

GWDG NACHRICHTEN 04-05|24

Mehrfaktor-
Authentifizierung

Migration auf KIX 18

GWDG Academy for Azubis

LLM Services and the
AAI Act of the EU

QA Catalogue

ZEITSCHRIFT FÜR DIE KUNDEN DER GWDG





GWDG NACHRICHTEN

04-05|24 Inhalt

.....

**4 Erhöhung der Sicherheit durch verpflichtende
Mehrfaktor-Authentifizierung 9 Migration
auf KIX 18: Vom Trouble-Ticket- zum Service-
Management-System 12 GWDG Academy for
Azubis 13 Kurz & knapp 16 LLM Services and
the AI Act of the EU 19 QA Catalogue – A Quality
Assessment Tool for Library Catalogues
25 Academy**

Impressum

.....

Zeitschrift für die Kunden der GWDG

ISSN 0940-4686
47. Jahrgang
Ausgabe 4-5/2024

Erscheinungsweise:
10 Ausgaben pro Jahr

www.gwdg.de/gwdg-nr

Auflage:
550

Fotos:
© andranik123 - stock.adobe.com (1)
© shuo - stock.adobe.com (11)
© WrightStudio - stock.adobe.com (15)
© Designsprache - stock.adobe.com (18)
© Maksym Yemelyanov - Fotolia.com (24)
© Robert Kneschke - Fotolia.com (25)
© MPI-NAT-Medienservice (3)
© GWDG (2)

Herausgeber:
Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen
Burckhardtweg 4
37077 Göttingen
Tel.: 0551 39-30001
Fax: 0551 39-130-30001

Redaktion und Herstellung:
Dr. Thomas Otto
E-Mail: thomas.otto@gwdg.de

Druck:
Kreationszeit GmbH, Rosdorf



Prof. Dr. Ramin Yahyapour
ramin.yahyapour@gwdg.de
0551 39-30130

*Liebe Kund*innen und Freund*innen der GWDG,*

die Themen Informationssicherheit und Weiterentwicklung unseres Sicherheitskonzeptes begleiten uns ständig. Gerade die Authentifizierung bei Diensten allein mit einem Nutzernamen und zugehörigem Passwort ist nicht mehr Stand der Technik und in der aktuellen Bedrohungslage nicht mehr ausreichend. Zu einfach kann ein Angreifer mit einem kompromittierten Passwort lange unentdeckt den Zugang zu Systemen bekommen und sich in der Infrastruktur ausbreiten.

*Eine Lösung ist die Nutzung von Mehrfaktor-Authentifizierung (MFA), wie man es auch vom Online-Banking kennt. Die GWDG hat sich seit Jahren auf ihre Einführung vorbereitet und MFA bereits für einige Dienste optional angeboten. Wir haben dies jedoch nicht flächendeckend verpflichtend eingeführt, da nicht alle Dienste dies unterstützen und insbesondere unsere Nutzer*innen hierzu erst befähigt werden müssen. In dieser Ausgabe berichten wir über weitere Schritte, um MFA sukzessive für alle Dienste einzuführen. Wir werden hierüber in Zukunft noch weiter berichten, wenn wir beginnen, Dienste nur noch über MFA anzubieten.*

Ramin Yahyapour

GWDG – IT in der Wissenschaft

Erhöhung der Sicherheit durch verpflichtende Mehrfaktor-Authentifizierung

Text und Kontakt:

Ralph Krimmel
ralph.krimmel@gwdg.de
Maximilian Wilhelm Scheid
maximilian-wilhelm.scheid@gwdg.de
See-Ling Wong
see-ling.wong@gwdg.de

Die zunehmende Digitalisierung hat die Notwendigkeit verstärkt, Online-Konten und sensible Informationen vor unbefugtem Zugriff zu schützen. Die Mehrfaktor-Authentifizierung (MFA) bietet eine zusätzliche Sicherheitsebene jenseits traditioneller Passwörter und ist zu einer wichtigen Sicherheitspraxis geworden. Die GWDG bietet bereits seit einigen Jahren die Möglichkeit, den Zugriff auf ihre Dienste mit MFA abzusichern. Nun werden wir schrittweise eine Verpflichtung einführen, sodass der Zugang zu ausgewählten Diensten dieser zusätzlichen Sicherheitsebene unterliegt. In diesem Artikel möchten wir darstellen, warum eine Mehrfaktor-Authentifizierung wichtig ist, wie die GWDG diese umgesetzt hat, wie Sie solch einen zweiten Faktor registrieren können und bis wann dies geschehen muss. Abschließend werden wir am Ende ein wenig Klarheit in den MFA-Vokabular-Dschungel bringen und die unterschiedlichen am Markt verfügbaren Methoden miteinander vergleichen.

WARUM MUSS ICH DAS MACHEN?

