
Die Veröffentlichung von Standardisierten Daten aus der Stadtklimaforschung

Anette Ganske¹, Vivien Voss², Amandine Kaiser³, Angelika Heil³ und Andrea Lammert³

¹Technische Informationsbibliothek (TIB)

²Universität Hamburg, Meteorologisches Institut, CEN

³Deutsches Klimarechenzentrum (DKRZ)

Eine Grundvoraussetzung für die Wiederverwendbarkeit von wissenschaftlichen Daten ist die Veröffentlichung der Daten in einer Art und Weise, die garantiert, dass sie auffindbar, zugänglich, untereinander vergleichbar und weiterverwendbar sind. Dies entspricht der Einhaltung der FAIR-Prinzipien.

Innerhalb des Projekts AtMoDat wurde ein Konzept erstellt, wie atmosphärische Modell-daten idealerweise für die Nachnutzbarkeit beschrieben werden sollten. Voraussetzungen hierfür sind die Nutzung des Datenformats NetCDF, die Anwendung des international etablierten Climate and Forecast (CF) Metadatenstandards und die Publikation der Daten mit einem DataCite DOI. Basierend auf dem neu erstellten ATMODAT Standard wurde ein Tool zur Überprüfung der Konformität von Daten und Metadaten, der *atmodat data checker*, entwickelt, welcher über GitHub mit freier Lizenz verfügbar ist.

Neben der klassischen Klimamodellierung gibt es viele Bereiche der atmosphärischen Modellierung, welche noch nicht im CF Metadatenstandard enthalten sind, wie z.B. die mikroskalige Stadtklimamodellierung. Durch die Nutzung des ATMODAT Standards lassen sich nun auch in diesen Bereichen standardisierte Datensätze erstellen. Dies wird exemplarisch anhand des Stadtklimamodells MITRAS vorgestellt.

Nach der Publikation der geprüften Daten können diese mit EASYDAB (Earth System Data Branding) hervorgehoben werden. EASYDAB ist eine neu entwickelte Qualitätskennzeichnung, die potentiellen Datennutzern anzeigt, dass es sich um einen geprüften Datensatz handelt, der einfach wiederzuverwerten ist. Dies wird durch das EASYDAB-Logo auf der Landing Page des Datensatzes angezeigt. Die innerhalb des AtMoDat-Projekts entwickelten Verfahren der Datenstandardisierung und Qualitätskennzeichnung können einfach auf Daten anderen Fachrichtungen übertragen werden und dort zu einer verbesserten Nachnutzbarkeit der Daten beitragen.

1 Einleitung

Klimamodelle sind komplexe Computerprogramme, welche Abschätzungen darüber liefern, wie sich das Klima unter bestimmten Annahmen verändern wird. Der Austausch

von Klimamodelldaten ist weit über die Klimaforschungsgemeinschaft von großer Bedeutung. Klimamodelle liefern sehr große Datenmengen, welche so gespeichert und zugänglich gemacht werden müssen, dass eine Nachnutzung möglichst effizient erfolgen kann. Im Rahmen des Coupled Model Intercomparison Project (CMIP) [1] arbeiten weltweit hunderte Klimaforschungseinrichtungen daran, koordinierte Simulationen mit globalen Erdsystem- bzw. Klimamodellen durchzuführen, die neuesten Modellergebnisse untereinander auszutauschen, zu vergleichen und zu analysieren. Die CMIP-Modellexperimente dienen als wissenschaftliche Basis für die Sachstandsberichte des Weltklimarats IPCC¹ (Intergovernmental Panel on Climate Change). Die IPCC-Berichte tragen den aktuellen Kenntnisstand über den menschengemachten Klimawandel zusammen und bieten Grundlagen für wissenschaftsbasierte klimapolitische Entscheidungen.

Um den Datenaustausch und die vergleichende Datenauswertung zu erleichtern, wurden für CMIP genaue Vorgaben für das Dateiformat, die Metadaten und die Datendokumentation gemacht und angewendet [2]. Auch wurde eine einheitliche Namensgebung der Variablen und Dateien vereinbart. Das CMIP Standard-Dateiformat ist NetCDF² (Network Common Data Form), ein binäres, systemunabhängiges Format, das sich besonders für gegitterte Daten eignet, und das – da es Daten zusammen mit den Metadaten abspeichern kann – selbstbeschreibend ist. Für CMIP wurde festgelegt, dass die Dateien die international vereinbarten *NetCDF Climate and Forecast Metadata Conventions*³ (im Folgenden CF Conventions genannt) erfüllen sollen. Ein zentrales Element der CF Conventions ist die CF-Standardnamentabelle. Standardnamen sind genau definierte Bezeichner für Größen; sie werden Variablen über das *standard_name* Attribut zugeordnet und machen sie so eindeutig und maschinenlesbar. Die CF-Standardnamentabelle schreibt auch für jeden Bezeichner eine physikalische Einheit vor und sie enthält eine kurze, menschenlesbare Erklärung, wie die einem Standardnamen zugeordnete Größe definiert ist. Derzeit umfasst die Tabelle etwa 4500 Einträge aus verschiedenen geowissenschaftlichen Bereichen. Zusätzlich definieren die CF Conventions wie räumliche und zeitliche Achsen von abgespeicherten Daten zu hinterlegen sind. Weiterhin wurden innerhalb von CMIP Richtlinien erarbeitet, wie die Daten technisch und inhaltlich geprüft werden und in welcher Art und Weise die Ergebnisse der Qualitätsprüfung für den Datennutzer sichtbar gemacht werden.

