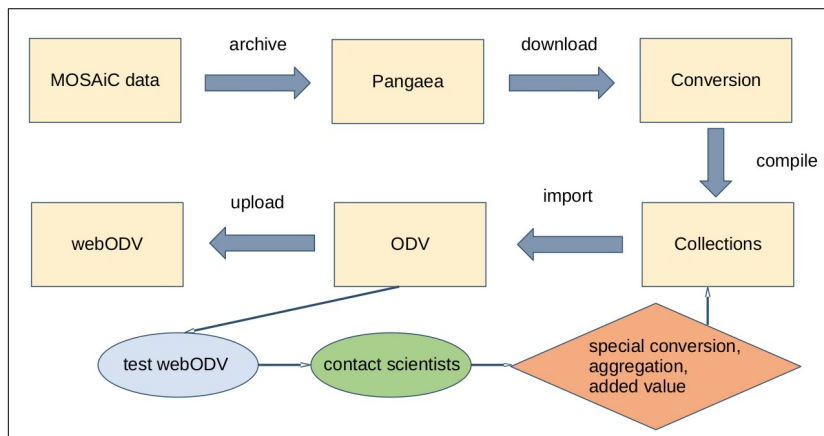


Abbildung 4: Schematische Darstellung der M-VRE Datenpipeline.

Die ersten MOSAiC-Datenkollektionen, die in WP2 erzeugt werden, können getestet werden und für den operationellen Einsatz wurde der webODV-Dienst im AWI Cloud-System bzw. der containerbasierten MSS Lösung vorbereitet. Die MOSAiC-Datenkollektionen wurden in das MSS System transferiert, um sie für webODV nutzbar zu machen. Abbildung 5 zeigt die Analyse eines MOSAiC Datensatzes mit webODV.

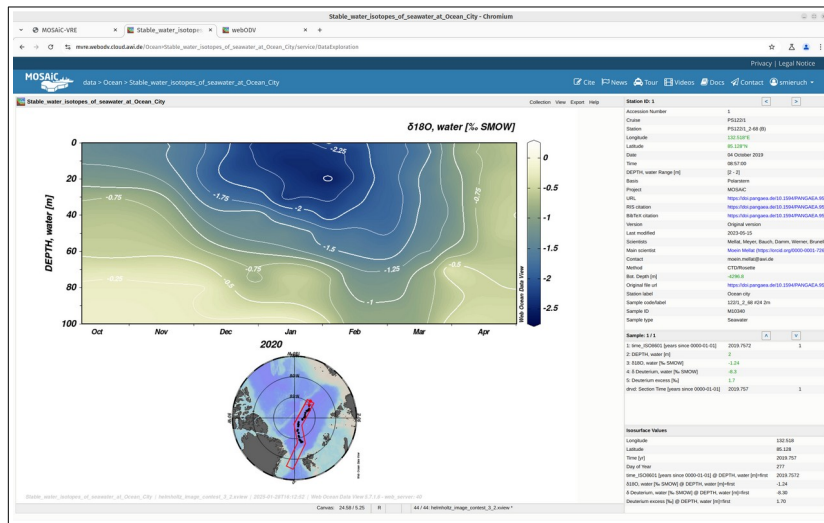


Abbildung 5: webODV Beispiel.

Aufgrund von Umstellungen im AWI-Rechenzentrum konnte der webODV-Dienst jedoch nicht sofort in den produktiven Betrieb genommen werden. Nach einigen Wochen der Problemlösung durch die Netzwerktechniker konnte die Installation Mitte Februar 2022 stabil betrieben werden.

Ein wichtiger Bestandteil von M-VRE, besonders von webODV, ist die Dokumentation, wobei diese direkt über den webODV-Dienst abrufbar ist, oder über die M-VRE Homepage, bzw. die Dokumentation unter <https://docs.mosaic-vre.org/webODV/> (Abbildung 6). Diese Dokumentationen bieten Anleitungen zur Nutzung der verschiedenen Funktionen des webODV „Data Extractor“ und „Data Explorer“ Services, die für die Analyse und Visualisierung der MOSAIC-Daten verwendet

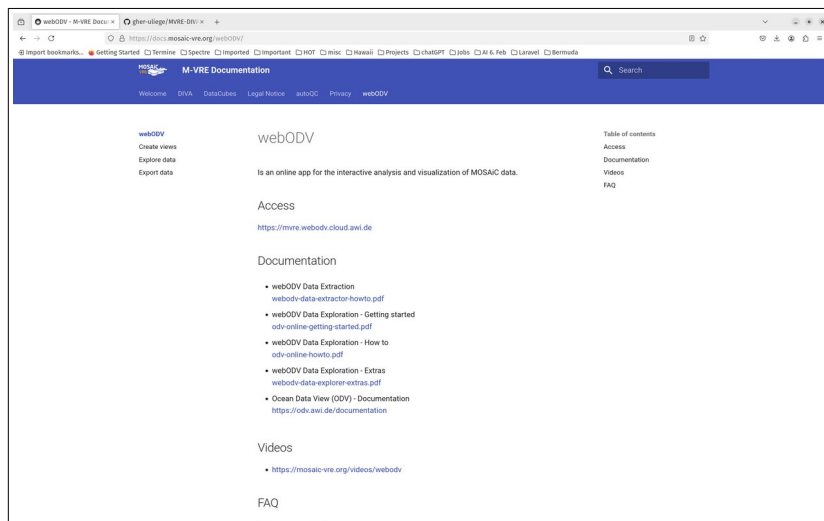


Abbildung 6: webODV Dokumentation.

werden können. Die Entwicklungen zur Aggregation von MOSAIC-Datensätzen werden vorangetrieben und einige Datenkollektionen können bereits vollautomatisiert von PANGAEA in webODV integriert werden, jedoch sind viele Datensätze aufgrund von nötigen Spezialkonvertierungen nicht sofort verfügbar. PANGAEA wird als das zentrale Repository für die Integration der MOSAIC-Daten genutzt. Es wurde festgestellt, dass die Daten oft nicht einheitlich

abgelegt sind, was die Aggregation erschwert. Um sinnvoll mit den Daten arbeiten zu können, müssen diese in ein einheitliches Format gebracht werden, was zusätzlichen Aufwand erfordert. Bis November 2022 waren etwa 2150 MOSAiC-Datensätze in PANGAEA identifiziert, von denen ca. 1160 integriert wurden. Nach der offiziellen Freigabe der Daten im Januar 2023 konnten etwa 3000 Datensätze identifiziert und ca. 2100 in webODV integriert werden. Die Integration verzögert sich aufgrund von Spezialkonvertierungen und Verzögerungen bei Laborarbeiten und Qualitätschecks, insbesondere im Bereich Bio-Geochemie. Die ursprünglich in Betracht gezogene Pan2Applic-Software kann nicht verwendet werden, da die MOSAiC-Daten zu heterogen abgelegt sind, somit wird unser PANGAEA2ODV weiter ausgebaut zu einem zuverlässigen Kern-Tool in M-VRE. Weiterhin wird die M-VRE-Homepage unter <https://mosaic-vre.org> fertiggestellt, und bietet einen wichtigen Startpunkt für neue Nutzer\*innen (Abb. 7).

