

Datenmanagementplan

SFB1368-B04



Clausthaler Zentrum
für Materialtechnik

13. JULI 2021

Clausthaler Zentrum für Materialtechnik

Prof. Dr. Wolfgang Maus-Friedrichs

Dr. Lienhard Wegewitz

M. Sc. Philipp Moritz

Datenmanagementplan SFB1368 – B04

1. Planung und Erstellung

1.1. Projektbeschreibung und Zusammenfassung

Projektbezeichnung	SFB1368 Teilprojekt B04 - Klebstoffbasierte Montageprozesse in XHV-adäquater Atmosphäre mit desoxidierten und oxidierten Fügepartnern
Fördermittelgeber und Kennzeichen	DFG, SFB1368, Projektnummer: 394563137, Teilprojekt B04
Projektlaufzeit	01.01.2020-31.12.2023
Projektleiter	Prof. Dr. Wolfgang Maus-Friedrichs, Prof. Dr.-Ing. Annika Raatz
Forschungseinrichtungen	Clausthaler Zentrum für Materialtechnik (CZM), TU Clausthal Institut für Montagetechnik (match), Leibniz Universität Hannover

1.2. Vorhandene Daten

Auf vorhandene Daten kann in diesem Projekt nicht zurückgegriffen werden, da es sich bei der sauerstofffreien Produktion um ein neues Forschungsfeld handelt. Lediglich bekannte Geräteeinstellungen der Messgeräte können wiederverwendet werden. Für einige Materialien und Messmethoden können Referenzdaten, die aus vergangenen Forschungsprojekten stammen, zum Vergleich herangezogen werden. Diese Daten sind auf einem NAS-System des Rechenzentrums der TU Clausthal mit Backup außerhalb Clausthals gespeichert und können vom Projektbearbeiter verwendet werden.

1.3. Arten von Daten

Im Laufe des Projekts werden folgende Arten von Datensätzen generiert:

XPS-Analyse

Es handelt sich um Spektroskopie-Daten, die mittels Röntgen-Photoelektronen-Spektroskopie (XPS) gewonnen werden. Anhand dieser Daten lassen sich die chemische Zusammensetzung und die chemischen Bindungen an einer Probenoberfläche bestimmen. Dazu werden Übersichtsspektren über einen weiten Energiebereich aufgenommen, die alle Elemente einer Oberfläche beinhalten und Detailspektren, die eine bessere Auflösung haben, allerdings nur einen kleinen Energiebereich zeigen.

MIES-UPS-Analyse

Es handelt sich um Spektroskopie-Daten, die mit den Methoden Spektroskopie mit metastabilen Edelgasatomen (MIES) und Ultraviolett-Photoelektronen-Spektroskopie (UPS)) gewonnen werden. Anhand der Daten lassen sich Aussagen über die elektronische Struktur einer Probenoberfläche und somit über die chemischen Gruppen in den äußersten Atomlagen treffen.

AFM-Analyse

Diese Datensätze werden mit Rasterkraftmikroskopen (AFM) erzeugt. Die Rohdaten beinhalten dabei eine Positions-, eine Höhen- und eine Phaseninformation. Aus diesen Daten werden topografische Abbildungen und Rauheitskennwerte von Probenoberflächen generiert.

REM/AES-Analyse

Diese Datensätze werden am Gerät Omicron NanoSAM erzeugt. Es handelt sich um eine Kombination aus Rasterelektronenmikroskopie-Aufnahmen (REM) und Auger-Elektronenspektren (AES). Die REM-Aufnahmen sind elektronenoptisch erzeugte Bildformate, die einen Eindruck topografischer Strukturen an Oberflächen vermitteln. Die Auger-Elektronenspektren ermöglichen die lokale qualitative und quantitative Bestimmung der chemischen Oberflächenzusammensetzung mit Informationstiefen in der Größenordnung kleiner 10 nm.

REM/EDX-Analyse

Diese Datensätze werden am Gerät Helios Nanolab 600 (FEI) erzeugt. Es handelt sich um eine Kombination aus Rasterelektronenmikroskopie-Aufnahmen (REM) und energiedispersiver Röntgenspektroskopie (EDX). Die REM-Aufnahmen sind elektronenoptisch erzeugte Bildformate, die einen Eindruck topografischer Strukturen an Oberflächen vermitteln. EDX-Spektren ermöglichen die lokale qualitative und quantitative Bestimmung der chemischen Zusammensetzung mit Auflösungen im Bereich 100 nm bis einige Mikrometer.