Die Motivation hinter der Einführung von MFA ist klar: Die herkömmliche Authentifizierung allein durch Benutzername und Passwort reicht oft nicht mehr aus, um die heutzutage verbreiteten „fortschrittlichen“ Angriffsmethoden abzuwehren. Passwörter können gestohlen, durch Phishing-Angriffe kompromittiert oder durch sogenannte Brute-Force-Angriffe geknackt werden. MFA fügt dem traditionellen Passwort eine weitere Sicherheitsebene hinzu. Dies kann ein Token, ein biometrisches Merkmal oder eine andere Authentifizierungsmethode sein. Ein Token ist ein physisches oder digitales Gerät, das zur Authentifizierung verwendet wird. Token erzeugen oft Einmalpasswörter oder andere Formen von kryptografischem Code. Bei der GWDG können Sie zwischen verschiedenen Methoden frei wählen: Authenticator App (TOTP), privacyIDEA Authenticator App (Push) und Sicherheitsschlüssel wie FIDO2 oder Passkeys.

Bestätigung Ihrer Identität in zwei Schritten: MFA bei Zugriff auf GWDG-Dienste

Die GWDG hat im Rahmen von MFA speziell die „Bestätigung in zwei Schritten“ eingeführt. Sie können die Sicherheit Ihres GWDG-Kontos im Accountportal <https://id.academiccloud.de> durch Hinzufügen eines zusätzlichen Authentifizierungsfaktors erhöhen. Konkret bedeutet dies, dass die Anmeldung in zwei aufeinanderfolgenden Schritten erfolgt: Zuerst geben Sie Ihr Passwort ein und dann bestätigen Sie Ihre Identität über den zweiten Faktor.

WIE RICHTE ICH MIR NUN EINEN ZWEITEN FAKTOR EIN?

Dies kann in der Academic Cloud auf verschiedene Arten erfolgen. Dazu gehen Sie im Accountportal auf das Menü „Sicherheit“ (siehe Abbildung 1). Dort haben Sie beispielsweise die Möglichkeit, Ihr Smartphone mit einer Authenticator-App zu verwenden oder Ihr Gerät über sogenannte FIDO2-Token/Passkeys für

Enhancing Security through Mandatory Multi-Factor Authentication

Increasing digitalisation has increased the need to protect online accounts and sensitive information from unauthorised access. Multi-factor authentication (MFA) offers an additional layer of security beyond traditional passwords and has become an important security practice. The GWDG has already been offering the option of securing access to its services with MFA for several years. Now we will gradually introduce an obligation so that access to selected services is subject to this additional layer of security. We will show in this article why multi-factor authentication is important, how the GWDG has implemented it, how you can register such a second factor and by when this must be done. Finally, we will tidy up the MFA buzzword jungle a bit and try to bring some clarity into it. For further inquiries or assistance, contact us at sso-support@gwdg.de.

Meine AcademicID DE EN Logout

- ACCOUNTINFORMATIONEN
- SICHERHEIT**
- EXTERNE DIENSTE
- ANDERES

PASSWORT

Hier können Sie Ihr Passwort ändern

Letzte Änderung: 24.10.2023 12:31:08

BESTÄTIGUNG IN ZWEI SCHRITTEN

Eine sichere Anmeldung erfolgt in zwei Schritten. Im ersten Schritt wird Ihr Passwort abgefragt, der zweite Schritt dient der sicheren Bestätigung Ihrer Identität bei der Anmeldung.

1_Im Accountportal unter „Sicherheit“ kann man sich die zweiten Faktoren registrieren.

BESTÄTIGUNG IN ZWEI SCHRITTEN

Eine sichere Anmeldung erfolgt in zwei Schritten. Im ersten Schritt wird Ihr Passwort abgefragt, der zweite Schritt dient der sicheren Bestätigung Ihrer Identität bei der Anmeldung.

Hier finden Sie einen Überblick über die Methoden, die für einen zweiten Schritt in Frage kommen.

- Authenticator App**
Sichern Sie Ihren Zugang mit einer Authenticator App auf Ihrem Smartphone. MEIN KONTO ABSICHERN
- Paskeys und Sicherheitsschlüssel**
Mit Hilfe von Paskeys können Sie sich mit einem Gerät auf sichere Weise authentifizieren. MEIN KONTO ABSICHERN

[Alle Optionen](#)

2_Bestätigung in zwei Schritten: Möglich mit einer Authenticator App oder einem Sicherheitsschlüssel auf dem genutzten Gerät.

den zweiten Schritt der Anmeldung zu registrieren.

Bei der Erstregistrierung wählen Sie den gewünschten Faktor aus (siehe Abbildung 2) und geben eine Beschreibung für diesen ein. Es ist empfehlenswert, das Gerätemodell einzutragen, mit dem Sie den zweiten Faktor registrieren möchten. Nach Festlegung der Beschreibung folgt die Registrierung des zweiten Faktors, bei der Sie den Anweisungen folgen müssen.