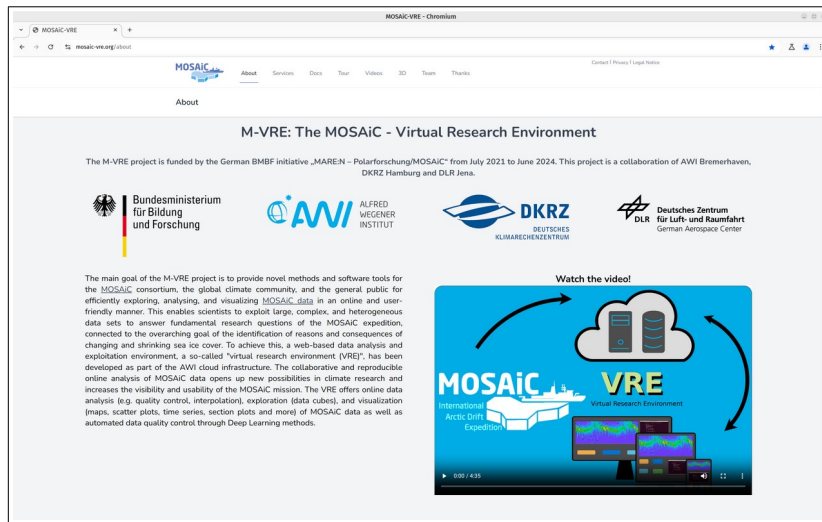


Abbildung 7: Die M-VRE Homepage

Hier sind auch die webODV Video Tutorial (<https://mosaic-vre.org/videos/webodv>) verfügbar, die im Rahmen von M-VRE produziert wurden. Abbildung 8 zeigt einen Ausschnitt aus einem Video Tutorial.

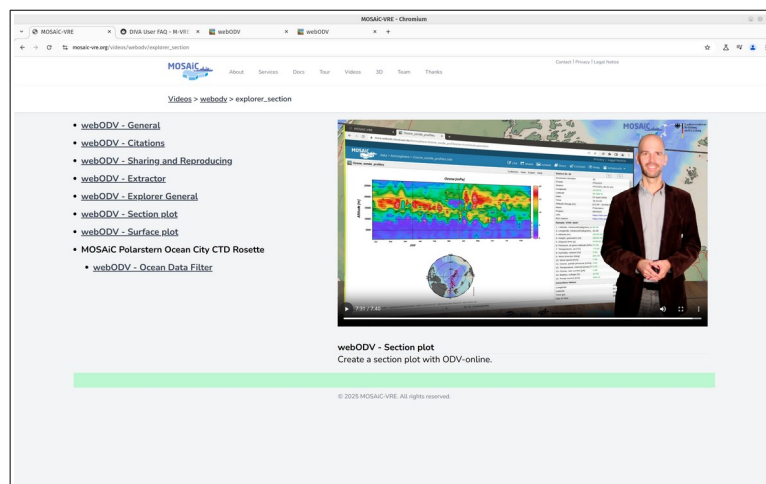


Abbildung 8: webODV Video Tutorials.

Zusätzlich wurden in Zusammenarbeit mit dem MOSAiC Team Ocean und Prof. Reiner Schlitzer (Entwickler der ODV Software) dreidimensionale (3d) Animationen von ausgewählten MOSAiC

Datensätzen erzeugt, welche unter [https://mosaic-vre.org/3d/pot\\_temp](https://mosaic-vre.org/3d/pot_temp) verfügbar sind. Diese Animationen kamen besonders bei MOSAiC Konferenzen als spannende Ausstellungsobjekte zum Einsatz. Abbildung 9 zeigt ein Beispiel für Temperatur-Messungen entlang der Drift.

Im Januar 2023 wurden die MOSAiC-Daten in PANGAEA und webODV öffentlich zugänglich gemacht. Seitdem ist die Anzahl der veröffentlichten Datensätze in PANGAEA gestiegen: Im November 2022 gab es etwa 2.150 Datensätze, von denen 1.160 für die Integration in webODV geeignet waren. Im Januar 2023 stieg die Zahl auf etwa 3.019 Datensätze, von denen 2.118 integrierbar waren. Bis Ende November 2023 wurden ca. 4.121 Datensätze veröffentlicht, von denen etwa 2.786 für die Integration in webODV geeignet sind. Abbildung 10 zeigt den Anstieg der veröffentlichten MOSAiC-Datensätze und die Anzahl der für webODV geeigneten Daten seit März 2022.

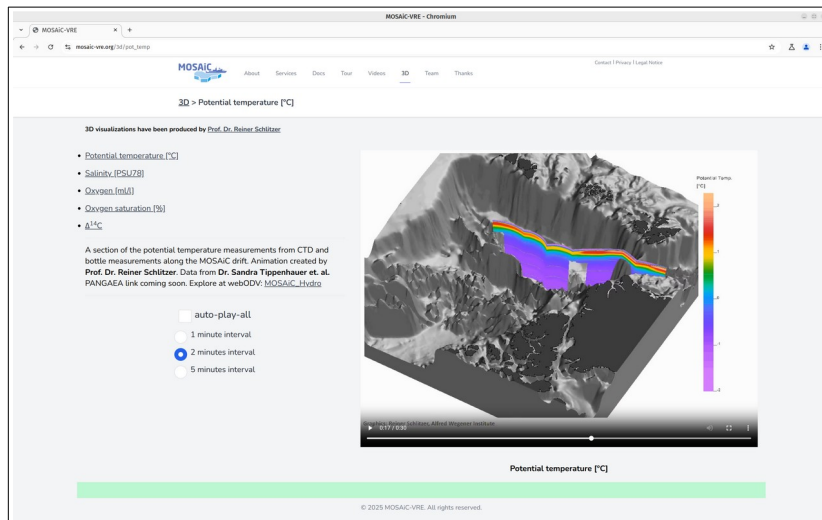


Abbildung 9: 3d Animation der MOSAiC Daten.

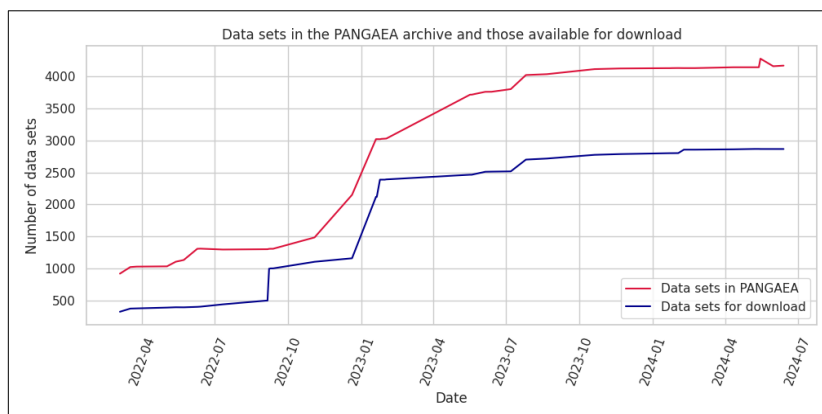


Abbildung 10: Anzahl der in PANGAEA veröffentlichten MOSAiC-Datensätze seit März 2022 bis Juli 2024. Die rote Linie stellt die in PANGAEA veröffentlichten Daten dar und die blaue Linie die Daten, die zur Integration in webODV geeignet sind.

Für die MOSAiC-Daten müssen spezielle Konvertierungsskripte erstellt werden, um die Daten in ein lesbares ODV-Format zu konvertieren. Aufgrund der unterschiedlichen Formate und Strukturen der gespeicherten Daten erfordert dies verschiedene Ansätze und zusätzlichen Aufwand. Im Laufe des Jahres wurden diese Skripte in den Datenkonvertierungsworkflow integriert.

