



Copyright: iStock.com/pidjoe

Zertifikatskurs Data Librarian

ZBIW.

Zentrum für Bibliotheks-
und Informationswissen-
schaftliche Weiterbildung

Technology
Arts Sciences
TH Köln

Zertifikatskurs Data Librarian

Der Alltag in Wissenschaftlichen Bibliotheken verändert sich rasant. Open Science, FAIR*-Prinzipien, Transparenz und Reproduzierbarkeit des Forschungsprozesses stellen Bibliotheken vor neue Herausforderungen. Die fortschreitende Digitalisierung der Wissenschaft wie auch die Entwicklung neuer Hochdurchsatz-Analyseverfahren führen zu einer rasant wachsenden Datenmenge im Forschungsbetrieb. Dies eröffnet viele neue Chancen, bringt aber auch zunehmend höhere Ansprüche mit sich. An dieser Stelle können Beschäftigte in Wissenschaftlichen Bibliotheken ihre traditionelle Rolle als Bewahrer*innen von Forschungsergebnissen neu definieren und ihr Angebotsportfolio von Literatur auf Forschungsdaten ausweiten. Dazu benötigen sie einen Kanon an Kompetenzen, die größtenteils noch nicht Inhalte der klassischen bibliothekarischen Studiengänge sind.

Der Zertifikatskurs „Data Librarian“ qualifiziert Beschäftigte aus Wissenschaftlichen Bibliotheken für neue Aufgaben in diesem Kontext und macht sie „datenkompetent“. Vermittelt werden grundlegende Kenntnisse zu Forschungsdaten, Digitalisierung, Metadaten, Suchmaschinen, Datenbanksystemen, maschinellem Lernen sowie künstlicher Intelligenz, auch unter Berücksichtigung rechtlicher und ethischer Aspekte.

Der Kurs findet im Format Blended Learning statt, einer Kombination aus einer online-gestützten Selbstlernphase über eine Plattform und Präsenz (vor Ort oder Live-Online). Er beinhaltet zahlreiche praktische Übungen wie Live-Coding und kleine anwendungsbezogene Projekte.

**Findable, Accessible, Interoperable, Re-usable*

Zielgruppe

Dieser Zertifikatskurs richtet sich an Beschäftigte Wissenschaftlicher Bibliotheken mit einer abgeschlossenen Berufsausbildung oder einem Studienabschluss und mindestens einjähriger Berufserfahrung, die Chancen und Möglichkeiten der Digitalisierung in Bibliotheken nutzen oder umsetzen wollen und in der Lage sein möchten, den Herausforderungen entgegenzusteuern.

Termine

Modul 1: 04.11.2024 - 06.11.2024 (3 Tage Präsenz)

Modul 2: 11.12.2024 (eine Stunde Live-Online) | 24.01.2025 (ein Tag Präsenz)

Modul 3: 27.01.2025 (eine Stunde Live-Online) | 07.03.2025 (6 Stunden Live-Online)

Modul 4: 10.03.2025 (eine Stunde Live-Online) | 11.04.2025 (ein Tag Präsenz)

Modul 5: 28.04.2025 (1,5 Stunden Live-Online) | 06.06.2025 (ein Tag Präsenz)

Modul 6: 11.06.2025 (2 Stunden Live-Online) | 29.08.2025 (7 Stunden Live-Online)

Inhaltlicher Aufbau

Die Inhalte werden in sechs Modulen vermittelt:

Modul 1: Hacken und experimentieren mit Daten

Modul 2: Daten strukturieren - beschreiben - wiederauffinden

Modul 3: Daten analysieren und darstellen

Modul 4: Forschungs(daten)prozess verstehen und unterstützen

Modul 5: Nachhaltig und verantwortungsvoll handeln im Umgang mit Daten

Modul 6: Projektmodul

Arbeitsaufwand

Insgesamt 8 Präsenztage zuzüglich Einführungsmeetings, Online-Fragestunden und E-Learning-Phasen innerhalb der Module (jeweils etwa 30-40 Zeitstunden Arbeitsaufwand); Absolvierung eines Praxisprojektes; Arbeitsaufwand insgesamt 240 Stunden; fakultativ: Erstellung einer Projektarbeit zur Erlangung eines Zertifikates mit 8 ECTS.