Im Projekt AtMoDat⁴ (Atmosphärische Modelldaten) wurde untersucht, ob die in CMIP angewandten Standards systematisch auf weitere Bereiche der Meteorologie und Klimaforschung übertragen werden können. So gibt es eine Vielzahl kleinerer, CMIP unterstützender Modellvergleichsprojekte (MIPs), wie z.B. die Initiative Aerosols, Clouds, Precipitation and Climate (ACPC)⁵. Eine vollständige Anpassung des CMIP Standards auf solch kleinere MIPs wäre sehr aufwendig und zu kostspielig, so dass die Entwicklung eines reduzierten Standards als sinnvoll erachtet wurde. Weiterhin gibt es den Bereich der Stadtklimaforschung. Hier gibt es bisher keinen etablierten Datenstandard. Daten aus

¹<https://www.ipcc.ch/>

²<https://doi.org/10.5065/D6H70CW6>

³<http://cfconventions.org>

⁴<https://www.atmodat.de/>

⁵<http://www.acpcinitiative.org/>

der Stadtklimaforschung sind im Vergleich zu regionalen oder globalen Klimamodelldaten sehr hoch aufgelöst und haben oft eine andere Raumbitterstruktur. Zudem beinhalten sie Variablen, die spezifisch für die Stadtklimaforschung sind, wie z.B. Wärmeflüsse an Gebäudewänden. Um Daten aus der Stadtklimaforschung nachnutzbar veröffentlichen zu können, müssen fachspezifische Standards neu entwickelt werden.

Innerhalb des AtMoDat Projekts wurden Kurationskriterien [3] formuliert, auf deren Basis die Vollständigkeit der Datensätze, die Kohärenz und Konsistenz von Daten- und Metadaten sowie die portalübergreifende Auffind- und Nutzbarkeit der Daten im Zuge des Archivierens und Teilens über Repositorien sichergestellt wird. Die im AtMoDat Projekt entwickelten Standardisierungen und Kurationskriterien sind auf andere Bereiche der Klimaforschung übertragbar; sie können die Reproduzierbarkeit und Überprüfbarkeit deutlich erhöhen und auch die interdisziplinäre Nachnutzung von Klimamodelldaten unterstützen. Diese Kriterien werden im ATMODAT Standard [4] zusammen gefasst. In diesem Standard folgen wir den FAIR-Datenprinzipien⁶ [21], welche fordern, dass Daten so veröffentlicht werden sollten, dass sie auffindbar, zugänglich, interoperabel und wiederverwendbar sind. Zudem wurde im Rahmen des AtMoDat Projekts das Earth System Data Branding⁷ (EASYDAB) entwickelt, um qualitativ hochwertige Datensätze deutlicher aus der Menge anderer Daten hervorzuheben.

Der ATMODAT Standard wird in Abschnitt 2 näher erläutert. In Abschnitt 3 werden die Resultate der Anwendung des Standards auf die Daten der Stadtklimaforschung beschrieben und in Abschnitt 4 wird EASYDAB detaillierter erklärt.

2 Der ATMODAT Standard

Der ATMODAT Standard ist eine Qualitätsrichtlinie für die FAIRe Veröffentlichung von atmosphärischen Modelldaten mit offenen Lizenzen. Er setzt eine Publikation der Daten mit einem DataCite DOI in einem Repository⁸ voraus. Zudem muss der DOI, wie von DataCite⁹ empfohlen, stets mit einer sogenannten Landing Page verknüpft sein. Der DOI kann für einzelne Datensätze vergeben werden, d.h. für je eine Datei mit wissenschaftlichen Daten und ihre Metadaten. Alternativ kann er auch einer ganzen Datensatz-Kollektion gemeinsam zugewiesen werden, die aus mehreren Datensätzen besteht, deren Daten z.B. mit einer einzigen Modellsimulation berechnet wurden.

Um die FAIRness der veröffentlichten Daten zu ermöglichen, werden im ATMODAT Standard Vorgaben für die Dateien, die Metadaten und die Landing Pages gemacht. Zudem gibt der Standard vor, dass alle veröffentlichten Daten als NetCDF Dateien vorliegen.

⁶FAIR=Findable, Accessible, Interoperable, Reusable. <https://doi.org/10.1038/sdata.2016.18>

⁷<https://www.easydab.de>

⁸<https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/repositorien/>

⁹<https://datacite.org/>

2.1 Vorgaben für die Landing Page

Wird eine DOI in einem konventionellen Webbrowser aufgerufen, führt diese zu einer menschenlesbaren HTML-Internetseite, die immer auch einen maschinenlesbaren Text, den Seitenquelltext, enthält. Diese Seite wird Landing Page genannt. Die Landing Page wird vom Repositorium, das die Daten archiviert hat, bereitgestellt und enthält grundlegende Metadaten. Die Landing Page kann in Unterseiten gegliedert werden, um sie übersichtlicher zu machen.