Authenticator-App (z. B. Google Authenticator, Microsoft Authenticator oder FreeOTP)

Laden Sie eine Authenticator-App Ihrer Wahl auf Ihr

Smartphone herunter. Die GWDG empfiehlt hier „FreeOTP“. Bekannte Alternativen sind „Google Authenticator“ und „Microsoft Authenticator“. Wir empfehlen den sogenannten „Cloud Sync“ abzuschalten, damit Ihre Token nicht auf die Server der Hersteller synchronisiert werden.

Scannen Sie den QR-Code, der Ihnen angezeigt wird, mit Ihrer Authenticator-App (siehe Abbildung 3). Anschließend erscheint direkt ein sechsstelliger Code mit der Beschreibung „academiccloud.de“. Zur Verifizierung müssen Sie diesen im Browserfenster eingeben. Danach sollte automatisch eine Datei heruntergeladen werden. Diese Datei enthält einen Wiederherstellungscode,

AUTHENTICATOR APP

Bitte scannen Sie den unten angezeigten QR-Code mit Ihrer Authenticator App.

Weitere Informationen zum Scannen des Codes mit Ihrer Authenticator App finden Sie [hier](#).



ABBRECHEN

QR-CODE GESCANNT

3. Bei der Registrierung der Authenticator App muss ein QR-Code gescannt werden.

mit dem Sie Ihre zweiten Faktoren alle löschen können, sollte Ihr Smartphone einmal kaputt oder verloren gehen.

FIDO2/Passkeys

Passkeys bieten eine weitere Möglichkeit für die Mehrfaktor-Authentifizierung bei der GWDG. Sie ermöglichen es Ihnen, sich zusätzlich über bestimmte Geräte wie Smartphones, Computer oder Hardware-Token zu authentifizieren. Die Einrichtung von Passkeys erfolgt normalerweise direkt auf dem Gerät selbst, daher muss dies auch für jedes Gerät neu eingerichtet werden. Hinterlegen Sie als Beschreibung des Faktors den Gerätenamen. Bei der Registrierung werden Sie gefragt, welches Gerät Sie zur Authentifizierung nutzen möchten (siehe Abbildung 4). Alle aktuellen und gängigen Betriebssysteme sowie mobile Geräte werden unterstützt.

YubiKeys als Alternative

Nutzer*innen, die kein Smartphone oder Passkey-fähiges Gerät besitzen oder dieses für diesen Zweck nicht nutzen möchten, haben die Möglichkeit, sogenannte Hardware-Token zu verwenden. Die GWDG empfiehlt die Nutzung von YubiKeys, die sich als zuverlässige Option bewährt haben. Als Partner von Yubico kann die GWDG YubiKey-Produkte zu günstigeren Konditionen beziehen. Kontaktieren Sie uns, falls Sie planen, Yubikeys in Ihrer Einrichtung einzusetzen.

privacyIDEA Authenticator App

Eine weitere Möglichkeit bietet die „privacyIDEA Authenticator App“, welche einen sogenannten Push-Faktor bietet. Bei jedem

neuen Login in die Academic Cloud erhält die privacyIDEA Authenticator App eine Push-Benachrichtigung, mit der Sie Ihre Identität einfach per Bestätigung verifizieren können. Welche Gefahren bei einem Push-Faktor bestehen, erklären wir weiter unten genauer.