Abschluss

Teilnahmebescheinigung oder Zertifikat mit 8 ECTS

Für die Anwesenheit bei allen Präsenzterminen wird eine Teilnahmebescheinigung ausgestellt. Der Abschluss mit Zertifikat und Erlangung von 8 ECTS kann im Rahmen von Modul 6 erreicht werden. Dazu wählen die Teilnehmer*innen ein Thema ihrer Wahl aus den Kursinhalten und erstellen eine Projektarbeit, die bewertet wird.

Die wissenschaftliche Leitung liegt bei Prof. Dr. Konrad Förstner, ZB MED – Informationszentrum Lebenswissenschaften und TH Köln, Institut für Informationswissenschaft.

Kursgröße

Die maximale Kursgröße beträgt 14 Personen.

Tools/Technik

Sie benötigen ein internetfähiges Endgerät (PC oder Laptop), um auf die Lernplattform Moodle zugreifen zu können. Dort finden Sie Materialien, Aufgaben und die Zugangslinks zu den Videomeetings, die in der Regel über Zoom stattfinden. Für Zoom benötigen Sie zudem eine Webcam sowie ein Headset oder ein anderes Audiosystem.

Das Vorhalten eines eigenen vorinstallierten Laptops ist notwendig. Die Arbeitsumgebung wird in den E-Learning-Phasen zu Hause oder am eigenen Arbeitsplatz benötigt. Darüber hinaus sollte die Möglichkeit bestehen, weitere Programme selbstständig installieren zu können.

Veranstaltungsort

ortsunabhängig (Online) und TH Köln, Campus Südstadt, Ubierring 48, 50678 Köln, Raum 211

Kursgebühr

2.000,- Euro

Modul 1 sowie die Abschlusstage der Module 2, 4 und 5 finden in Präsenz vor Ort in Köln statt. Der Preis beinhaltet zwei Zwischenübernachtungen und Verpflegung bei Modul 1 sowie die Verpflegung zu den Abschlusstagen der Module 2, 4 und 5. Weitere Vorabübernachtungen und anfallende Reisekosten zu den Präsenzterminen in Köln sind von den Teilnehmer*innen selbst zu buchen und zu tragen.

Der Kurs ist nur als Gesamtpaket buchbar.

Kontakt

ZBIW der TH Köln

Marvin Lanczek

Tel.: 0221/8275-3695

E-Mail: marvin.lanczek@th-koeln.de

Die Module

Der Zertifikatskurs „Data Librarian“ besteht aus sechs Modulen, die inhaltlich eng miteinander verzahnt sind. Auf den folgenden Seiten finden Sie eine detaillierte Beschreibung aller Module.

Modul 1:	Hacken und experimentieren mit Daten	5
Modul 2:	Daten strukturieren - beschreiben - wiederauffinden.....	6
Modul 3:	Daten analysieren und darstellen.....	7
Modul 4:	Forschungs(daten)prozess verstehen und unterstützen	8
Modul 5:	Nachhaltig und verantwortungsvoll handeln im Umgang mit Daten	10
Modul 6:	Projektmodul	12