Die Vorgaben von DataCite für die Landing Page sind in Best Practices für DOI Landing Pages¹⁰ beschrieben. Auf der Landing Page muss im menschenlesbaren Format immer eine zitierbare vollständige bibliographische Angabe (incl. des DOI) über dem Datensatz selbst stehen, so dass er eindeutig vom Menschen identifiziert werden kann. Zudem soll die DOI im maschinenlesbaren Teil der Landing Page so gekennzeichnet werden, dass Suchmaschinen sie finden können.

Weiterhin muss eine Landing Page immer Informationen darüber enthalten, wie man auf die Daten zugreifen kann. Falls der Datensatz selbst nicht mehr existiert, muss dies auf der Landing Page vermerkt werden (Tombstone Page).

2.2 Vorgaben für die Metadaten

Metadaten müssen die Daten so beschreiben, dass potentielle Datennutzer in der Lage sind zu entscheiden, ob diese Daten für ihre Anwendung von Nutzen sind. Dafür müssen auch der Prozess der Datenerzeugung und die verwendeten Eingangsdaten dokumentiert werden.

Für die Datenveröffentlichung mit einem DataCite DOI werden in den Dateien, beim DOI und auf der Landing Page Metadaten benötigt: Alle notwendigen Informationen können entweder direkt in die Metadaten geschrieben werden oder es kann mit Links auf externe Dokumente verwiesen werden - vorzugsweise über persistente Identifikatoren (PIDs). Ein solches externes Dokument könnte z.B. die Dokumentation des numerischen Modells sein, das zur Berechnung der Daten verwendet wurde. Es wird dringend empfohlen, dass die Metadaten sowohl in maschinenlesbarer als auch maschineninterpretierbarer Form hinterlegt sind. Dies gewährleistet, dass aus den Metadaten automatisierte Listen, z.B. für Auswertungen von Institutionen erstellt werden können. Externe Dokumente selbst, zumindest aber deren Metadaten, sollten ebenfalls maschinenlesbar sein.

Der ATMODAT Standard gibt die folgenden Prinzipien vor, um die Maschinenlesbarkeit und -interpretierbarkeit und damit die FAIRness aller Metadaten zu ermöglichen:

- Alle Angaben zu Personen und Institutionen sollten möglichst immer mit einem PID ergänzt werden, d.h. einem ORCID für Personen oder einem ROR für Institutionen (falls vorhanden). Vorschläge für geeignete PIDs findet man in [6].

¹⁰<https://support.datacite.org/docs/landing-pages>

- Alle Links auf Dokumente, Homepages usw. sollten persistent angegeben werden, z.B. Dokumente mit einem DOI und Homepages vorzugsweise mit einer URN.
- Alle zeitlichen Informationen müssen in standardisierter Form angegeben werden, z.B. nach ISO 8601 [7] oder ISO 19108 [8].
- Das räumliche Referenzsystem muss immer angegeben werden, z.B. WGS84 [9].
- Schlagwörter, Orte und Arbeitsgebiete sollten aus kontrollierten Vokabularen (CVs) entnommen werden, z.B. geografische Namen aus geonames¹¹.

2.2.1 Datei-Metadaten

In den CF Conventions werden Spezifikationen für Namen, Einheiten und andere Parameter vorgegeben, um die Dateiinhalte zu vereinheitlichen und sie automatisch verarbeitbar zu machen. Deshalb schreibt der ATMODAT Standard vor, dass die CF Conventions (ab Version 1.4) für die Beschreibung aller Daten verwendet werden. Jedoch werden in den CF Conventions nicht für alle Variablen von atmosphärischen Modellen die sogenannten *standard_names* vorgegeben, wie z.B. für spezielle Variablen von Stadtklimamodellen (siehe Kapitel 3). In diesem Fall kann den Variablen ein eigener *long_name* zuweisen werden, welcher die Vorschriften der CF Conventions für die Bildung von *long_names* befolgt.

Neben der Einhaltung der CF Conventions fordert der ATMODAT Standard zusätzlich, dass die NetCDF-Dateien eine Beschreibung der Zeit-, Koordinaten- und vertikalen Achsen sowie bestimmte globale Attribute enthalten (siehe [4], Tabelle 11).

2.2.2 DOI-Metadaten

DOI-Metadaten werden für den DOI an DataCite übermittelt und beschreiben die veröffentlichten Daten. Falls ein DOI für mehrere Datensätze (Datensatz-Kollektion) zusammen vergeben wurde, beschreiben die DOI-Metadaten die Datensatz-Kollektion. Die DOI-Metadaten werden in die Metadatenfelder des DataCite Metadatenschemas [10] eingetragen.

Neben den allgemeinen Grundsätzen für alle Metadaten gibt es im ATMODAT Standard zusätzliche Empfehlungen für diese Metadaten:

- Die DataCite-Metadatenfelder *Contributor* und *Creator* sind Pflichtfelder, welche ausgefüllt werden müssen, damit Zusammenfassungen über die Publikation eines einzelnen Forschers, aller Forscher einer Institution oder aller Publikationen innerhalb eines Projekts automatisch zusammengestellt werden können. Aus den gleichen Gründen wird dringend empfohlen, auch die Förderung unter *FundingReference* anzugeben (falls vorhanden).