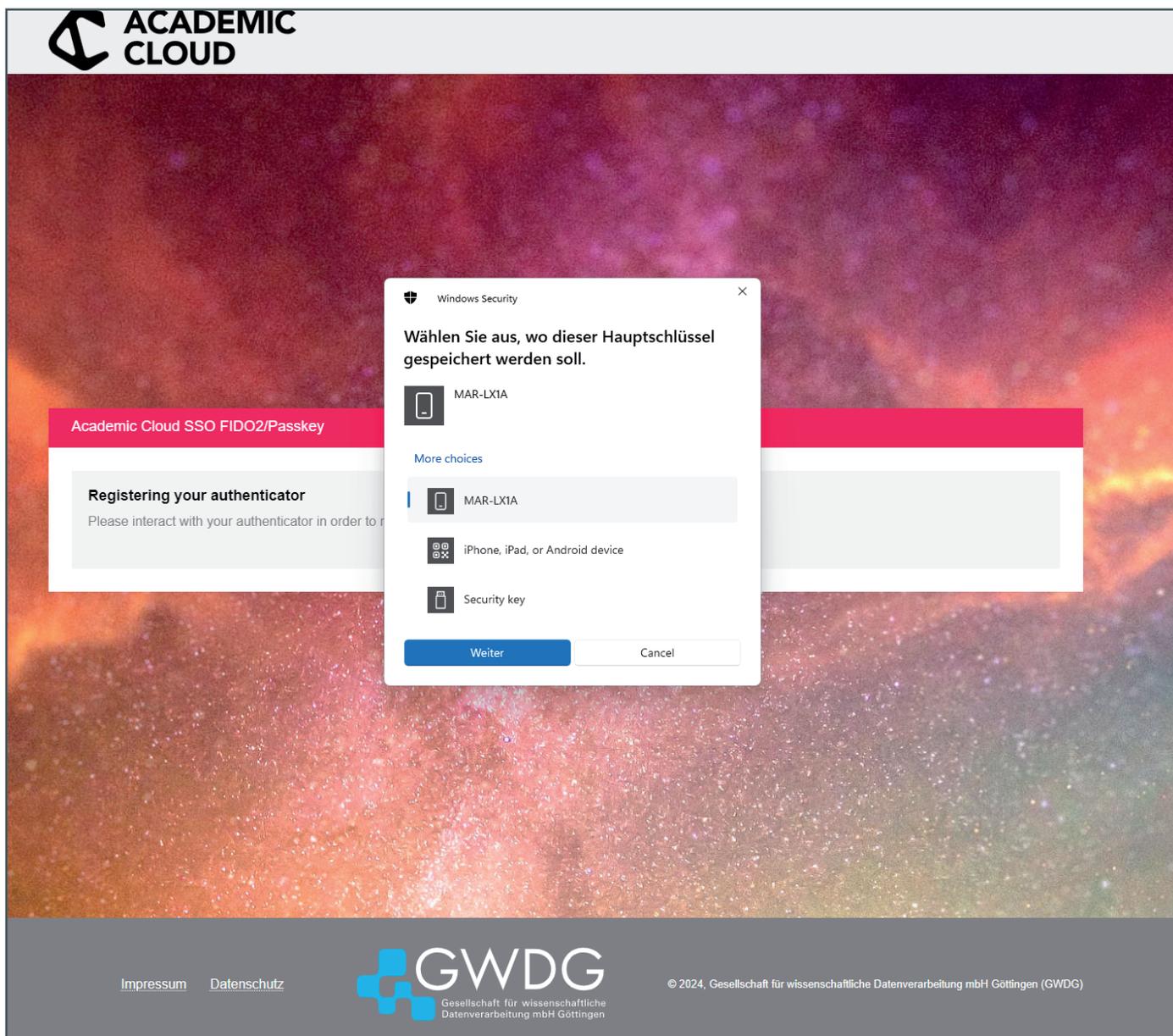
Wechsel von privacyIDEA zu eduMFA – eine hochschulübergreifende MFA-Lösung

Wie bereits in der letzten Ausgabe 3/2024 der GWDG-Nachrichten erwähnt wurde, wurde das Projekt eduMFA ins Leben gerufen, an der neben vier weiteren Partnern auch die GWDG beteiligt ist. Dabei handelt es sich um eine speziell auf die Bedürfnisse von Bildungseinrichtungen zugeschnittene Lösung zur Implementierung von MFA.

Die GWDG hat bereits das Open-Source-Projekt „privacyIDEA“ als Grundlage der MFA-Infrastruktur im Einsatz, plant aber einen Wechsel zu eduMFA und wird daher im Laufe dieses Jahres die privacyIDEA Authenticator App ersetzen. Der eduMFA-Faktor wird dann über eine separate eduMFA-App für alle gängigen Betriebssysteme verfügbar sein. Für diesen Übergang wird es eine umfassende Kommunikation sowie ausreichend Zeit zur Umstellung geben. Alle weiteren zweiten Faktoren sind vom Wechsel auf eduMFA nicht betroffen.

Was tun, wenn der zweite Faktor nicht mehr funktioniert?

Wir empfehlen besonders die Einrichtung eines zusätzlichen weiteren zweiten Faktors. Sollten Sie alle Optionen erschöpft haben, können Sie unter <https://id.academiccloud.de/mfareset> den Wiederherstellungstoken verwenden, den Sie nach der Registrierung bekommen haben. Dadurch werden alle bestehenden



4_Bei der Einrichtung eines Sicherheitsschlüssels (FIDO2/Passkeys) authentifiziert man sich direkt am Gerät. Dies ist individuell am Gerät durchzuführen und kann z. B. durch einen Fingerabdruck erfolgen.

Token gelöscht und Sie können einen neuen zweiten Faktor direkt einrichten.

Weitere Informationen zur Einrichtung finden Sie in unserer umfassenden Nutzerdokumentation: <https://docs.gwdg.de> > IT-Sicherheit > Mehrfaktor-Authentifizierung (<https://s.gwdg.de/KPmOrr>).

WANN MUSS ICH DAS MACHEN?