Modul 1: Hacken und experimentieren mit Daten	
Inhalt	<ul style="list-style-type: none"> » Unix Shell » Python und wichtige Programmierbibliotheken » Funktionen und objektorientierte Programmierung » Webscraping » Datenaufbereitung und Visualisierung » git und GitHub » Literate Programming mit Jupyter Notebooks » Plain-Text Dateien » Text-Editoren » Nutzung von APIs » Automatisierung » Computational Thinking
Lernziele	<p>Am Ende des Moduls können Sie</p> <ul style="list-style-type: none"> » in der Unix Shell navigieren, Programme ausführen und Datenprozessierung automatisieren, » kleine Python-Skripte schreiben, » Projekte mit git versionieren und diese auf GitHub teilen sowie zusammen mit anderen bearbeiten, » verschiedene Plain-Text-basierte Dateiformate bearbeiten, » Web-APIs und Programmier-Bibliotheken zum automatischen Beziehen von Daten auswählen, » mögliche arbeitsintensive Schritte in Ihrer Institution erkennen und durch Softwarelösungen vereinfachen.
Methoden	Live-Coding, Einzel- und Gruppenarbeit, Diskussion, Fallbeispiele, Erfahrungsaustausch
Dozent*innen	<p>Prof. Dr. Konrad Förstner ZB MED – Informationszentrum Lebenswissenschaften und TH Köln, Institut für Informationswissenschaft</p> <p>Rabea Müller ZB MED – Informationszentrum Lebenswissenschaften</p>
Arbeitsaufwand	7 Std. E-Learning, ein Live-Online-Technik-Check á eine Std., eine Live-Online-Fragestunde á eine Std. und 3 Präsenztage á 7 Std. = 30 Std. für Modul 1
Ort	ortsunabhängig (Online) und TH Köln, Campus Südstadt, Ubierring 48, 50678 Köln, Raum 211
Termin	<p>04.11.2024 - 15.12.2024 (Modullaufzeit)</p> <p>Technik-Check 29.10.2024 13.00 – 14.00 Uhr</p> <p>Präsenztage vor Ort 04.11.2024 10.00 - 17.00 Uhr 05.11.2024 9.00 - 16.00 Uhr 06.11.2024 9.00 - 16.00 Uhr</p> <p>Live-Online-Meeting (Anwesenheit erforderlich) Fragestunde 09.12.2024 13.00 – 14.00 Uhr</p>

Modul 2: Daten strukturieren - beschreiben - wiederauffinden	
Inhalt	<ul style="list-style-type: none"> » Strukturierte Daten und Metadaten <ul style="list-style-type: none"> - Datenexport aus öffentlichen Web-Quellen - Datenformate - Transformation von Datenformaten (z.B. über Python-Skripte) » Suchmaschinentechnologie, Information Retrieval <ul style="list-style-type: none"> - Einführung in das Retrieval mit PyTerrier - Datenimport und Indexierung - Suchmöglichkeiten in PyTerrier
Lernziele	<p>Am Ende des Moduls können Sie</p> <ul style="list-style-type: none"> » offene Web-Quellen herunterladen und die Daten skriptgesteuert in (bibliothekarische) Standardformate überführen, » eine Standard-Suchmaschine installieren, konfigurieren und eigene Daten indexieren, » beliebige Anfragen an das System stellen sowie die Ergebnisse einordnen und evaluieren, » Python-Skripte und PyTerrier in Ihrem Arbeitsalltag einsetzen.
Methoden	Flipped Classroom - Materialien für das Selbststudium und Bearbeitung von Aufgaben in der E-Learning-Phase, Live-Coding, Einzel- und Gruppenarbeit, Diskussion, Erfahrungsaustausch
Dozent*innen	<p>Björn Engelmann TH Köln, Institut für Informationswissenschaft</p> <p>Fabian Haak TH Köln, Institut für Informationswissenschaft</p>
Arbeitsaufwand	28 Std. E-Learning, ein Live-Online-Modulauftakt á eine Std., eine Live-Online-Fragestunde á eine Std. und ein Präsenztage á 6 Std. = 36 Std. für Modul 2
Ort	ortsunabhängig (Online) und TH Köln, Campus Südstadt, Ubierring 48, 50678 Köln, Raum 211
Termin	<p>09.12.2024 - 24.01.2025 (Modullaufzeit)</p> <p>Live-Online-Meetings (Anwesenheit erforderlich) Startmeeting 11.12.2024 16.00 - 17.00 Uhr Fragestunde 08.01.2025 16.00 - 17.00 Uhr</p> <p>Präsenztage vor Ort 24.01.2025 10.00 - 16.00 Uhr</p>