Elektronisches Laborbuch (eLabFTW)

Im elektronischen Laborbuch werden Metadaten zu allen Experimenten, die im Rahmen des Projekts stattfinden, gesammelt. Über eine eindeutige ID, die automatisch im elektronischen Laborbuch vergeben wird, können die Datensätze aus Analysen und die im Laborbuch festgehaltenen Metadaten immer eindeutig zugeordnet werden.

Fotos und Videos

Fotos und Videos entstehen bei der Dokumentation von Proben, deren Präparation und dem Ablauf von Prozessen, zum Beispiel einer Plasmabehandlung. Sie werden gemeinsam mit der zum Projekt abgelegten Auswertung gespeichert und sind Proben und Experimenten eindeutig über die eindeutige ID aus dem elektronischen Laborbuch zuordenbar.

Auswertung, Präsentationen und Ergebnisberichte

Dabei handelt es sich um alle Arbeiten über die Rohdaten hinaus, wie die Auswertung und Interpretation, sowie das Erstellen von Präsentationen, Publikationen und Ergebnisberichten.

1.4. Datenformate und Datenerzeugung

XPS-Analyse

Die Rohdaten zu XPS-Analysen liegen als Textdatei mit CSV-Format vor. Sie werden mittels eigener Software auf LabView-Basis erzeugt. Außerdem werden die Daten in das im Bereich XPS gängige VAMAS-Format konvertiert gespeichert. Die Analyse der Rohdaten erfolgt mit dem Programm CasaXPS, das wiederum VAMAS-Dateien verwendet. Publikationsreife Grafiken werden mit OriginPro erstellt und liegen im OPJU-Dateiformat, sowie gängigen Bildformaten wie PNG, TIFF, JPG oder EMF vor.

MIES-UPS-Analyse

Die Rohdaten zu MIES- und UPS-Analysen liegen als Textdatei mit CSV-Format vor. Sie werden mittels eigener Software auf LabView-Basis erzeugt. Außerdem werden die Daten in das VAMAS-Format konvertiert gespeichert. Die Analyse der Rohdaten erfolgt mit dem Programm CasaXPS, das wiederum VAMAS-Dateien verwendet. Publikationsreife Grafiken werden mit OriginPro erstellt und liegen im OPJU-Dateiformat, sowie gängigen Bildformaten wie PNG, TIFF, JPG oder EMF vor.

AFM-Analyse

Die Rohdaten zu AFM-Analysen liegen in proprietären Dateiformaten der Gerätehersteller vor und werden in diesem Zustand gespeichert. Zusätzlich werden bei der Auswertung mit der Software SPIP (Image Metrology) PNG bzw. JPG Dateien erzeugt, die die Oberflächentopografie und -strukturen zeigen. Für die Bestimmung von Höhenunterschieden werden ASCII-Dateien mit den entsprechenden Messpunkten gespeichert.

REM/AES-Analyse

Die Rohdaten zu AES-Analysen liegen als MTRX-Datei in einem proprietären Format vor. Die AES-Spektren werden zusätzlich im offenen VAMAS-Format gespeichert. Die Auswertung der AES-Spektren erfolgt mit CasaXPS und wird ebenfalls im VAMAS-Format abgelegt. REM-Bilder werden im TIF-Format gespeichert.

REM/EDX-Analyse

Die Rohdaten zu EDX-Analysen liegen als IPJ-Datei in einem proprietären Format vor. Die EDX-Spektren werden zusätzlich im CSV-Format gespeichert. Quantitative Analysen werden in XLSX- oder Word-Dateien abgelegt. REM-Bilder werden im TIF-Format gespeichert.