¹¹<https://www.geonames.org>

- Alle Zeitangaben über die Erstellung oder Veröffentlichung des Datensatzes/der Datensatz-Kollektion sind wichtig für den Datennutzer. Sie werden über das Metadaten-Feld *Date* und den zugehörigen Unterfeldern *dateType* erfasst. Eine Ausnahme bildet das Veröffentlichungsjahr, welches in dem Feld *PublicationYear* angegeben wird.

Damit die Daten selbsterklärend beschrieben sind, wird dringend empfohlen, alle für die entsprechende Fachdisziplin sinnvollen Metadatenfelder aus dem DataCite Metadaten-schema zu verwenden.

2.2.3 Landing Page-Metadaten

Der ATMODAT-Standard schreibt vor, dass alle Metadatenfelder, die für den DOI angegeben werden, auch auf der Landing Page aufgelistet werden. Dabei müssen die Bezeichnungen der einzelnen Metadatenfelder nicht übernommen werden. Zusätzlich sollen bei Datensatz-Kollektionen oder bei Datensätzen mit mehreren Variablen auch die Beschreibungen der einzelnen Datensätze bzw. Variablen auf der Landing Page stehen. Falls die Datensätze/Datensatz-Kollektionen vor der Veröffentlichung geprüft wurden, sollte das Ergebnis der Prüfung ebenfalls auf der Landing Page vermerkt werden. Diese Prüfungen können ganz unterschiedlich definiert sein: so können z.B. Messdaten auf fehlerhafte Einträge geprüft werden. Es können aber auch Metadaten auf ihre Vollständigkeit oder auf Einhaltung einer Qualitätsrichtlinie geprüft werden.

2.3 Überprüfung der Einhaltung der Vorgaben

Der ATMODAT Standard richtet sich an Repositorien (Bereich Datenkuration), aber auch an die Wissenschaft (Bereich Datenproduktion) und enthält Checklisten, mit denen direkt nach der Datenproduktion und während der Datenkuration die Einhaltung des Standards geprüft werden kann. Zur Erleichterung und vor allem zur Automatisierung dieses Prüfprozesses auch für große Datenmengen wurde der *atmodat data checker* entwickelt. Der Checker ist ein modular aufgebautes Python-Programmpaket, welches mit freien Lizenzen auf Github¹² veröffentlicht wurde. Somit ist der *atmodat data checker* für alle Interessierten nutzbar und hilfreich bei der Produktion und Publikation standardisierter Daten. Ein Beispiel seiner Anwendung wird in Kapitel 3.2 gezeigt. Ein entsprechender Checker für die Prüfung der DOI-Metadaten ist in Arbeit.

3 Anwendung des Standards auf Daten der Stadtklimaforschung

Klima- und Atmosphärenmodelle, die bei CMIP verwendet werden und an die CF Conventions angepasst sind, werden über sehr große Regionen angewendet wie z.B. die ganze

¹²https://github.com/AtMoDat/atmodat_data_checker

Erde oder ganz Europa. Die horizontale Auflösung der Modellgitter liegt üblicherweise zwischen 50 und 200 km. Die Modelle rechnen typischerweise in Zeitschritten von Minuten bis Stunden und die Simulationen umfassen Zeiträume von Jahren bis Jahrtausenden. Diese Eigenschaften sind für die Simulation kleinskaliger Prozesse, wie z.B. in der Stadtklimamodellierung, nicht geeignet, denn solch grobe Gitterzellen und Zeitschritte können die städtische Struktur nicht darstellen und kurzlebige Prozesse nicht auflösen. Im Rahmen von AtMoDat Projekts soll eine Erweiterung des CF Standards an die speziellen Anforderungen der mikroskaligen Stadtklimamodellierung angeregt werden.

3.1 Besonderheiten der Stadtklimamodellierung

Die Stadtklimaforschung beinhaltet die Untersuchung von Phänomenen und Prozessen, die im urbanen Raum eine Rolle spielen. Themen in der Stadtklimaforschung sind beispielsweise Effekte des Klimawandels im urbanen Raum [12] oder die Verbesserung der Lebensqualität innerhalb von Städten durch Studien zur Ventilation, Ausbreitung von Luftschadstoffen (z.B. [13]) oder Temperaturverteilungen innerhalb Straßenschluchten [14]. Die Untersuchung solcher Effekte erfolgt zum Einen über die Erhebung und Auswertung von Messdaten (z.B. [12] und [13]), zum Anderen über die Verwendung von Modellen, die solche Prozesse simulieren können.

Modelle, die atmosphärischen Prozesse im städtischen Raum simulieren sollen, müssen Strukturen und Besonderheiten der Stadt darstellen können. Verwendet werden hierbei sogenannte mikroskalige Modelle, wie zum Beispiel ENVI-MET¹³, MISKAM¹⁴, MITRAS [15] oder PALM-4U [16]. Die dabei verwendeten Modellgebiete decken Bereiche in der Größenordnung weniger Quadratkilometer ab. Meist umfassen die Gebiete typische Stadtstrukturen, wie zum Beispiel einzelne Stadtteile oder einzelne Straßenzüge. Die Gitterweiten betragen wenige Metern und somit können Hindernisse wie Gebäude oder Bäume dargestellt werden. Die zeitliche Auflösung der Simulationen beträgt wenige Sekunden, so dass kurzlebige Prozesse gut aufgelöst werden können. Insgesamt umfasst eine Simulation meist nur Stunden oder wenige Tage. Beispielsweise lassen sich so Zirkulationen in einzelnen Straßenschluchten oder die Ausbreitung von Schadstoffen innerhalb der Stadt simulieren.