Modul 3: Daten analysieren und darstellen	
Inhalt	Praktische Einführung in Methoden und Begriffe: <ul style="list-style-type: none"> » der deskriptiven Statistik (Mittelwerte, Varianz, Korrelation, Häufigkeiten) » der schließenden Statistik (Verteilungen, Stichproben, Variablen, Mittelwerttests) » des maschinellen Lernens (überwachtes Lernen, unüberwachtes Lernen, Reinforcement, Mustererkennung, Clustering, Klassifikation, Dimensionsreduktion) » der Datenvisualisierung (Tabellen, Diagrammtypen)
Lernziele	Am Ende des Moduls können Sie <ul style="list-style-type: none"> » Datensätze beschreiben und charakterisieren, » Tabellendaten in Python einlesen, » grundlegende deskriptive Statistiken in Python berechnen, » einfache statistische Diagramme in Python erstellen, » Verfahren des maschinellen Lernens mittels Python anwenden, » die Python Bibliotheken pandas, seaborn und scikit-learn nutzen, » die grundlegende Funktionsweise von Methoden des maschinellen Lernens nachvollziehen.
Methoden	Flipped Classroom - Materialien für das Selbststudium und Bearbeitung von Aufgaben in der E-Learning-Phase, Austausch, Trainer-Feedback und kollegiale Beratung beim virtuellen Präsenztermin
Dozent*innen	Prof. Dr. Konrad Förstner ZB MED – Informationszentrum Lebenswissenschaften und TH Köln, Institut für Informationswissenschaft Dr. Ania López Heinrich-Heine-Universität Düsseldorf
Arbeitsaufwand	38 Std. E-Learning, ein Live-Online-Modulauftakt á eine Std., eine Live-Online-Fragestunde á eine Std. und ein Live-Online-Präsenztag á 6 Std. = 46 Std. für Modul 3
Ort	ortsunabhängig (Online)
Termin	27.01.2025 - 07.03.2025 (Modullaufzeit) Live-Online-Meetings (Anwesenheit erforderlich) Startmeeting 27.01.2025 17.00 - 18.00 Uhr Fragestunde 13.02.2025 16.00 - 17.00 Uhr Präsenztag 07.03.2025 10.00 - 16.00 Uhr

Modul 4: Forschungs(daten)prozess verstehen und unterstützen

Inhalt	<p>Wie funktioniert Forschung und was hat das mit Bibliotheken zu tun?</p> <ul style="list-style-type: none">» Einführung in den wissenschaftlichen Publikationsprozess: Stationen des Publikationsprozesses vom Schreiben über die Einreichung und Begutachtung bis zur Publikation» Werkzeuge für Forscher*innen: Publikationsplattformen und Forschungsdatenrepositorien, Auswahlkriterien der Zeitschriften, Review, Predatory Publishing» Was bedeuten Open Access und Co? Welchen Einfluss hat das auf die Reputation der Forscher*innen?» Welche Handreichungen gibt es? Wo finde ich was, um Forscher*innen zu beraten?» Vorteile paralleler Publikationen (Text/FD) <p>Herausforderungen im Forschungsdatenmanagement (FDM):</p> <ul style="list-style-type: none">» Einführung in den Forschungszyklus und Definition von Forschungsdaten» Austauschbarkeit und Speicherung von Forschungsdaten, FAIR Data» Planung des Forschungsprozesses» Rechtliche Fragestellungen» Rolle der Bibliotheken im Management von Forschungsdaten <p>Bibliometrie:</p> <ul style="list-style-type: none">» Vorstellung der Historie und der ursprünglichen Intention» Verwendung von bibliometrischen Indikatoren sowie Bedeutung für die wissenschaftliche Reputationsbildung» Berechnungsweise gängiger bibliometrischer Indikatoren (z.B. Journal Impact Factor, h-Index etc.) sowie kritische Diskussion» Umgang mit Zitationsdaten und -Datenbanken (z.B. Web of Science, Scopus)» Altmetrics als Ergänzung zu zitationsbasierten Indikatoren» Diskussion eines möglichen Spektrums von bibliothekarischen Dienstleistungen im Hinblick auf Bibliometrie
Lernziele	<p>Am Ende des Moduls können Sie</p> <ul style="list-style-type: none">» Open Access definieren und die verschiedenen Wege (Gold und Grün) erklären,» die zurzeit wichtigsten Hilfsmittel (Internetseiten) zur Beratung von Forscher*innen auflisten (Wo finde ich was?),» den wissenschaftlichen Publikationsprozess erläutern und verstehen,» die Diskussion um Predatory Publishing nachverfolgen und eine kritische Einschätzung zum Diskurs geben,» die wichtigsten Herausforderungen im Management von Forschungsdaten benennen und diese in den Forschungszyklus einordnen,» das Management von Forschungsdaten in den wissenschaftlichen Publikationsprozess einbetten,» wichtige Hilfsmittel zur Beratung von Forscher*innen zur Planung des Forschungsdatenmanagements auflisten,» die Begriffe FAIR und OPEN Data definieren,» mit Zitationsdatenbanken umgehen und kennen die wichtigsten bibliometrischen Indikatoren sowie die damit verbundenen Kritikpunkte.