Seit Mitte 2022 haben Sie die Möglichkeit, Ihre Academic ID mit einem zweiten Faktor zu sichern. Bis Ende 2024 beabsichtigen wir, alle aus dem Internet erreichbaren Dienste mit MFA zu schützen, wodurch die Anmeldung nur noch mit einem zweistufigen Prozess möglich sein wird. Ab Anfang Juni werden wir beginnen, alle unsere Nutzer*innen über die bevorstehende verpflichtende Mehrfaktor-Authentifizierung über verschiedene Kommunikationskanäle zu informieren. Außerdem wird auch die Einführungsphase beginnen, in der zuerst der VPN-Zugriff ausschließlich mit MFA möglich sein wird.

MFA-VOKABULAR-DSCHUNDEL: OTP, TOKEN, TAN, PUSH, FIDO2, PASSKEYS, YUBIKEYS – WAS IST DAS ALLES NUN?

In der heutigen digitalen Welt werden wir mit einer Vielzahl von Begriffen konfrontiert, die oft im Zusammenhang mit der Mehrfaktor-Authentifizierung verwendet werden. Von OTP über Token bis hin zu FIDO2 und YubiKeys scheint der MFA-Vokabular-Dschungel manchmal verwirrend zu sein. Doch was bedeuten diese Begriffe eigentlich und wie tragen sie zur Sicherheit unserer digitalen Identität bei?

MFA und OTP: Wieso brauche ich zwei Passwörter?

MFA oder Mehrfaktor-Authentifizierung ist ein Sicherheitsmechanismus, der zusätzliche Authentifizierungsfaktoren neben dem traditionellen Passwort erfordert. Diese zusätzlichen Faktoren können etwas sein, das der Benutzer weiß (z. B. eine Sicherheitsfrage), etwas, das er besitzt (z. B. ein Smartphone) oder ein biometrisches Merkmal (z. B. der Fingerabdruck).

OTP steht für Einmalpasswort (One-Time Password) und bezeichnet eine Art von Passwort, das nur für eine begrenzte Zeit gültig ist. OTPs können auf verschiedene Weise generiert werden, einschließlich Hardware-Token, mobile Apps oder SMS. Da diese Methode meist in Verbindung mit einer Authenticator App steht, finden Sie diese auch unter diesem Stichwort. Wir nutzen ein zeitbasiertes Einmalpasswort (TOTP: Timebased One-Time Password). Durch einen speziellen Algorithmus wird alle 30 Sekunden ein neuer sechsstelliger Code erstellt, welcher für maximal eine Minute gültig ist.

Auf den ersten Blick könnte man denken, dass MFA bedeutet, dass man zwei Passwörter benötigt. Dies ist jedoch nicht ganz korrekt. Dabei geht es darum, dass der andere Faktor, z. B. neben dem Passwort, welches man auswendig gelernt hat, unabhängig von diesem ist. Deshalb sind zwei Passwörter im Sinne von MFA nicht zwei Faktoren, da es sich bei beidem um einen Wissensfaktor handelt.

TAN: Alt und bewährt

Eine TAN ist eine Transaktionsnummer, die zur Bestätigung von Transaktionen in Online-Banking-Systemen verwendet wird. Sie wird oft als zusätzlicher Sicherheitsfaktor neben Benutzername und Passwort eingesetzt, um sicherzustellen, dass Transaktionen authentisch sind und nicht von unbefugten Personen durchgeführt werden. Früher gab es sogenannte TAN-Listen, die Banken ausgestellt haben. TAN-Listen (Transaktionsnummern-Listen) enthalten eine Reihe von einmaligen Codes, die zur Bestätigung von Transaktionen oder Aktionen in Online-Banking-Systemen verwendet werden. Jedes Mal, wenn eine Transaktion durchgeführt werden soll, fordert das System den Benutzer auf, eine TAN aus der Liste einzugeben, die zuvor vom Bankinstitut bereitgestellt wurde. Der Benutzer wählt dann einen verfügbaren TAN-Code aus der Liste aus und gibt ihn in das entsprechende Feld ein. TAN-Listen gelten als ältere Methode der Transaktionsauthentifizierung und werden zunehmend durch moderne Authentifizierungsmethoden wie OTP oder Push ersetzt.

Push: Super einfach, aber mit Nachteilen

Bei einem Push-Token bekommt der/die Nutzer*in eine Push-Benachrichtigung auf sein/ihr Mobilgerät. Diese Benachrichtigung enthält normalerweise eine Aufforderung, eine spezifische Aktion durchzuführen, z. B. einen Knopf zu drücken, um den Login zu bestätigen. Die GWDG bietet dies momentan mit der privacyIDEA Authenticator App an.