Angesichts des raschen Anstiegs der MOSAiC-Daten in PANGAEA wurde der Konvertierungsworkflow optimiert. Ein teilautomatischer Prozess wurde entwickelt, der das Herunterladen, die Aggregation und die Konvertierung nur für neu veröffentlichte oder geänderte Datensätze durchführt. Zudem wurde ein Skript implementiert, das den DataType und das DataField jedes Datensatzes definiert, um die Strukturierung der Daten für die Konvertierung zu erleichtern.

Zusätzlich wurden Informationsdateien für jede Datensammlung in webODV erstellt, die eine Zusammenfassung der veröffentlichten Datensätze enthalten und den Benutzer\*innen als erste Informationsquelle dienen. Diese Dateien sind über ein '?' neben dem Titel der Datensammlung in webODV zugänglich. Abbildung 11 zeigt einen Ausschnitt der webODV Startseite mit den Listen der vorhandenen Datenkollektionen und den Informations Pop-Up Fenstern. Hier haben wir den Original Abstract der Datensätze aus PANGAEA integriert, und so können Nutzer\*innen sofort erste Informationen zu den Datensätzen bekommen.

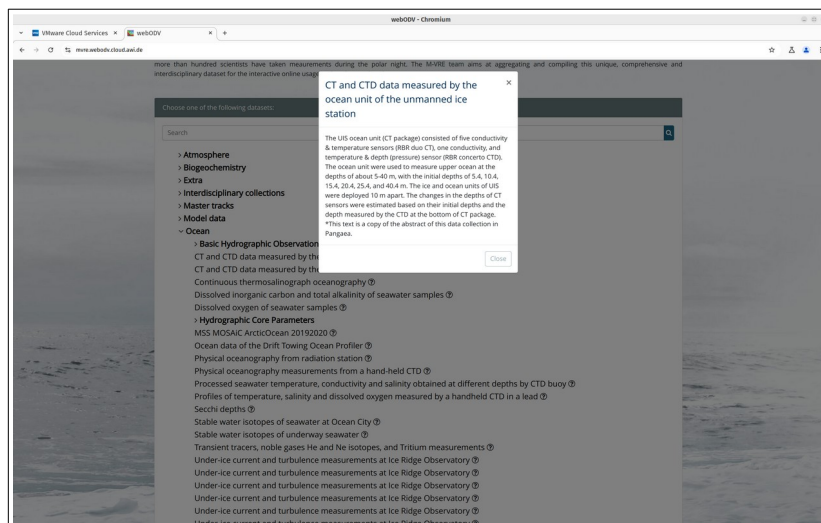


Abbildung 11: Infobox für einen MOSAiC Datensatz in webODV.

M-VRE erhebt den Anspruch ein virtuelles Forschungslabor zu sein, mit den zwei Kern-Tools webODV und Jupyter (inkl. DIVAnd) sowie autoQC. Ein essentieller Aspekt dabei ist, dass die MOSAiC Daten sowohl in webODV als auch in Jupyter zugänglich sind. Hier profitieren wir von den drei verschiedenen Ausgabeformaten von ODV, wobei wichtig zu erwähnen ist, dass die Struktur in allen drei Formaten identisch ist. Für optimalen Zugriff liegen die Daten im webODV als .odv Kollektionen vor, das binäre, optimierte ODV-interne Format. Im Jupyter Notebook liegen die Daten als einfache .txt Formate vor. Dies hat den Vorteil, dass die Daten leicht gelesen werden können und so auch einfach mit anderen Tools interoperabel sind. Zusätzlich bieten wir die Daten im netCDF Format an, ein weit verbreitetes (besonders in der Klimamodellierungs-Community) binäres Format, welches auch für das DIVAnd Tool genutzt wird.

Um neuen Benutzer\*innen ohne Python-Erfahrung den Einstieg zu erleichtern, wurden Jupyter Notebooks erstellt, die Anleitungen zum Lesen der Daten sowie zur Durchführung grundlegender statistischer Analysen und zur Erstellung von Visualisierungen bieten. Diese Notebooks dienen als Vorlagen für Wissenschaftler\*innen und können leicht ausgebaut werden. Abbildung 12 zeigt eines der Vorlagen Notebooks zum Einlesen der MOSAiC Daten.

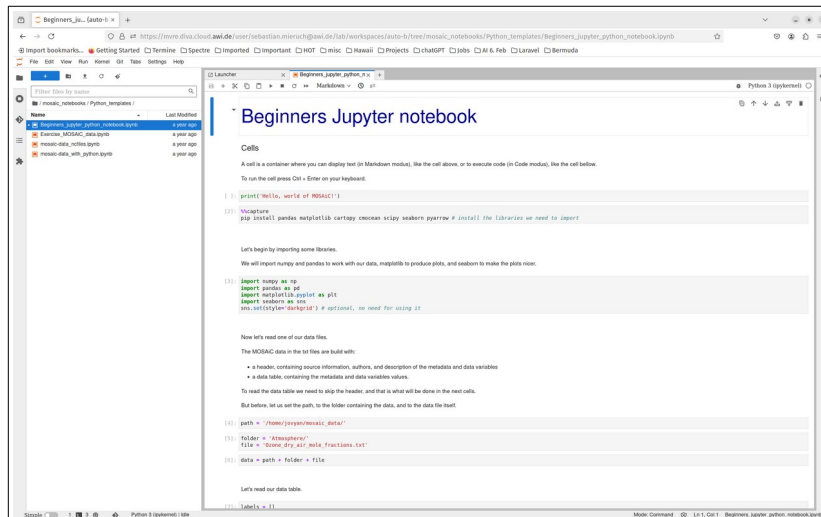


Abbildung 12: Eines, der von uns erstellten Vorlagen-Notebooks für den einfachen Einstieg in die Analyse der MOSAiC Daten.

Ein weiterer wichtiger Punkt in M-VRE ist die Entwicklung und Implementierung von Deep-Learning-Modellen zur semiautomatischen Qualitätskontrolle von ozeanographischen Messdaten, insbesondere für Temperatur- und Salinitätsmessungen. Im Rahmen des Projekts wurde eine NVIDIA Tesla V100S 32GB Grafikkarte (GPU) angeschafft, um die Deep-Learning-Modelle zu entwickeln. Es werden verschiedene Python-Frameworks verwendet, darunter TensorFlow, Keras und scikit-learn, um die Modelle zu implementieren und zu trainieren. Der erste Schritt im Machine-Learning-Prozess besteht darin, die Daten so aufzubereiten, dass sie von den Algorithmen verarbeitet werden können. Dies umfasst die Bereinigung und Strukturierung der Daten, um sicherzustellen, dass sie für das Training der Modelle geeignet sind. Der von uns entwickelte und vorhandenen SalaciaML-Algorithmus (<https://salacia-ml.awi.de/>) wurde an die neue GPU-Infrastruktur angepasst, und erste Trainingszyklen wurden mit dem UDASH-Datensatz (<https://doi.pangaea.de/10.1594/PANGAEA.872931>) durchgeführt. Der UDASH-Datensatz enthält qualitativ hochwertige Temperatur- und Salzgehaltsdaten für den Arktischen Ozean und die nördlichen subpolaren Meere, die von 1980 bis 2015 gesammelt wurden. Es wurden vergleichende Analysen mit verschiedenen Machine-Learning-Algorithmen durchgeführt. Dabei kam die Optimierung der Hyperparameter zur Anwendung, um die Leistung der Modelle zu verbessern. Zu den untersuchten Algorithmen gehören:

- Multilayer Perceptron (MLP)
- Support Vector Machine (SVM)
- Random Forest (RF)
- Naive Bayes (NB)
- K-Nearest Neighbors (KNN)

Als Ergebnis des Vergleichs der verschiedenen Algorithmen hat sich das MLP, also das Neuronale Netzwerk, als stärkstes Tool etabliert, wobei mit diesem Typ von Deep Learning Algorithmus fortgefahren wird.