Zur Veranschaulichung der hohen räumlichen Auflösung der mikroskaligen Stadtklimamodelle zeigen wir hier eine Simulation mit MITRAS für die Hamburger Innenstadt (Abbildung 1). Die damit berechneten Windgeschwindigkeiten werden in Abbildung 2 gezeigt. Dabei wurde eine horizontale Gitterweite von 2.5 m verwendet. In der Simulation wurde in rund 1 km Höhe eine südwestliche Anströmung mit einer Windgeschwindigkeit von 3 m/s vorgegeben. Abbildung 2 zeigt die simulierten Windgeschwindigkeiten in vier verschiedenen Höhenschichten, von bodennah (2a) bis zur maximalen Höhe der Gebäude im Gebiet (2d). Es lässt sich deutlich erkennen, wie Gebäude das Windfeld in allen drei Raumrichtungen (horizontale Ebene und in der Vertikalen) beeinflussen. In Bodennähe

¹³<http://www.envi-met.net/documents/papers/geosim2001.pdf>

¹⁴https://www.lohmeyer.de/site/assets/files/4559/manual_miskam_63.pdf

blockiert eine enge Bebauung die Strömung, was sich in niedrigen Windgeschwindigkeiten äußert. Auf Plätzen und breiteren Straßen treten höhere Windgeschwindigkeiten auf. Mit zunehmender Höhe blockieren immer weniger Gebäude den Wind; in 72,5 m Höhe (2d) haben nur noch die Kirchturmspitzen und Hochhäuser direkten Einfluss auf das Windfeld, was sich in der niedrigeren Windgeschwindigkeit im Nachlauf der Strömung widerspiegelt.

Für spezifische Eigenschaften der Stadt, wie die Hindernisse und dazugehörige Variablen, z.B. Wandtemperaturen, Strahlungsflüsse an Gebäudeoberflächen oder Niederschlag auf Dächern, werden in den bisherigen CF Conventions noch keine Vorgaben zur Standardisierung gemacht. Daher beschäftigt sich ein Teil des AtMoDat Projektes damit, typische Variablen aus der Stadtklimaforschung zu identifizieren und ihre Aufnahme in die CF Conventions zu beantragen.

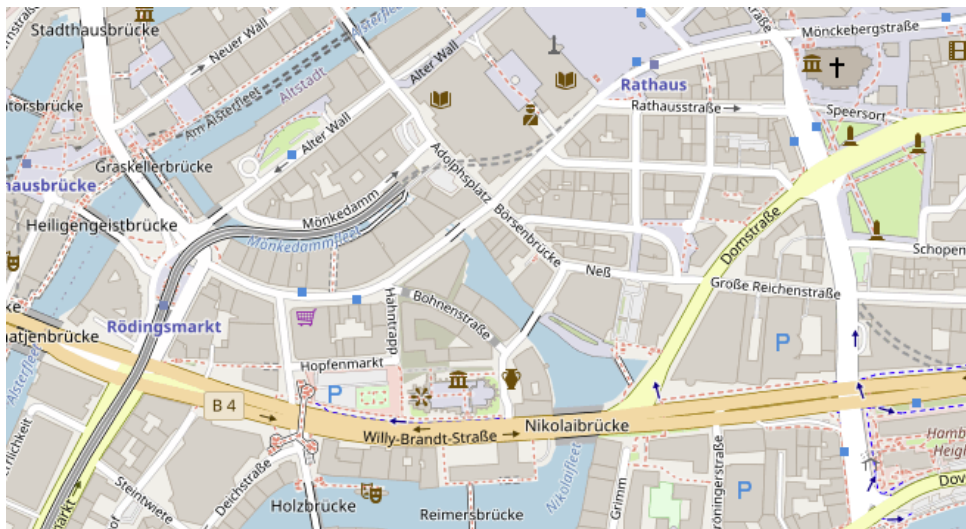


Abbildung 1: Ausschnitt der Hamburger Innenstadt, der für die Simulation mit dem Modell MITRAS verwendet wird. Karte: www.openstreetmap.org

3.2 Anwendung des atmodat data checker

Erste Tests zur Anwendung und Umsetzung des ATMODAT Standards wurden mit Modellausgaben von MITRAS durchgeführt. MITRAS schreibt die Simulationsergebnisse in binäre Dateien, die mit Hilfe eines Post-Processors in NetCDF-Dateien umgewandelt werden. Bisher fehlten dem Post-Processor die Attribute *long_name* und *standard_name*, die die eindeutige Beschreibung und Zuordnung der Variablen definieren. Beide Attribute wurden in den Post-Processor eingebaut. Variablen, die noch nicht in den CF Conventions geführt werden, erhalten zunächst nur das Attribut *long_name*. Weiterhin wurden zusätzliche Metadaten nach Empfehlungen aus dem ATMODAT Standard zur besseren Nachnutzbarkeit hinzugefügt.