Zurzeit gibt es vermehrt Angriffe, die als „Push-Bombing“ bekannt sind und im Zusammenhang mit der Mehrfaktor-Authentifizierung stehen. Dabei kennt ein Angreifer in der Regel bereits einen der Authentifizierungsfaktoren, wie beispielsweise Benutzername und Passwort. Anschließend kann der Angreifer Push-Benachrichtigungen anfordern, die dann an das Smartphone des Nutzers bzw. der Nutzerin gesendet werden. Häufig erscheinen diese Benachrichtigungen als neues Fenster oder Popups auf dem Smartphone und fordern dazu auf, die Anfrage zu bestätigen oder abzulehnen. Diese „Bist du das wirklich?“-Bildschirme sind so verbreitet, dass Benutzer*innen oft einfach darauf klicken, um sie zu schließen und mit ihrer Arbeit fortzufahren.

FIDO2 und Passkeys: Technische Details zu modernen Authentifizierungsformen

FIDO2 (Fast IDentity Online) ermöglicht eine sichere und nutzerfreundliche Authentifizierung ohne Passwörter. Bei der Verwendung von FIDO2-basierten Authentifizierungsmethoden werden anstelle von herkömmlichen Passwörtern kryptografische Schlüsselpaare verwendet. Ein Schlüsselpaar besteht aus einem öffentlichen und einem privaten Schlüssel. Der öffentliche Schlüssel wird mit dem Dienst, bei dem Sie sich anmelden möchten, geteilt, während der private Schlüssel auf Ihrem Gerät gespeichert bleibt. Wenn Sie sich auf einer Webseite oder bei einem Dienst anmelden, sendet der Dienst eine Anfrage an Ihr Gerät, um sich zu authentifizieren. Ihr Gerät verwendet dann Ihren privaten Schlüssel, um eine Signatur für die Anfrage zu generieren, die der Dienst mit Ihrem öffentlichen Schlüssel überprüfen kann. Auf diese Weise erfolgt die Authentifizierung ohne die Offenlegung des privaten Schlüssels.

Passkeys bei der GWDG

Passkeys sind ein Teil des FIDO2-Authentifizierungsstandards. Tatsächlich unterscheidet sich das, was die GWDG als „Passkeys“ bezeichnet, technisch gesehen etwas von den herkömmlichen Passkeys. Dennoch beruhen beide auf dem gleichen Prinzip. Im eigentlichen Sinne impliziert die Verwendung von Passkeys bereits eine zweistufige Authentifizierung, bei der sowohl Identifikation (Wer bin ich?) als auch die Authentifizierung (Bin ich wirklich ich?) in einem Schritt übermittelt werden. Bei der GWDG dient der Faktor „Passkeys“ als zweiter Faktor im klassischen Sinne und ist technisch betrachtet ein FIDO2-Token. Das bedeutet, dass der erste Schritt der Anmeldung weiterhin das traditionelle Passwort erfordert und im zweiten Schritt der Token zum Einsatz kommt.

YubiKey: Der vielgenutzte Hardware-Token

Der YubiKey ist ein kompaktes Gerät, das in einen USB-Anschluss gesteckt wird. Er kann auch in Form eines NFC-fähigen Schlüsselanhängers oder einer Karte erhältlich sein. Sobald er mit einem Computer oder einem mobilen Gerät verbunden ist, kann der Benutzer durch Drücken einer Taste auf dem YubiKey die Mehrfaktor-Authentifizierung durchführen.

WEITERE INFORMATIONEN UND SUPPORT

Wie bereits erwähnt, werden wir ab Anfang Juni beginnen, alle unsere Nutzer*innen über die bevorstehende verpflichtende Mehrfaktor-Authentifizierung über verschiedene Kommunikationskanäle zu informieren. Beginnen werden wir mit der MFA-Verpflichtung für unseren VPN-Dienst. Darauf folgend wird die Anmeldung mit MFA an weiteren Diensten verpflichtend werden, wie z. B. Uni-Dienste (eCampus, studIP etc.). Am Ende des Jahres sollen dann alle aus dem Internet erreichbaren Dienste mit MFA abgesichert sein. Über den aktuellen Fortschritt halten wir Sie auf <https://status.gwdg.de> sowie über E-Mail-Verteiler und Mailinglisten auf dem Laufenden.

Bei weiterführenden Fragen zum Thema Mehrfaktor-Authentifizierung bei der GWDG oder auftretenden Problemen kontaktieren Sie uns gerne direkt über sso-support@gwdg.de. ●

Migration auf KIX 18: Vom Trouble-Ticket- zum Service-Management- System

Text und Kontakt:

Dr. Konrad Heuer
konrad.heuer@gwdg.de
0551 39-30313

Mit dem anstehenden Versionswechsel auf KIX 18 wird am Standort Göttingen das bewährte Trouble-Ticket-System KIX 17 um neue Funktionen erweitert. Seitens der GWDG wird durch neue Strukturen in KIX 18 angestrebt, den Nutzer*innen-Support weiter zu verbessern. Zeitplan und Umfang der Arbeiten werden nachfolgend beschrieben.