Methoden	Flipped Classroom - Materialien für das Selbststudium und Bearbeitung von Aufgaben in der E-Learning-Phase, Diskussion während des virtuellen Präsenztages
Dozent*innen	<p>Prof. Dr. Ursula Arning ZB MED – Informationszentrum Lebenswissenschaften und TH Köln, Institut für Informationswissenschaft</p> <p>Dr. Ania López Heinrich-Heine-Universität Düsseldorf</p> <p>Dr. Jasmin Schmitz ZB MED – Informationszentrum Lebenswissenschaften</p>
Arbeitsaufwand	30 Std. E-Learning, ein Live-Online-Modulaufakt á eine Std. und ein Präsenztage á 6 Std. = 37 Std. für Modul 4
Ort	ortsunabhängig (Online) und TH Köln, Campus Südstadt, Ubierring 48, 50678 Köln, Raum 211
Termin	<p>10.03.2025 - 11.04.2025 (Modullaufzeit)</p> <p>Live-Online-Meeting (Anwesenheit erforderlich) Startmeeting 10.03.2025 16.00 - 17.00 Uhr</p> <p>Präsenztage vor Ort 11.04.2025 10.00 - 16.00 Uhr</p>

Modul 5:	Nachhaltig und verantwortungsvoll handeln im Umgang mit Daten
Inhalt	Lizenzen (Open Access, Open Data, Open Source) <ul style="list-style-type: none"> » Definition und Anwendungsbereiche offener Lizenzen » Entscheidungskriterien für die Wahl der „richtigen“ Lizenz Ethische Aspekte im Umgang mit Daten <ul style="list-style-type: none"> » Vermittlung von Grundlagenwissen zu Ethik allgemein und im speziellen zur Informationsethik » Ethische Fragestellungen im Umgang mit (Forschungs-)Daten: Nachhaltigkeit, Datenschutz, Zugang Digitale Langzeitarchivierung (DLZA) <ul style="list-style-type: none"> » Vermittlung von Grundlagenwissen: was bedeutet DLZA, was nicht (konzeptionelle Grundlagen, Ebenen der DLZA) » Bewusstsein schaffen für die verschiedenen Herausforderungen, denen man in der DLZA gegenübersteht » Anhand von Tools grundlegende DLZA-Prozesse im Umgang mit Dateiformaten (Formatidentifizierung und -validierung) kennenlernen
Lernziele	Am Ende des Moduls <ul style="list-style-type: none"> » kennen Sie die Unterschiede zwischen den verschiedenen Lizenztypen, die verschiedenen Anwendungsbereiche und wissen, welche Lizenz sich wofür eignet, » haben Sie eine ethische Dilemma-Diskussion im informationswissenschaftlichen Kontext geführt, » können Sie den Umgang mit (Forschungs-)Daten anhand ethischer Fragestellungen reflektieren und verändern bzw. steuern, » können Sie die konzeptionellen Grundlagen und wichtigsten Standards der Digitalen Langzeitarchivierung benennen, » können Sie die verschiedenen Herausforderungen in der Digitalen Langzeitarchivierung erläutern, » können Sie die Vor- und Nachteile unterschiedlicher Erhaltungsstrategien darlegen, » sind Sie in der Lage, die Relevanz des Themas „Dateiformate“ für die Digitale Langzeitarchivierung einzuschätzen sowie Tools zur Dateiformaterkennung und -validierung zu nutzen.
Methoden	Flipped Classroom - Materialien für das Selbststudium und Bearbeitung von Aufgaben in der E-Learning-Phase, Austausch und kollegiale Beratung beim virtuellen Präsenztermin, Feedback
Dozent*innen	Alexandra Jobmann Bibliothek der HafenCity Universität Hamburg Pia Rudnik TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften
Arbeitsaufwand	30 Std. E-Learning, ein Live-Online-Modulauftakt á 1,5 Std. und ein Präsenztag á 5,5 Std. = 37 Std. für Modul 5
Ort	ortsunabhängig (Online) und TH Köln, Campus Südstadt, Ubierring 48, 50678 Köln, Raum 211