Im weiteren Verlauf wird eng mit Dr. Myriel Vredenburg, einer Ozeanographin und Expertin im Bereich Qualitätskontrolle von Ozeandaten zusammengearbeitet, um das maschinelle Lernsystem und den UDASH-Datensatz zu evaluieren. Wichtige weitere Treffen und Diskussionen werden mit Dr.

Axel Behrendt, dem Erzeuger des UDASH Datensatzes abgehalten sowie mit Dr. Steffen Seitz von der Universität Dresden, mit dem seit 2019 regelmäßige Treffen zu KI Themen stattfinden. Die Tests konzentrierten sich auf die Erkennung von Temperaturgradientenfehlern in Ozeandaten. Die Tests zeigten die Wirksamkeit des Systems, identifizierten jedoch auch Bereiche, die verbessert werden müssen, insbesondere die Reduzierung von falsch positiven und falsch negativen Ergebnissen. Ziel der automatisierten Qualitätskontrolle ist die operationelle Implementierung eines Web-Services, den wir „autoQC“ nennen. Nach dem Training, der Validierung und dem Testen unserer Systeme haben wir die benutzerfreundliche Webschnittstelle SalaciaML-arctic entwickelt, um den Zugriff auf unseren KI-Algorithmus zu erleichtern. Die Online-App, die unter <https://mvre.autoqc.cloud.awi.de/> verfügbar ist, ermöglicht es Benutzer\*innen, ihre Daten einfach zu verarbeiten und bietet umfassende Dokumentation. Verarbeitete Daten können als .csv-Dateien exportiert werden. Die App verarbeitet die Daten im Hintergrund und benachrichtigt die

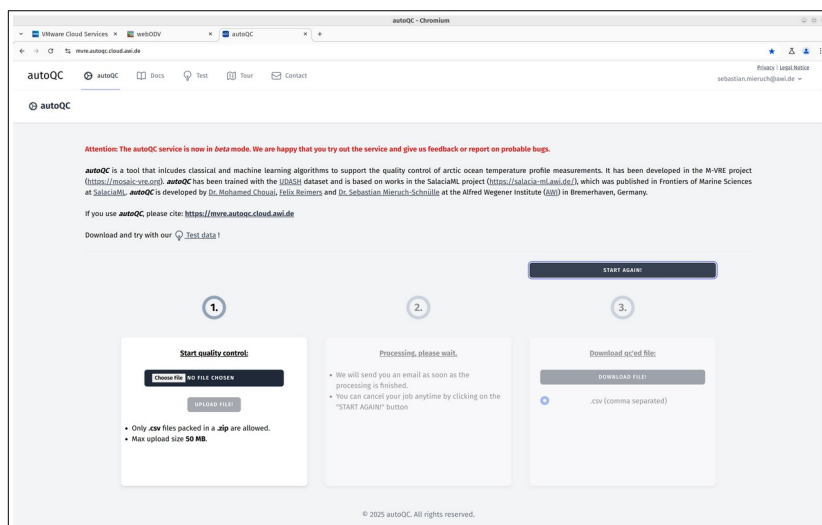


Abbildung 13: autoQC App unter <https://mvre.autoqc.cloud.awi.de/>.

Benutzer\*innen per E-Mail, sobald die Qualitätskontrolle abgeschlossen ist. Der finale Code ist auf GitHub (<https://github.com/mchouai27/SalaciaML-Arctic>) veröffentlicht, und alle Codeblöcke sind in einer Python-Datei namens 'all\_in\_one.py' zusammengefasst, die die Vorverarbeitung, Verarbeitung und Nachbearbeitung der Daten übernimmt. Im Zuge der Datenoptimierung für unsere KI Anwendung wurde eine neue Version des UDASH Datensatzes erzeugt und auf PANGAEA veröffentlicht (<https://doi.pangaea.de/10.1594/PANGAEA.973235>). Mit der freien Verfügbarkeit des Algorithmus und der Daten unterstützen wir aktiv die FAIR-Datenprinzipien, z.B. zur Reproduktion unserer Ergebnisse. Abbildung 13 zeigt einen Screenshot der autoQC App.

## 2. Wichtigste Positionen des zahlenmäßigen Nachweises

### **0838 Reisekosten**

Die Reisekosten wurden vollständig ausgeschöpft und sogar noch aus 0847 etwas erhöht. Die Reisen zu nationalen und internationalen Konferenzen und Meetings waren ein essentieller Teil des M-VRE Projektes. Ganz besondere Bedeutung hatte die Teilnahme an den drei MOSAiC Konferenzen, 2022 in Potsdam, 2023 in Boulder und 2024 wieder in Potsdam. Diese Konferenzen waren nicht direkt eingeplant, konnten aber anstatt geplanter Reisen realisiert werden. Während der MOSAiC Konferenzen hat das M-VRE Team eine wesentliche Rolle zum Thema Datenauswertung und Visualisierung gespielt. Durch Vorträge und intensive Workshops konnten wir die Nutzung von M-VRE vorantreiben, und wichtige Kontakte zu den MOSAiC Teams knüpfen.

Genauso wichtig waren natürlich auch unsere projektinternen M-VRE Treffen, bei denen wir typischerweise den Stand des Projektes, das weitere Vorgehen sowie direkte Arbeiten diskutiert und auch durchgeführt haben.

Weitere Konferenzen, wie z.B. zur EGU und IMDIS, waren wichtig, um im Austausch mit der Community zu bleiben und die Sichtbarkeit von M-VRE zu erhöhen.

### **0847 vorhabenspezifische Abschreibungen**

Hier wurde für das M-VRE Projekt ein Rechnersystem Dell R640 angeschafft. Zum einen konnte das System in die vorhandene AWI Cloud integriert werden, wodurch unserem Projekt exklusive Rechenressourcen zur Verfügung standen, die absolut essentiell für die Verarbeitung der MOSAiC Daten waren. Eine Hauptkomponente des Systems ist die NVIDIA Tesla V100S 32GB Grafikkarte (GPU), die etwa 10.000 € des Preises ausmacht. Ohne diese leistungsstarke GPU wäre die Realisierung und der operationelle Betrieb der autoQC App, also der KI-betriebenen Qualitätskontrolle von Ozean-Temperaturmessungen, wahrscheinlich nicht möglich gewesen. Die GPU beschleunigt das Training der KI-Algorithmen bis zum hundertfachen dessen, was standardmäßige CPU Einheiten leisten. Die Entscheidung und Zusage für den Kauf der GPU, und damit dem Exklusivzugriff darauf im AWI war eine absolut richtige und wichtige Entscheidung und hat signifikant zum Erfolg des Projektes beigetragen. In diesem Punkt konnte sogar noch etwas Geld gespart werden, welches für wichtige Reisen eingesetzt werden konnte.

### **0850 sonstige unmittelbare Vorhabenkosten**

Ursprünglich wurden hier Kosten für Publikationen eingeplant. Wir haben allerdings im Verlauf des Projektes festgestellt, dass unsere Prioritäten nicht im Veröffentlichen von wissenschaftlichen Publikationen, also Artikeln in Fachzeitschriften, ausgerichtet sind. Sondern vielmehr auf den persönlichen und/oder direkten Kontakt mit Nutzer\*innen abzielen. Deshalb haben wir entschieden die Mittel für andere Punkte zu investieren, die aus unserer Sicht relevanter für den Erfolg des Projektes waren. Dies sind zum einen die Reisen zu den MOSAiC Konferenzen und unsere dortigen Workshops. Dazu gehört z.B. auch die Anfertigung von Flyern zur Auslage. Abbildung 14 zeigt die Vorderseite des Flyers, den wir auf hochqualitativem Papier im DinA6 Format drucken lassen haben. Auf der Rückseite befindet sich ein QR-Code und die Web-Adresse <https://mosaic-vre.org/>.