Elektronisches Laborbuch

Als elektronisches Laborbuch wird eLabFTW verwendet. eLabFTW ist eine kostenfreie, aus der Wissenschaftscommunity heraus entwickelte OpenSource-Software. Die Daten können als CSV-, PDF- oder ZIP-Dateien exportiert werden. In eLab werden von Experimentierenden unter anderem folgende Angaben gespeichert:

- Eindeutige Bezeichnung einer Probe

- Durchführende Person und Datum der Durchführung
- Probe und Probenmaterial
- Probentransport
- Präparation der Probe (z.B. Plasmabehandlung, Reinigen, Schneiden, Beschichten, usw.)
- Durchgeführte Analysen mit dazugehörigen Parametern der Messgeräte
- Kommentare, Beobachtungen und Notizen

Fotos und Videos

Fotos und Videos werden in den Formaten JPG bzw. TIF, PNG und MP4 abgelegt. Sie werden gemeinsam mit der zum Projekt abgelegten Auswertung gespeichert und sind Proben und Experimenten eindeutig über die eindeutige ID aus dem elektronischen Laborbuch zuordenbar.

Auswertung, Präsentationen und Ergebnisberichte

Dabei handelt es sich um alle Arbeiten über die Rohdaten hinaus, wie die Auswertung und Interpretation, das Erstellen von Präsentationen und Publikationen, sowie Ergebnisberichten. Sie werden in den Formaten VAMAS, OPJU und PDF-Export, XLSX, DOCX und PPTX gespeichert. Alle Datensätze dieser Kategorie werden in einem Projektordner in der Cloud des Rechenzentrums der Technischen Universität Clausthal abgelegt.

1.5. Datenvolumen

Für das Datenvolumen, das im Laufe des Projekts entsteht, kann nur eine Schätzung angegeben werden. Entscheidungen, an welchen Proben wie viele Analysen durchgeführt werden müssen, können erst während der Projektlaufzeit gegeben werden.

Tabelle 1: Erzeugtes Datenvolumen im Projekt.

Datensatz	Anzahl	Rohdaten / MB	Auswertung / MB	Gesamt / GB
XPS-Analyse	200	2	3	1,0
MIES/UPS-Analyse	100	2	4	0,6
AFM-Analyse	200	60	30	18,0
REM/AES-Analyse	25	50	30	2,0
REM/EDX-Analyse	50	250	300	27,5
el. Laborbuch	350	2		0,7
Fotos u. Videos	300	5		1,5
Ergebnisse / Berichte	Schätzung		8,0	8,0
				59,3

1.6. Qualitätssicherung

Für alle in Analysen oder als Metadaten im elektronischen Laborbuch erzeugten Daten gelten klare Regeln für deren Benennung und Speicherung. Die Datenhaltung wird im Fall der Veröffentlichung von den Autoren bis zu den Rohdaten zurückverfolgt, um die korrekte wissenschaftliche Praxis sicherzustellen.

2. Auswahl und Aufbewahrungsdauer

2.1. Gründe für die Aufbewahrung

Da noch keine gesetzliche Grundlage für die Aufbewahrung von Forschungsdaten existiert, ist die Entscheidung für die Aufbewahrung eine Selbstverpflichtung. Mit der Aufbewahrung wird die Nachnutzung in Folgeprojekten oder durch andere Wissenschaftler möglich. Außerdem ist die Aufbewahrung und das zugänglich machen aller Daten, auf denen wissenschaftliche Publikationen beruhen, eine der Säulen guter wissenschaftlicher Praxis.

2.2. Datenauswahl

Die Verwendbarkeit der generierten Daten kann erst nach der Auswertung abgeschätzt werden. Aufgrund des maßvollen Datenvolumens werden alle generierten Daten gespeichert. Die Verantwortung für die Auswahl der für die Publikation geeigneten Datensätze liegt beim Projektmitarbeiter und dem Projektleiter.

2.3. Aufbewahrungsdauer und Datenentsorgung

Die Daten werden für mindestens 10 Jahre auf den Servern des Rechenzentrums der Technischen Universität Clausthal gespeichert, danach wird überprüft, ob eine weitere Speicherung der Daten sinnvoll und notwendig ist. Datensätze zu Publikationen werden im Repositorium Zenodo abgelegt, sind mit einer DOI versehen und werden auf unbestimmte Zeit gespeichert (siehe 4.1).

3. Einspeisen und Datenübergabe

Die Datenerzeugung (Einspeisen), Datenspeicherung, Datenweitergabe und die Erzeugung von Metadaten sind in den Abschnitten 1.4, 4.1, 6 und 10 beschrieben.