Termin	28.04.2025 - 06.06.2025 (Modullaufzeit) Live-Online-Meeting (Anwesenheit erforderlich) Startmeeting 28.04.2025 16.00 - 17.30 Uhr Präsenztag vor Ort 06.06.2025 10.00 - 15.30 Uhr
---------------	--

Modul 6: Projektmodul	
Inhalt	Sie arbeiten an einem praktischen Projekt, in dem die Inhalte eines oder mehrerer Module genutzt werden, um ein Problem des eigenen Arbeitsumfeldes zu lösen. Dies kann z.B. ein Konzept zu einem Forschungsdaten-Management-Plan, ein Programm zur Prozessierung von Daten oder Ähnliches sein. Dabei erhalten Sie während der Bearbeitung Feedback von Ihren Peers.
Lernziele	Am Ende des Moduls » haben Sie ein Projekt genutzt um einzelne Inhalte dieses Zertifikatskurses praktisch anzuwenden und zu transferieren, » sind Sie in der Lage, Problemstellungen in Ihrem beruflichen Alltag zu analysieren und zu lösen.
Methoden	Projektarbeit, Peer-Feedback
Dozent*innen	Prof. Dr. Konrad Förstner ZB MED – Informationszentrum Lebenswissenschaften und TH Köln, Institut für Informationswissenschaft
Arbeitsaufwand	45 Std. Projektarbeit, ein Live-Online-Modulauftakt á 2 Std. und ein Live-Online-Präsentationstag á 7 Std. = 54 Std. für Modul 6
Ort	ortsunabhängig (Online)
Termin	11.06.2025 - 29.08.2025 (Modullaufzeit) Live-Online-Meetings (Anwesenheit erforderlich) Startmeeting 11.06.2025 16.00 - 18.00 Uhr Projektpräsentation 29.08.2025 10.00 - 17.00 Uhr

Charakteristika und Vorteile von Zertifikatskursen

Sich berufsbegleitend qualifizieren, neue fachliche Kompetenzen sowie Schlüsselqualifikationen erwerben, anregenden beruflichen Austausch erleben – es gibt viele Gründe, die für die Teilnahme an einem Zertifikatskurs sprechen.

Zertifikatskurse bieten ganz gezielt wissenschaftlich fundierte, praxisrelevante und aktuelle Weiterbildung in kompakter Form auf Hochschulniveau, die präzise auf die Bedürfnisse der Praxis zugeschnitten sind. Zertifikatskurse sind in sich sowohl thematisch als auch organisatorisch geschlossen und bestehen aus mehreren aufeinander aufbauenden Modulen, wobei die Inhalte im Berufsalltag sofort Anwendung finden können und bereits während der Weiterbildung einen Mehrwert für den Alltag liefern. Als berufsbegleitende Weiterbildung sind die Zertifikatskurse exakt auf das enge Zeitbudget von Berufstätigen ausgerichtet und lassen sich so flexibel in ein Gesamtkonzept des lebensumspannenden Lernens integrieren.

Das Zertifikat der TH Köln eines erfolgreich abgeschlossenen Zertifikatskurses ist sowohl national als auch international ein anerkannter Leistungsnachweis. Zudem ist anhand der vergebenen ECTS eine Anrechnung der Leistungen bei einem anschließenden oder parallelen Hochschulstudium grundsätzlich möglich. Das Zertifikat wird nach bestandener Projektarbeit ausgestellt.