Weiterhin wurden Mittel in Höhe von ca. 900 € für die Anschaffung von Material zur Produktion der Video-Tutorials eingesetzt. Abbildung 15 zeigt das Produktionsset. Wichtige Komponenten wie z.B. der Greenscreen, die Leuchten, Mikrofone und der Teleprompter waren essentiell für die professionelle Umsetzung.

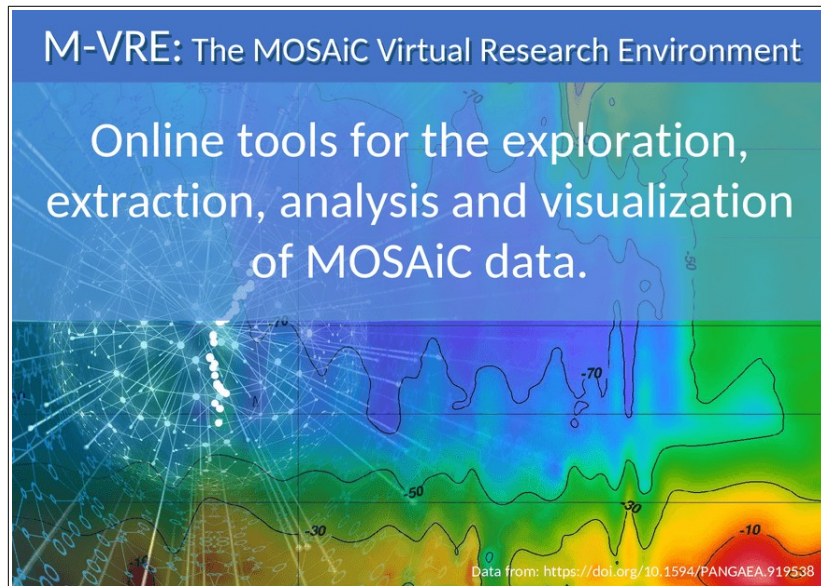


Abbildung 14: Der M-VRE Flyer, den wir auf Konferenzen und Meetings verteilt haben. Auf der Rückseite ist ein QR Code abgedruckt und die Web-Adresse: <https://mosaic-vre.org/>.



Abbildung 15: Produktion der Video-Tutorials.

### 3. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Alle Arbeiten in M-VRE waren notwendig und angemessen. Das Projektziel wurde erreicht, da schon in der Planungsphase sehr weitsichtige und realistische Abschätzungen der Aktivitäten und Möglichkeiten vorlagen. Oberstes Ziel in M-VRE war von Anfang an die echte Nutzbarkeit der online Services. Deshalb wurde auch mit Hochdruck an der möglichst schnellen Verfügbarkeit gearbeitet. Höchste Priorität hatte die Konvertierung der MOSAIC Daten von PANGAEA nach ODV. Schon im Februar 2022 konnte der webODV Service in den operationellen Betrieb gehen. Heute haben wir ein mächtiges virtuelles Forschungslabor mit den MOSAIC Daten und den Services: webODV, Jupyter (DIVAnd, Python) und autoQC zur Verfügung. Natürlich verlaufen in einem komplexen Projekt wie M-VRE nicht alle Entwicklungen absolut gradlinig und Justierungen waren nötig. Zum Beispiel die Erkenntnis, dass das MOSAIC Central Storage für unsere Anwendungen nicht nutzbar ist. Aber auch dieser Einsicht, die zunächst negative scheint, war notwendig und kostete Arbeitszeit. Schlussendlich

hat es sich herausgestellt, dass PANGAEA die für uns entscheidende Datenquelle ist. Bereits in der Antragsphase wurde gesehen, dass PANGAEA eine wichtige Rolle spielen könnte. Der Data Cube Service hat es leider nicht in den operationellen Betrieb geschafft. Gründe dafür wurden oben schon besprochen. Dennoch konnten wichtige Ergebnisse im Testbetrieb erzielt werden und Erfahrungen gesammelt werden, die in zukünftigen Forschungsvorhaben von Nutzen sein werden.

#### 4. Voraussichtlicher Nutzen und Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

##### Wirtschaftliche Verwertung:

- Stärkung der AWI Rechenzentrumsinfrastruktur: Zu Beginn des M-VRE Projektes im Sommer 2021 spielten unser GPU System, und unsere Aktivitäten im KI Bereich eine Vorreiterrolle am AWI. Zu der Zeit gründeten wir mit weiteren Kolleg:innen die „AI-Group“, eine Gruppe, die seit dem regelmäßige Treffen zum Thema KI abhält. Seit dem hat das AWI massiv in KI investiert, in Hardware und in wissenschaftliche Projekte. Wir sehen unseren Beitrag hier als einen signifikanten Schub. Im März 2024 konnten wir durch die Arbeiten in M-VRE zum Thema KI ein internes Projekt in Höhe von 90.000 € einwerben, und dort die Arbeiten zum Thema autoQC fortführen.
- Unterstützung von Helmholtz Cloud-Diensten, Initiativen uva.: Während des M-VRE Projektes haben wir immer die Nähe zu Helmholtz Aktivitäten gesucht, und uns besonders im Rahmen von HIFIS eingebracht. Dort haben wir durch unsere Erfahrungen in M-VRE ein spezielles webODV unter <https://hifis.webodv.awi.de> implementiert, zum Import von eigenen, privaten Daten, welches natürlich auch von der MOSAiC Community genutzt werden kann. M-VRE Mitarbeiter konnten zudem an Austauschprogrammen im Rahmen von HIDA teilnehmen und so wichtige Fähigkeiten erlernen. Weiterhin haben wir z.B. beim Helmholtz Imaging Contest 2024 und 2025 teilgenommen, mit einer künstlerischen Collage von webODV Visualisierungen der MOSAiC Daten (Abbildung 16)

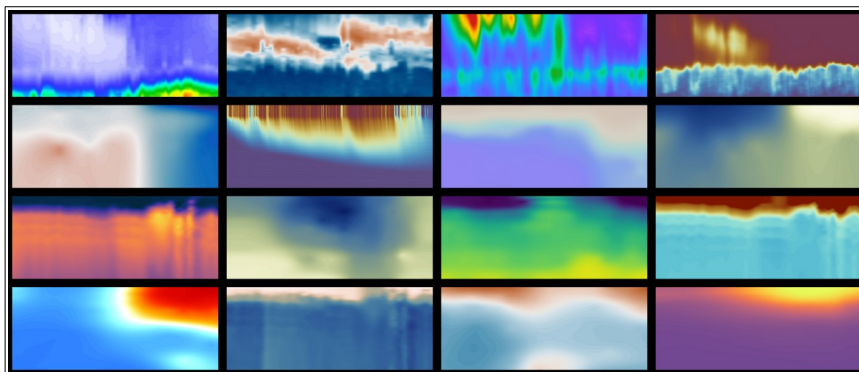


Abbildung 16: webODV Beitrag für den Helmholtz Imaging Contest.

##### Wissenschaftliche-technische Verwertung:

- M-VRE: Der operationelle Betrieb der M-VRE Services wird auch über die Dauer des Projektes also nach Juni 2024 fortgeführt. Die M-VRE Services konnten im Rahmen der Projektdauer in einen produktiven, stabilen und wartungsarmen Zustand gebracht werden. Somit stehen der wissenschaftlichen Community und Öffentlichkeit mächtige Tools frei zur Verfügung, die im Rahmen des BMBF geförderten Projektes M-VRE entwickelt wurden.