4. Speicherung und Infrastruktur

4.1. Datensicherung

Die Rohdaten werden während des Projekts auf dem NAS-System des Rechenzentrums der TU Clausthal gespeichert. Es existieren festgelegte Dateipfade und Benennungen für die Rohdaten, sodass diese immer eindeutig einer Analyse zugeordnet werden können. Zudem sind sie mit einer eindeutigen ID mit den Metadaten aus dem elektronischen Laborbuch verknüpft. Auswertung und Ergebnisse, die auf den Rohdaten basieren, werden während der Projektlaufzeit in Projektordnern mit zugewiesenem Zugriff in der Cloud der TU Clausthal gespeichert (TUCloud/ownCloud). Die Daten sind durch Backups außerhalb Clausthals abgesichert. Rohdaten, Auswertung und Ergebnisse, die für die Projektpartner in Hannover

nutzbar sind, werden in einem gemeinsam zugänglichen Bereich des seafire-Systems der LU Hannover abgelegt. Alle relevanten Daten zu Publikationen werden in dem von der EU und dem CERN betriebenen Repositorium Zenodo (<https://zenodo.org/>) abgelegt und für die eindeutige Zuordnung mit einer DOI versehen. Diese DOI wird in der Publikation nach Möglichkeit mit angegeben. Die DOI wird im Zusammenhang mit der Publikation im CKAN-System des SFB als Datensatz gespeichert.

4.2. Infrastruktur

Das Rechenzentrum der Technischen Universität Clausthal stellt für die Rohdatenablage und die Archivierung Speicherplatz auf dem NAS-System der TU Clausthal zur Verfügung. Für die Daten der Auswertung, Berichte, usw. steht die vom Rechenzentrum der TU Clausthal gehostete TUCloud zur Verfügung. Eine TU-eigene Instanz des elektronischen Laborbuchs eLabFTW wird vom Rechenzentrum gehostet. Die Daten sind durch Backups außerhalb Clausthals abgesichert.

Die Infrastruktur für publizierte Datensätze wird von der EU und dem CERN in Form des Repositoriums Zenodo zur Verfügung gestellt (<https://zenodo.org/>).

Die Infrastruktur für den Datenaustausch mit den Kooperationspartnern in Hannover (Seafire, CKAN) wird von den Leibniz Universität IT Services bereitgestellt. Dienstinhaber des CKAN-Systems ist das INF-Projekt des Sonderforschungsbereichs.

5. Erhaltungsmaßnahmen

Die eingesetzte Software und die verwendeten Dateiformate sind unter 1.4 beschrieben. Alle Abhängigkeiten zwischen Geräten, Rohdaten und anderen Formaten werden über die eindeutige ID im elektronischen Laborbuch abgebildet und sind auf diese Weise verfolgbar.

Die Daten werden für mindestens 10 Jahre auf den Servern des Rechenzentrums der Technischen Universität Clausthal gespeichert, danach wird die Aufbewahrungswürdigkeit überprüft, ob eine weitere Speicherung der Daten sinnvoll und notwendig ist. Im Repositorium Zenodo abgelegte Daten zu Publikationen sind mit einer DOI versehen und werden auf unbestimmte Zeit gespeichert.

6. Zugriff und Nutzung

6.1. Nachnutzung und Suchbarkeit

Die Daten zu allen Publikationen werden vollständig (Rohdaten, Auswertung, Ergebnisse, SI) im Repositorium Zenodo gespeichert und mit einer DOI versehen und nach Veröffentlichung des Artikels allen Externen zugänglich gemacht. Daten aus Experimenten, die nicht in einer Veröffentlichung verwendet wurden, können SFB-intern verwendet werden.

6.2. Zugangsbeschränkungen

Für die Daten, die Grundlage von Publikationen sind, bestehen keine Zugangsbeschränkungen. Sobald eine Publikation veröffentlicht ist, werden die Datensätze mit OpenAccess auf Zenodo freigegeben. Alle anderen Datensätze sind auf den Servern der TU Clausthal gespeichert und können innerhalb des SFB zur Verfügung gestellt werden.

6.3. Interoperabilität

Soweit möglich werden die Daten in nicht-proprietären Dateiformaten gespeichert und zur Verfügung gestellt. Für einige Datensätze ist dies aufgrund proprietärer Software der Gerätehersteller nicht möglich (siehe 1.4). Eine weitere Ausnahme sind Beispiele von Microsoft Office-Formaten (vgl. 1.4), welche zwar proprietäre Datenformate repräsentieren, jedoch im vollen genutzten Funktionsumfang mit OpenSource-Software geöffnet und bearbeitet werden können.