Die Datensätze wurden mit Hilfe des *atmodat data checkers* (siehe Abschnitt 2.3) auf die Konformität mit dem ATMODAT-Standard überprüft. Der Checker überprüft neben der

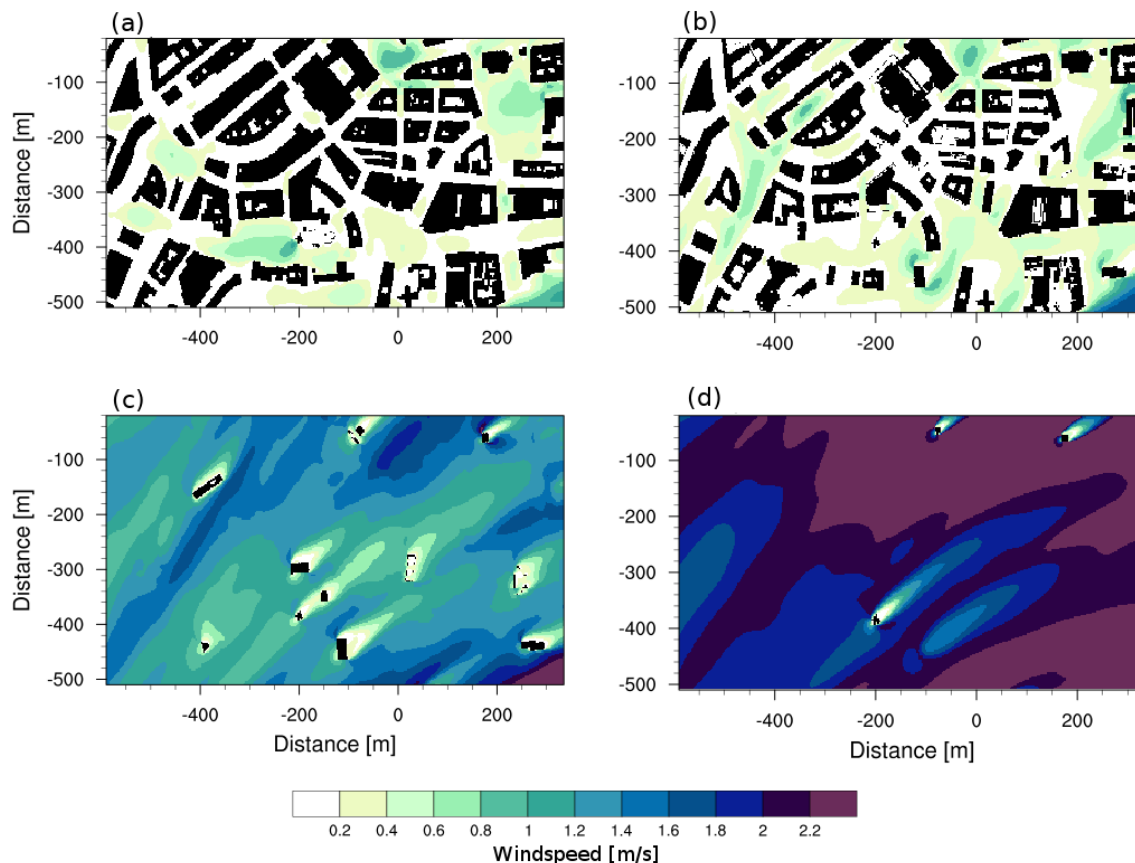


Abbildung 2: Darstellung des Windes in der Hamburger Innenstadt am 21.06.2000 um 4:30 Uhr morgens, simuliert mit dem Modell MITRAS. Farbige Flächen zeigen verschiedene Windgeschwindigkeiten, die schwarzen Flächen beschreiben die Gebäude in dem Modellgebiet. Die Abbildungen zeigen die Windgeschwindigkeiten in (a) 2,5 m, (b) 22,5 m, (c) 48,5 m und (d) 72,5 m Höhe.

Einhaltung der CF Conventions die Dateien auch auf lückenhafte Metadatenangaben und fehlerhafte Dateien. Werden Fehler entdeckt, gibt der Checker eine Meldung aus, in der zwischen Fehlern und Warnungen differenziert wird. Zudem enthält der Output klare Hinweise zur Korrektur der Fehler. Bei den Metadaten unterscheidet der Checker zwischen zwingend erforderlichen (mandatory), empfohlenen (recommended) und sonstigen optionalen Metadatenangaben. Der Output des *atmodat data checkers* ist maschinenlesbar, was z.B. eine automatische Auswertung bei größeren Datenmengen ermöglicht.

Um den *atmodat data checker* anzuwenden zu können, muss er lokal auf dem Rechner installiert werden. Anweisungen zur Installation und zur Anwendung des Checkers sind auf Github¹⁵ hinterlegt. Die Software lässt sich problemlos auf verschiedenen Betriebssystemen installieren und ist einfach anzuwenden. Im Test mit MITRAS-Ausgabedateien konnte der *atmodat data checker* einwandfrei alle fehlerhaften NetCDF-Dateien, die zum Beispiel keinen *standard_name* oder *long_name* enthielten, identifizieren. Des Weiteren

¹⁵https://github.com/AtMoDat/atmodat_data_checker

gibt der *atmodat data checker* hilfreiche Anweisungen, wie die Metadaten der Dateien um im ATMODAT Standard empfohlene Attribute ergänzt werden können, um so die Nachnutzung der Daten zu erleichtern.