- AutoQC: Die Algorithmen des autoQC Service wurden in GitHub veröffentlicht und es wurde über Zenodo eine DOI generiert: 10.5281/zenodo.11535790.
- Eine neue Version des UDASH Datensatzes wurde auf PANGAEA veröffentlicht: <https://doi.pangaea.de/10.1594/PANGAEA.973235>.

#### Wissenschaftliche und wirtschaftliche Folgeentwicklungen:

- autoQC: Hier konnte durch die autoQC Aktivitäten ein AWI internes Projekt von 90.000 € eingeworben werden, um dieses Thema weiter voranzutreiben.
- webODV und Datenkonvertierung: Basierend auf den Arbeiten, Erfahrungen und erworbenem KnowHow konnte von Juli 2024 bis September 2024 ein Projekt mit der Universität Hawaii durchgeführt werden (ca. 50.000 €, AWI-interne Finanzierung) und ein webODV für die Hawaii Ocean Time-series aufgesetzt werden (<https://hot.webodv.awi.de>), wobei ein Schwerpunkt, wie bei MOSAiC, auf der Konvertierung der Originaldaten in das ODV Format lag.
- BlueCloud 2026: Die webODV Entwicklungen werden im EU BlueCloud 2026 weitergeführt, wobei Mittel in Höhe von 100.000 € eingeworben wurden.
- Im Rahmen von Helmholtz Enterprise und dem AWI Technologie-Transfer wurde eine „Field Study Fellowship“ für drei Monate mit Umfang 23.000 € eingeworben. Hier geht es um die Frage, ob webODV kommerziell besser verwertet werden kann.

#### Wissenstransfer:

- webODV workshops: Seit 2023 veranstalten wir jährlich ODV / webODV Workshops (zwei Halbtage) im Marum in Bremen für neue PhD Student\*innen und alle Interessierten.
- Zukunftstag: Im Jahr 2024 wurde ein Zukunftstag Projekt mit Schüler\*innen der 7. bis 9. Klassen durchgeführt, wobei MOSAiC Daten mit webODV ausgewertet wurden.
- Kunstprojekt: Seit 2022 arbeiten wir mit dem Künstlerduo Nina und Adnan Softic aus Berlin zusammen. Zusammen mit weiteren Künstler\*innen entwickelte das Team um die Softic's ein Musikinstrument, das „Klimaton“ (Abbildung 17), (<https://softic.info/klimaton-ARCTIC-2020>), welches MOSAiC Daten zur Klangerzeugung nutzt. Das Projekt nutzt webODV, um einfach und effizient MOSAiC Daten in ihr System zu integrieren.



Abbildung 17: Das Klimaton, ein Musikinstrument, angetrieben mit MOSAiC Daten aus webODV.

- Museum: Durch eine Neuausrichtung hat sich das Deutsche Museum in Bonn zu einem Forum für künstliche Intelligenz entwickelt. Wir arbeiten mit dem Deutschen Museum in Bonn im Rahmen der autoQC Aktivitäten zusammen, und konnten dort z.B. in einer Abendvorstellung über M-VRE und im besonderen über autoQC Vortragen (<https://www.deutsches-museum.de/bonn/aktuell/klimaforschung-ki-und-klangkunst>).

## 5. Während der Durchführung bekannt gewordene Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Die online Datenauswertung, die Generierung von Community Datensätzen und KI Anwendungen in der Klimaforschung sind natürlich ein sich schnell entwickelndes Gebiet. Dennoch sind uns im Projektzeitraum keine für uns signifikanten Fortschritte bei anderen Stellen bekannt geworden.

## 6. Erfolgte und geplante Veröffentlichungen des Ergebnisses

### • Webseiten:

- Die wichtigsten Veröffentlichungen bzw. Anlaufstellen für MOSAiC Datennutzer\*innen sind unsere M-VRE Webseite <https://mosaic-vre.org>, sowie die Services:
  - webODV: <https://mvre.webodv.cloud.awi.de>
  - DIVAnd: <https://mvre.diva.cloud.awi.de>
  - autoQC: <https://mvre.autoqc.cloud.awi.de>

### • Daten- und Softwarepublikationen:

- Chouai, Mohamed; Mieruch-Schnülle, Sebastian; Behrendt, Axel; Vredenburg, Myriël; Rabe, Benjamin (2024): UDASH-AI: Unified Database for Arctic and Subarctic Hydrography Optimized for Artificial Intelligence Applications [dataset]. PANGAEA, <https://doi.org/10.1594/PANGAEA.973235>
- mchouai27. (2024). mchouai27/SalaciaML-Arctic: SalaciaML-Arctic v1.0.0 (Zenodo). Zenodo. <https://doi.org/10.5281/zenodo.11535790> (<https://github.com/mchouai27/SalaciaML-Arctic/tree/Zenodo>)
- [https://github.com/smieruch/webodv\\_helmholtz\\_image\\_contest\\_2025](https://github.com/smieruch/webodv_helmholtz_image_contest_2025)
- [https://github.com/smieruch/mosaic\\_hydrographic\\_core\\_parameters](https://github.com/smieruch/mosaic_hydrographic_core_parameters)

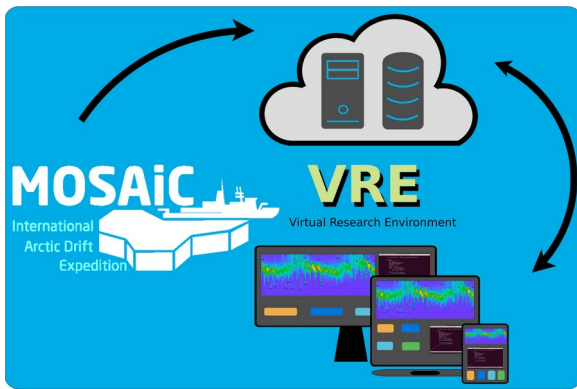
### • Publikationen in wissenschaftlichen Zeitschriften:

- Schulz, K., Koenig, Z., Muilwijk, M., Bauch, D., Hoppe, C. J. M., Droste, E. S., Hoppmann, M., Chamberlain, E. J., Laukert, G., Stanton, T., Quintanilla-Zurita, A., Fer, I., Heuzé, C., Karam, S., Mieruch-Schnülle, S., Baumann, T. M., Vredenburg, M., Tippenhauer, S., and Granskog, M. A. (2024): The Eurasian Arctic Ocean along the MOSAiC drift in 2019-2020: An interdisciplinary perspective on physical properties and processes. *Elementa: Science of the Anthropocene* (2024) 12 (1): 00114, doi: [10.1525/elementa.2023.00114](https://doi.org/10.1525/elementa.2023.00114) .

### • Konferenzbeiträge:

- Linck Rosenhaim, I., Mieruch-Schnülle, S., and Schlitzer, R.: Data preparation for the development of a user-friendly, free, online, and interactive platform for the visualization and analysis of interdisciplinary data, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-14352, <https://doi.org/10.5194/egusphere-egu23-14352>, 2023.
- Mieruch, S., Linck Rosenhaim, I., and Schlitzer, R.: The MOSAiC webODV: Interactive online data exploration, visualization and analysis, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-13807, <https://doi.org/10.5194/egusphere-egu23-13807>, 2023.
- Schoening, Timm, Ulrike Kleeberg, and Stephan Frickenhaus. "Data Science Symposium 2023-Book of abstracts." GEOMAR, 2023 ([https://oceanrep.geomar.de/id/eprint/58897/1/DataScienceSymposium8-GEOMAR\\_book\\_of\\_abstracts.pdf](https://oceanrep.geomar.de/id/eprint/58897/1/DataScienceSymposium8-GEOMAR_book_of_abstracts.pdf))
- Chouai, M., Simon Reimers, F., and Mieruch-Schnülle, S.: Deep learning to support ocean data quality control , EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-15185, <https://doi.org/10.5194/egusphere-egu23-15185>, 2023.