7. Organisation, Management und Richtlinien

7.1. Organisation und Management

Die Daten werden innerhalb der Institution gespeichert und Datensätze zu Publikationen werden im Repositorium Zenodo mit OpenAccess-Zugang abgelegt. Eine genaue Beschreibung findet sich in Abschnitt 4.1. Am Forschungsdatenmanagement sind das Rechenzentrum der Technischen Universität Clausthal, die Projektbearbeitenden, die Projektleitung, der FDM-Beauftragte Dr. Lienhard Wegewitz und das INF-Projekt des SFB beteiligt. Das FDM-System und der FDM-Prozess wurden für verschiedene Datensätze in diesem Datenmanagementplan beschrieben.

7.2. Richtlinien

Es gelten die Leitlinien zum Forschungsdatenmanagement der DFG, die Vorgaben der DFG zum Forschungsdatenmanagement in Sonderforschungsbereichen, die Forschungsdaten-Leitlinie der Technischen Universität Clausthal (<https://www.fdm.tu-clausthal.de/erste-schritte/forschungsdaten-leitlinie>) und die Richtlinie zum Umgang mit Forschungsdaten im SFB 1368.

7.3. Einhaltung der Vorgaben und Planung

Dieser Datenmanagementplan wird bei Bedarf durch den Projektbearbeiter und den FDM-Beauftragten Dr. Lienhard Wegewitz aktualisiert. Die Einhaltung des FDM-Systems wird wie unter 8.2 beschrieben nachgewiesen.

8. Kosten

8.1. Kosten- und Aufwandsschätzung

Die Infrastruktur wird vom Rechenzentrum der Technischen Universität Clausthal kostenfrei zur Verfügung gestellt. Die Nutzung des Repositoriums Zenodo ist kostenfrei, es wird von der EU und dem CERN finanziert.

Der zusätzliche Personalaufwand für die Umsetzung des Forschungsdatenmanagements in diesem Projekt ist gering. Bei der Entwicklung der in diesem Datenmanagementplan beschriebenen Vorgehensweise wurde darauf geachtet, dass möglichst viele Aufgaben, zum Beispiel das Eintragen im elektronischen Laborbuch, den Anforderungen des Forschungsdatenmanagements genügen, aber kaum oder keinen Zusatzaufwand verursachen. Insgesamt hat sich gezeigt, dass das FDM-System der Arbeitsgruppe keine Mehrarbeit verursacht, sondern im Gegenteil den Gesamtaufwand, da Daten schneller gefunden und zugeordnet werden können.

8.2. Anreize

Die in diesem Datenmanagementplan beschriebenen Regeln und Vorgehensweisen sind in der Arbeitsgruppe des Antragstellers verpflichtend durchzuführen. Dies wird sichergestellt, indem Experimente nur dann für Abschlussarbeiten und Publikationen genutzt werden dürfen, wenn von der Sicherung der Rohdaten bis zur Ablage der Auswertung und der Ergebnisse, sowie dem Eintrag im elektronischen Laborbuch alle Informationen vorhanden sind.

Bei der Einführung des in diesem Datenmanagementplan beschriebenen Vorgehens und der dazugehörigen Software wurden alle Mitglieder der Arbeitsgruppe miteinbezogen. Ein Kernteam hat Vorschläge unterbreitet, die dann gemeinsam bewertet und diskutiert wurden. Bei der Entwicklung der Vorgehensweise zum Forschungsdatenmanagement in der Arbeitsgruppe war ein Hauptaugenmerk möglichst viele Erleichterungen zu schaffen und nur wenig zusätzliche Verwaltungsarbeit zu schaffen. Zu einem Stichtag wurde die Entscheidung getroffen das System in seiner derzeitigen Funktionsweise zu nutzen, diese Entscheidung gilt für alle Arbeitsgruppenmitglieder.

9. Rechtliche Aspekte von Forschungsdaten

9.1. Datenschutz - Personenbezogene Daten

Bis auf die Namen der Experimentierenden, die Namen von Autoren einer Publikation und die Institutszugehörigkeit werden keinerlei personenbezogene Daten erhoben und gespeichert.