4 EASYDAB

Mit dem Earth System Data Branding (EASYDAB) können spezielle Datensätze aus der Erdsystemforschung, die mit einem DataCite DOI veröffentlicht wurden, hervorgehoben werden: Das EASYDAB Logo auf der Landing Page zeigt an, dass die Datensätze eine offene Lizenz haben, den FAIR Data Principles genügen und vom zuständigen Repository auf Einhaltung einer Qualitätsrichtlinie geprüft wurden. So können z.B. Datensätze, die die Qualitätsanforderungen des ATMODAT Standards einhalten, mit EASYDAB publiziert werden. Alternativ können Repositorien für eine EASYDAB Veröffentlichung auch auf eigene Qualitätsrichtlinien zurückgreifen, falls mit diesen eine vergleichbar hohe Qualität bei den veröffentlichten Daten erreicht wird, wie das im ATMODAT Standard definiert wird. Die Bedingungen für die Qualitätsrichtlinien findet man unter <https://www.easydab.de/about-easydab/easydab-guideline>.

Das EASYDAB Logo ist geschützt. Um es verwenden zu dürfen, müssen Repositorien einen Vertrag mit der Technischen Informations Bibliothek (TIB)¹⁶ abschließen. In diesem verpflichten sich die jeweiligen Repositorien, dass sie das EASYDAB Logo nur auf den Landing Pages von den Datensätzen/Datensatz-Kollektionen zeigen, bei denen die EASYDAB Richtlinien eingehalten werden.

Mithilfe des EASYDAB Logos können Repositorien anzeigen, dass sie Datensätze unter Berücksichtigung der FAIRness Prinzipien sorgfältig kuratieren und nachnutzbar machen. Forschende können gut beschriebene Datensätze leichter finden, evaluieren und für sie relevante Daten nutzen. In Zusammenarbeit mit dem World Data Center for Climate¹⁷ (WDCC) werden erste Datensätze aus der Stadtklimaforschung und von einem kleineren Modellvergleichsprojekt (MIP) zur Veröffentlichung mit EASYDAB vorbereitet.

5 Fazit und Ausblick

Klimamodelldaten sind eine unverzichtbare Grundlage für Klimaforscher und politische Entscheidungsträger. Klimamodelldaten geben zum Beispiel Aufschluss darüber, wie sich atmosphärische Prozesse auf globaler oder regionaler Ebene mit einem Klimawandel verändern; sie geben auch Hinweise, welche Anpassungsmaßnahmen zum Schutz urbaner Gebiete sinnvoll wären.

¹⁶siehe www.easydab.de

¹⁷<https://www.dkrz.de/up/systems/wdcc>

Die Klima- und Atmosphärenforschung hat umfangreiche Standardisierungen bezüglich der Dateiformate, Metadatengestaltung und Dokumentation etabliert, um die Daten von globalen und regionalen Klimamodellberechnungen zu vereinheitlichen und automatisiert auswertbar zu machen. Im Rahmen des AtMoDat Projekts werden diese Standardisierungen systematisch auf Daten hochauflösender Modelle angepasst. Der hierbei entwickelte ATMODAT Standard wird, wie in diesem Beitrag beschrieben, unter anderem an Daten des Stadtklimamodells MITRAS erprobt.

Nach erfolgreicher Anwendung erfolgt die exemplarische Veröffentlichung erster Datensätze inklusive DataCite DOI im Langzeitarchiv des WDCC. Dies hat besonders im Bereich der Stadtklimaforschung Pionier- und Vorbildcharakter, da im WDCC bisher keine standardisierten Daten aus dieser Disziplin gespeichert werden. Die entwickelten Standardisierungsverfahren lassen sich leicht auf weitere Disziplinen übertragen und, wenn erforderlich, erweitern. Sie stellen somit einen wesentlichen Beitrag dar, die Kompatibilität von Datenstandards sowohl innerhalb einzelner Fachrichtungen als auch disziplinübergreifend zu verbessern und hierdurch den intra- und interdisziplinären Datenaustausch erheblich zu erleichtern.

Um qualitativ hochwertige Datensätze wie solche, die dem ATMODAT Standard entsprechen, in Repositorien deutlich erkennbar hervorheben zu können, wurde EASYDAB, das Earth System Data Branding entwickelt. Das geschützte EASYDAB Logo hilft Nutzern, diese Datensätze schnell zu erkennen und auszuwählen. Der einfache Zugang zu standardisierten, qualitätsgeprüften Forschungsdaten ermöglicht nicht nur eine effizientere Nachnutzung. Er führt auch zu einem erhöhtem Vertrauen der Nutzer in das Repository, das diese Daten zur Verfügung stellt, und so wird mit dem EASYDAB Branding auch die Wertigkeit des DataCite DOIs gestärkt. Das längerfristige Ziel ist, dass qualitätsgesicherte Datenpublikationen, wegen ihrer hohen Vorteile, zunehmend von Nutzern sowie Förderorganisationen gewürdigt und eingefordert werden.

Acknowledgements

Das AtMoDat Projekt wurde im Rahmen vom “Forschungsvorhaben zur Entwicklung und Erprobung von Kurationskriterien und Qualitätsstandards von Forschungsdaten” finanziert vom Deutschen Bundesministerium für Bildung und Forschung (BMBF; FKZ: 16QK02A, 16QK02B, 16QK02C, 16QK02D). Wir danken dem DataCite Team, der CF Conventions Community und anderen AtMoDat Partnern für fruchtbare Diskussionen.