- Kreps, Gastón & Mieruch-Schnülle, Sebastian & Chouai, Mohamed. (2024). autoQC: An AI based online app for ocean data quality control. AGU 2024
- Chouai, Mohamed & Mieruch-Schnülle, Sebastian & Kreps, Gastón. (2024). AutoQC: An AI Based Online App for Ocean Data Quality Control. AGU 2024
- Simoncelli, S., Vernet, M., & Coatanoan, C. (2024). International Conference on Marine Data and Information Systems - Proceedings Volume. Miscellanea INGV, 80. <https://doi.org/10.13127/MISC/80>
- **Andere:**
  - Klimaton:
    - <https://softic.info/klimaton-ARCTIC-2020>
    - <https://calls.ars.electronica.art/2023/prix/winners/8208/>
    - <https://www.silent-green.net/en/programme/detail/2023/12/10/klimaton-listening-to-the-disappearing-landscape>
    - <https://www.experimenta.fr/exposition-xpa/klimaton-arctic%E2%89%882020/>
  - Deutsches Museum Bonn:
    - <https://www.deutsches-museum.de/bonn/aktuell/klimaton-klanginstallation-jetzt-im-museum-erleben>



### Teil III Erfolgskontrollbericht

Förderkennzeichen: 03F0888A

Zuwendungsempfänger: Alfred-Wegener-Institut Helmholtz Zentrum für Polar- und Meeresforschung, Projektleitung: Dr. Sebastian Mieruch-Schnülle

Verbundprojekt: MOSAIC 2 - M-VRE: Online Services zur Erkundung, Analyse und Visualisierung von MOSAIC Daten

Vorhaben: Implementierung interdisziplinärer Analysen

Laufzeit: 01.07.2021 – 31.12.2024

Berichtszeitraum: 01.07.2021 – 31.12.2024

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 03F0888A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen / den Autoren.

#### 1. **Wissenschaftlich-technisches Ergebnis des Vorhabens, eingereichte Nebenergebnisse und gesammelte Erfahrungen**

Das M-VRE Projekt ist mit dem Anspruch angetreten ein virtuelles Forschungslabor zu entwickeln, welches die Analyse, Auswertung und Visualisierung sowie die Verbreitung und effiziente Nutzung der MOSAIC Daten unterstützt. Ziel war nicht ein prototypisches Experiment, sondern ein operationelles und nutzerfreundliches System schnell aufzubauen und so signifikant die Forschung zu unterstützen. Dabei kann das M-VRE System auf zwei Ebenen betrachtet werden, die Arbeiten im Hintergrund (Backend), die den Nutzer\*innen verborgen bleiben, und die Entwicklungen im Vordergrund (Frontend), also die sichtbaren online Services, Tools, Videos, Dokumentation usw. Für die Realisierung eines solchen Systems wurde ein hochqualifiziertes Team aus verschiedenen Expert\*innen zusammengestellt, welche auf <https://mosaic-vre.org/team> vorgestellt werden, und so wurden tiefgreifende Erfahrungen und Kenntnisse in Computerwissenschaften, Datenwissenschaften, Statistik, Webprogrammierung, Physik, Mathematik, Geowissenschaften uva. zusammengebracht. Zusätzlich wäre ein solches Vorhaben nicht möglich ohne die Unterstützung eines Rechenzentrums, in unserem Fall das AWI Rechenzentrum, also den Mitarbeitenden in den Bereichen Cloud Computing, Storage, Netzwerk, GPU und mehr. Weiterhin war der Kontakt und die direkte Unterstützung und Zusammenarbeit mit den MOSAIC Teams essentiell, um die webODV, Jupyter und autoQC Services zu entwickeln.

Die technischen Entwicklungen im Backend, also besonders die Datenaufbereitung, Aggregation, Harmonisierung und die damit verbundene Softwareentwicklung, aber auch die Implementierung der Services in der AWI Cloud, die zusammen das Kernstück und die Grundlage des Projektes bilden, haben das M-VRE Team signifikant vorangebracht. So konnte Dr. Ingrid Linck Rosenhaim nach dem M-VRE Projekt einen neuen Arbeitsplatz im PANGAEA Projekt bekommen. Irfan Khan konnte sich im DKRZ behaupten und eine neue Projektstelle für weitere vier Jahre bekommen.

Aufgrund der realistischen und zielgerichteten Planung konnte das M-VRE Team den webODV Service in der frühen Projektphase bereits in den operationellen Betrieb nehmen, die Jupyter (DIVAnd) und autoQC Services folgten etwas später. So konnten wir der MOSAiC Community, der Öffentlichkeit und weiteren Nutzer\*innen, z.B. dem Klimaton Team, die M-VRE in der aktuellen und sehr aktiven Phase der Auswertung der MOSAiC Daten anbieten. Wichtig war hier die Durchführung von Workshops und die Erstellung von Video Anleitungen, um Nutzer\*innen die Einführung in die Tools so einfach wie möglich zu gestalten. Durch diese gewonnene Expertise und Erfahrung konnte Dr. Sebastian Mieruch weitere Projekte und Drittmittel einwerben, z.B. HOT (50.000 €), SalaciaML (90.000 €) und BlueCloud (100.000 €). Dr. Mohamed Chouai konnte eine Stelle an der Technischen Hochschule in Ingolstadt bekommen.

Während der Hochphase der Auswertung der MOSAiC Daten konnten wir bis zu 100 Nutzer\*innen pro Monat in der webODV App „sehen,“ wobei sich jetzt, fünf Jahre nach der Expedition die Nutzerzahlen bei ca. 30 pro Monat einzupendeln scheinen. Es wird sich zeigen, wie lange MOSAiC Daten noch analysiert werden, aber sicher ist, dass webODV und die Jupyter Services eine der nutzerfreundlichsten Anlaufstellen für die Analyse, schnelle und einfache Exploration, Visualisierung und Download von MOSAiC Daten darstellen.