EINLEITUNG

In den GWDG-Nachrichten vom April und September 2005 [1, 2] wurde die neue Funktions-E-Mail-Adresse support@gwdg.de vorgestellt, über die Nutzer*innen Anfragen an die GWDG stellen konnten, ohne GWDG-Personal direkt zu kontaktieren. Aber erst mit der April-Ausgabe im Folgejahr [3] ließ die GWDG die „Katz aus dem Sack“ und beschrieb, dass hinter dieser Funktionsadresse ein vollständiges Trouble-Ticket-System in den Probebetrieb gestartet war; aus jeder E-Mail-Nachricht an support@gwdg.de wurde ein Trouble-Ticket.

GWDG-intern waren längere Beratungen und Gespräche auf allen Ebenen vorausgegangen, ob mit einem solchen System nicht der bewährte persönliche oder telefonische Kontakt zwischen Nutzer*innen und GWDG-Personal geschwächt werden würde, der damals noch eine wesentlich größere Rolle spielte, waren doch zu der Zeit die öffentlichen Räume der GWDG noch deutlich stärker frequentiert als in späteren Jahren.

Aber die Vorteile eines Trouble-Ticket-Systems bewährten sich im Laufe der Zeit, u. a.:

- Anfragen erreichen während der gesamten Öffnungszeiten der GWDG (werktags von 7:00 bis 21:00 Uhr, samstags und sonntags von 10:00 bis 18:00 Uhr) immer eine planmäßig eingeteilte Person, welche eingehende Tickets liest, eventuell gleich beantwortet oder an eine zuständige Gruppe von Mitarbeiter*innen weiterleiten kann.
- Jedes Ticket hat eine im System nachvollziehbare Historie und kann mit Kommentaren versehen werden und so die Zusammenarbeit unterschiedlicher Bereiche oder Personen erleichtern.
- Automatische Antworten signalisieren Anfragenden einen erfolgreichen E-Mail-Eingang, Antwortvorlagen erleichtern dem GWDG-Personal die Beantwortung von Standardanfragen.
- Eine Häufung von Anfragen zu einem bestimmten Thema ist leicht erkennbar und signalisiert organisatorische oder technische Probleme oder Störungen.

Da die GWDG wie viele andere IT-Einrichtungen auch ein Service-Management nach den „best practices“ der „Information Technology Infrastructure Library (ITIL)“ anstrebt, ist ein Trouble-Ticket-System tatsächlich unverzichtbar.

Seitens der GWDG fiel damals die Entscheidung für eine auf Open Source basierende Lösung, das „Open Ticket Request System (OTRS)“ [4]. Neben einer reinen Open-Source-Basisversion existiert nach wie vor eine kommerzielle Variante der OTRS AG. In Göttingen erfolgte vor einigen Jahren dann jedoch schon der Wechsel von OTRS zur Software KIX der damaligen Cape IT GmbH, die in zwischen ihren Namen in KIX Service Software GmbH [5] geändert hat.