Literaturverzeichnis

- [1] Eyring, V. , S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer and K. E. Taylor. “Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6)

- experimental design and organization.” Geoscientific Model Development, Band 9, Heft 5 (2016): 1937-1958. <https://doi.org/10.5194/gmd-9-1937-20161937-1958>
- [2] Juckes, M., K.E. Taylor, P.J. Durack, B. Lawrence, M.S. Mizielinski, A. Pamment, J.-Y. Peterschmitt, M. Rixen and S. S en esi. “The CMIP6 Data Request (DREQ, version 01.00.31).” Geosci. Model Dev. Band 13, Heft 1 (2020): 201–224. <https://doi.org/10.5194/gmd-13-201-2020>
- [3] Ganske, A., D. Heydebreck, H. H ock, A. Kraft, J. Quaas and A. Kaiser. “A short guide to increase FAIRness of atmospheric model data.” Meteorologische Zeitschrift Band 29, Heft 6 (2020): 483-491. <http://dx.doi.org/10.1127/metz/2020/1042>
- [4] Ganske, A., A. Kraft, A. Kaiser, D. Heydebreck, A. Lammert, H. H ock, H. Thiemann, V. Voss, D. Grawe, B. Leitl, K. H. Schl unzen, J. Kretzschmar and J. Quaas. “AT-MODAT Standard (v3.0).” World Data Center for Climate (WDCC) at DKRZ(2021). https://doi.org/10.35095/WDC/atmodat_standard_en_v3_0
- [5] Wilkinson, M. D., et al. “The FAIR Guiding Principles for scientific data management and stewardship.” Scientific Data, Band 3 (2016): 160018. <https://doi.org/10.1038/sdata.2016.18>
- [6] Madden, F., R. van Horik, S. van de Sandt, A. Lavasa and H. Cousijn. “Guides to Choosing Persistent Identifiers - Version 2.” Zenodo (2020). <https://doi.org/10.5281/zenodo.3956569>.
- [7] ISO 8601-1:2019(en). “Date and time — Representations for information interchange — Part 1: Basic rule.”(2019). <https://www.iso.org/standard/70907.html>
- [8] ISO 19108:2002(en). “ Geographic information — Temporal schema.” (2002). <https://www.iso.org/standard/26013.html>
- [9] Department of Defense. “World geodetic system 1984 – its definition and relationships with local geodetic systems.” Technical Report 3rd Edition, Amendment 1, Geodesy and Geophysics Department, National Imagery and Mapping Agency (NIMA) (2000). https://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html
- [10] DataCite Metadata Working Group. “ DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.3.” DataCite e.V. (2019). <https://doi.org/10.14454/7xq3-zf69>
- [11] Scherer, D., F. Antretter, S. Bender, J. Cortekar, S. Emeis, U. Fehrenbach, G. Gross, G. Halbig, J. Hasse, B. Maronga, S. Raasch and K. Scherber. “Urban climate under change [UC] 2-A national research programme for developing a building-resolving atmospheric model for entire city regions.” Meteorologische Zeitschrift Heft 28, Band 2 (2019): 95-104. <https://doi.org/10.15488/10423>
- [12] Scherer, D., F. Ament, S. Emeis, U. Fehrenbach, B. Leitl, K. Scherber, C. Schneider and U. Vogt “Three-dimensional observation of atmospheric processes in cities.” Meteorologische Zeitschrift, Band 28, Heft 2 (2019): 121-138. <https://doi.org/10.1127/metz/2019/0911>

- [13] Martin, D., C.S. Price, I.R. White, G. Nickless, K.F. Petersson, R.E. Britter, A.G. Robins, S.E. Belcher, J.F. Barlow, M. Neophytou, S.J. Arnold, A.S. Tomlin, R.J. Smalley and D.E. Shallcross. “Urban tracer dispersion experiments during the second DAPPLE field campaign in London 2004.” *Atmospheric Environment*, Band 44, Heft 25 (2010): 3043-3052. <https://doi.org/10.1016/j.atmosenv.2010.05.007>
- [14] Bohnenstengel, S., K. H. Schlünzen, and D. Grawe. “Influence of thermal effects on street canyon circulations.” *Meteorologische Zeitschrift* Band 13, Heft 5 (2004): 381-386. <https://doi.org/10.1127/0941-2948/2004/0013-0381>
- [15] Salim, M. H., K. H. Schlünzen, D. Grawe, M. Boettcher, A. M. U. Gierisch and B. H. Fock. “The microscale obstacle-resolving meteorological model MITRAS v2.0: model theory.” *Geosci. Model Dev.*, Band 11 (2018): 3427–3445. <https://doi.org/10.5194/gmd-11-3427-2018>
- [16] Maronga, B., G. Gross, S. Raasch, S. Banzhaf, R. Forkel, W. Heldens, F. Kanani-Sühring, A. Matzarakis, M. Mauder, D. Pavlik, J. Pfaffenrott, S. Schubert, G. Seckmeyer, H. Sieker and K. Winderlich. “Development of a new urban climate model based on the model PALM-Project overview, planned work, and first achievements.” *Meteorologische Zeitschrift*, Band 28, Heft 9 (2019): 1-15. <https://doi.org/10.1127/metz/2019/0909>