Natürlich stellt sich die Frage, was besser hätte gemacht werden können in M-VRE. Generell ist M-VRE aus unserer Sicht ein Erfolg. Die von uns erzeugte Auswahl bzw. Aggregation von MOSAiC Daten in einer einheitlichen, übersichtlichen, interoperablen Form ist eine immense Wertschöpfung, die der Öffentlichkeit zur freien Verfügung steht, genauso wie die freie Nutzbarkeit der Services. Das M-VRE Team war sehr gut aufgestellt und hat eine große Bandbreite an Aktivitäten abgedeckt, von der sehr technischen Backend-Entwicklung über nutzerfreundliche Frontend-Entwicklung bis zur Produktion von Video Tutorials und intensiven „in person“ Workshops und Konferenzbeiträgen. Allerdings hätte in M-VRE aus unserer Sicht ein Aspekt intensiver ausfallen können, die direkte Interaktion mit den MOSAiC Teams. In diesen Punkt haben wir schon das Maximum investiert, hatten aber im Nachhinein betrachtet nicht genug Ressourcen dafür eingeplant. Ob diese Ressourcen finanziert worden wären ist natürlich auch im Nachhinein nicht sicher zu beantworten. Angenommen wir hätten eine extra FTE, also Mitarbeiter\*in im Projekt gehabt, die 100% für den Kontakt, die Beratung und Begleitung der MOSAiC Teams während der Datenauswertung zuständig gewesen wäre, dann hätten wir definitiv bessere Hilfestellung und Unterstützung bereitstellen können. Unsere Mitarbeiter\*in hätte also die verschiedenen MOSAiC Teams direkt begleitet und hätte sofort Probleme und Schwierigkeiten bei der Datenakquisition und Analyse erkannt und versucht mit unseren Tools zu helfen. Neue Tools haben immer eine gewisse Barriere und Nutzer\*innen müssen entscheiden, ob sie genug Benefit erzielen, wenn sie Zeit und Arbeit investieren, neue Tools kennenzulernen. Diese Barriere zu überwinden ist die größte Herausforderung in einem Projekt wie M-VRE. Die Hauptaufgabe der Mitarbeiter\*in für Kontakt, Beratung und Begleitung, die wir nicht hatten, wäre gewesen die MOSAiC Teams zu unterstützen, so dass mehr Nutzer\*innen die Barriere der M-VRE Nutzung überwinden. Damit es nicht falsch verstanden wird, M-VRE ist ein sehr nutzerfreundliches System und es gibt wahrscheinlich kein anderes Tool, welches hoch komplexe Visualisierungen generieren kann und gleichzeitig so einfach und schnell zu erlernen ist wie ODV / webODV. Dennoch, von alleine können weder webODV noch Jupyter, Klimaanalysen generieren und ein gewisser Arbeitseinsatz ist nötig, um die Bedienung der Tools zu erlernen, der sich aber definitiv auszahlt.

## **2. Fortschreibung des Verwertungsplans**

### **2.1. Erfindungen/ Schutzrechtsanmeldungen und erteilte Schutzrechte**

Keine.

### **2.2. Wirtschaftliche Erfolgsaussichten nach Projektende**

- Wie im Teil II des Berichtes beschrieben gibt es folgende wirtschaftliche Ergebnisse und Erfolge:
  - autoQC, AWI internes Projekt, 90.000 €
  - HOT, 50.000 €
  - BlueCloud 2026, 100.000 €
  - Helmholtz Enterprise, 23.000 €
- Weiterhin tragen die ehemaligen M-VRE Mitarbeitenden die erworbenen Kenntnisse und Fähigkeiten weiter und so profitieren weitere deutsche Institute und im Endeffekt die Gesellschaft.
- Die webODV und Jupyter Services werden nach Projektende vom AWI und DKRZ weiter betrieben.
- Im Rahmen von Transferstrategien soll hier noch einmal auf das Helmholtz Enterprise „Field Study Fellowship“ Projekt hingewiesen werden. In diesem Projekt wird, zusammen mit dem AWI Technologie-Transfer, versucht mögliche Anwendungsfelder in Industrie und Wirtschaft von webODV, Datenanalyse und Cloud Computing zu erkunden.

### **2.3. Wissenschaftliche und/oder technische Erfolgsaussichten nach Projektende**

Die M-VRE Services sind auch nach dem Ende des Projektes verfügbar und werden so lange wie möglich im operationellen Modus betrieben. Somit bietet M-VRE neben PANGAEA (<https://pangaea.de/>) und US Archiven wie z.B. ARM (<https://www.arm.gov/data/>) MOSAiC Daten an. Wie stark MOSAiC Daten in Zukunft genutzt werden ist im Moment nicht abzusehen. Dennoch ist davon auszugehen, dass die Daten noch Jahrzehnte genutzt werden. M-VRE hat den Anspruch hier eine signifikante Rolle bei der Bereitstellung der Daten zu spielen, wobei webODV den einfachsten, schnellsten und effizientesten Zugang, im Vergleich mit anderen, bietet.

Im Rahmen des autoQC Teilprojektes haben wir im AWI eine wichtige Stellung im Vorantreiben von Methoden der künstlichen Intelligenz eingenommen. Hier sehen wir großes Zukunftspotential und einen signifikanten Beitrag.

### **2.4. Wissenschaftliche und wirtschaftliche Anschlussfähigkeit**

Im April 2024 haben wir einen Antrag für eine zweite Phase von M-VRE unter dem Kennzeichen 03V01866 eingereicht. Kernziel war, weitere MOSAiC Daten in das M-VRE System, also webODV und Jupyter einzupflegen, sowie die Community stärker zu unterstützen. Bisher haben wir keine Entscheidung darüber erhalten. Dennoch hat das Dokument heute die gleiche Wichtigkeit und Gültigkeit wie vor einem Jahr. Es gibt in PANGAEA einige interessante Datensätze, die bisher nicht in das M-VRE System übertragen wurden. Zusätzlich wären Optimierungen und Verbesserungen der aggregierten Daten-Kollektionen sinnvoll.

Ebenso wäre eine erneute Aufnahme des Kontaktes mit den MOSAiC Teams sinnvoll, um gezielt Projekte zur Datenauswertung mit unseren Tools zu unterstützen.

### **3. Arbeiten, die zu keiner Lösung geführt haben**

Aufgrund guter Planung konnten Arbeiten ohne Lösung minimiert werden. Schnell wurde klar, dass das MOSAiC Central Storage System nicht geeignet ist, um die Daten Kollektionen in Arbeitspaket 2 zu erzeugen. Hier hatten wir von Beginn an das PANGAEA Archiv eingeplant, welches zur Grundlage der Datenquelle wurde. Ebenso schnell hat sich gezeigt, dass das Tool Pan2Applic nicht genutzt werden kann und wir konnten voll auf das PANGAEA2ODV Tool setzen. Geplante Ideen für eine webODV API, also einen programmatischen Zugang zu webODV, wurden frühzeitig verworfen, da in der Community dafür kein Bedarf war. In Zusammenarbeit mit dem DKRZ konnte die einfache containerbasierte Lösung schnell umgesetzt werden, als klar wurde, dass Kubernetes zu komplex und nicht nötig war.

### **4. Einhaltung der Ausgaben- und Zeitplanung**

Alle geplanten Investitionen und Nutzungen der Mittel wurden eingehalten, darunter fallen die Kosten für das Rechnersystem, die Reisen und Material für Video Produktion und Workshops.

Die Zeitplanung wurde eingehalten, MOSAiC Daten wurden gleich zu Beginn des Projektes aggregiert und harmonisiert, um diese schnell bereit zu stellen. Die Services wurden schnellstmöglichst in den operationellen Modus überführt. Bis Dezember 2022 wurden die M-VRE Services, also der Zugriff auf die Daten, über einen Login geschützt, so dass nur die MOSAiC Teams im Zeitraum des Daten Moratoriums Zugriff hatten. Dieser Login wurde im Januar 2023 aufgelöst und webODV ist seit dem auch ohne Login zugänglich. Die Jupyter und autoQC Services benötigen einen Login, da Nutzer\*innen z.B. im Jupyter einen eigenen privaten Workspace haben, und den autoQC Service betreiben wir mit Login, um missbräuchlichen Nutzen zu minimieren.

Durch den krankheitsbedingten Ausfall von Arne Osterthun vom DLR wurde eine kostenneutrale Verlängerung des M-VRE Projektes bis zum 31.12.2024 beantragt und genehmigt. In dieser Zeit konnten auch die AWI und DKRZ Teams in kleinerem Umfang an der Stabilisierung und Zuverlässigkeit der Services arbeiten, um für die Zukunft ein möglichst wartungsarmes System betreiben zu können.