9.2. Urheberrecht

Es wird keine Fremdsoftware als Produkt des Forschungsprozesses verwendet. Für die Auswertung eingesetzte Fremdsoftware wird in den Publikationen benannt. Es werden im Projekt keine fremden Forschungsdaten außerhalb der Kooperation im SFB verwendet. Bisher sind keine urheberrechtlichen Einschränkungen bekannt, welche die in diesem Projekt generierte Daten betreffen. Möglicherweise unterliegen im Einzelfall Grafiken und Tabellen, die in wissenschaftlichen Fachzeitschriften veröffentlicht sind, Eigentumsrechten des publizierenden Verlags. Dies wird im Einzelfall betrachtet.

10. Metadaten

Metadaten zu Datensätzen aus den Analysen werden an zwei Stellen erfasst. Erstens beinhalten die Rohdaten der meisten Analysemethoden gerätespezifische Parameter, die für die Auswertung notwendig sind und in maschinenlesbaren Formaten in den Rohdatendateien abgelegt sind. Zweitens werden Metadaten zu Proben und Experimenten im elektronischen Laborbuch angegeben. Dabei handelt es sich unter anderem um Probenbezeichnung, Datum des Experiments und eine eindeutige ID. Außerdem existiert für das elektronische Laborbuch ein Schlagwortsystem, für jede Probe bzw. jedes Experiment werden die Messmethoden und projektabhängig weitere Schlagwörter vergeben, die eine Sortierung und Kategorisierung der Proben und Experimente gestatten.

11. Identifikatoren und Informationsobjekte

Die wichtigsten Informationsobjekte in der Arbeitsgruppe sind Experimente. Eine Übersicht über die Informationsobjekte im FDM-System der Arbeitsgruppe von Prof. Dr. Wolfgang Maus-Friedrichs ist in Abbildung 1 gezeigt. Jedes Experiment wird im elektronischen Laborbuch dokumentiert und automatisch mit einer eindeutigeneinzigartigen ID (Identifikator) versehen, alle weiteren Informationsobjekte sind von dort aus verlinkt, zugänglich und über die Unique ID eindeutig zuordenbar.

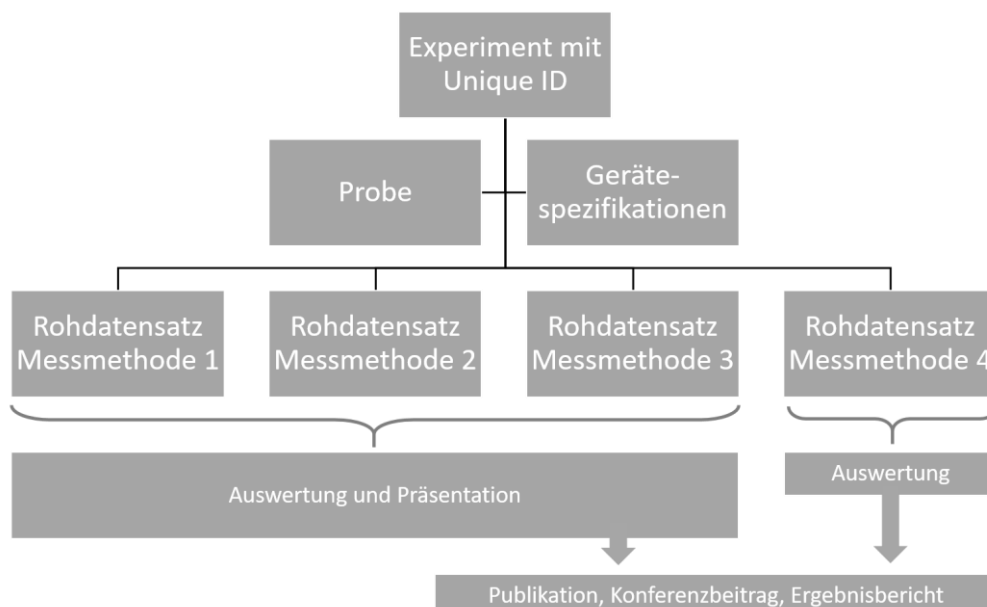


Abbildung 1: Informationsobjekte im FDM-System der AG Maus-Friedrichs.