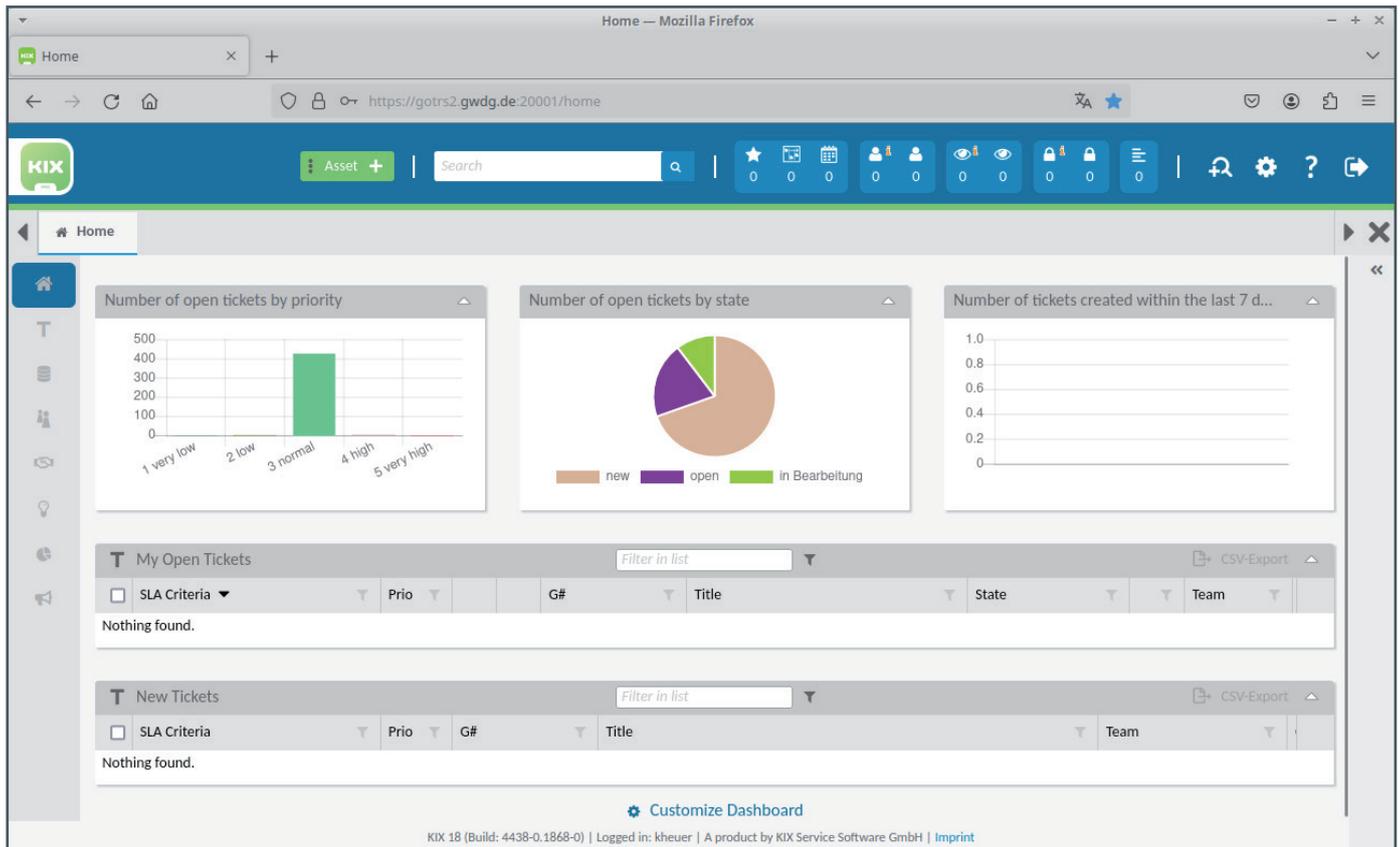
Wichtiges Kriterium für die GWDG war von Anbeginn, dass die Bedienung des Systems mit einem Webbrowser wie Firefox u. a. möglich ist (siehe Abbildung 1), also kein dediziertes Client-Programm auf den Endgeräten aller Ticket-Bearbeiter*innen installiert werden muss.

KIX startete als OTRS-Ableger [6]. Inzwischen wird immer mehr OTRS-Code durch Neuentwicklungen ersetzt, und das System wird um neue Funktionalitäten erweitert, so dass die Software heute als Service-Management-System auftritt.

Das KIX-System wird offiziell von der GWDG betrieben, jedoch technisch stark von der Abteilung IT der Universitäts-Zentralverwaltung unterstützt. KIX ist mandantenfähig, kann also unterschiedliche nutzende Einrichtungen gegeneinander abschotten, und neben der GWDG und der Abteilung IT nutzen fast 20 weitere

Migrating to KIX 18: Trouble Ticket System Evolves to Service Management System

With upgrading KIX to version 18 on the campus, the proven trouble ticket system KIX 17 will be extended and will provide new features. The GWDG will try ambitiously to improve user support by new internal KIX structures. The schedule and scope of the work are described below.



1_Startseite („Dashboard“) für Ticket-Bearbeiter*innen in KIX 18

Einrichtungen die Göttinger Installation.

BESONDERE ZIELE DER MIGRATION

Migrations- und Upgradeprozesse bei Softwareprodukten haben in aller Regel grundsätzlich das Ziel, diese Produkte auf einem ausreichend aktuellen Stand der Technik und in einem wartbaren und gewarteten Zustand zu halten.

Mit der Migration von KIX 17 auf KIX 18 wird am Standort Göttingen jedoch auch das Ziel verfolgt, die Erweiterung vom reinen Trouble-Ticket-System hin zum Service-Management-System zu vollziehen, so dass neben den Tickets auch weitere Daten im System verwaltet werden. Dazu werden insbesondere ein gemeinsamer Servicekatalog der IT-Dienstleister auf dem Campus und eine Buchführung über eingesetzte Endgeräte und deren Ausstattung gehören (Configuration Management Database – CMDB).

HERAUSFORDERUNGEN DER MIGRATION

Der Wechsel von der jetzt eingesetzten Version KIX 17 nach KIX 18 bringt durchaus Herausforderungen mit sich, denn die Version 18 enthält viele Neuentwicklungen und ist deutlich weiter vom OTRS-Ursprung entfernt.

Sowohl für die Administration des Systems als auch für das IT-Support-Personal sind daher teils umfangreiche Schulungen und Einweisungen erforderlich und vorgesehen. Insbesondere die Rechteverwaltung wird deutlich flexibler und komplexer.

Aus der Sicht der Hilfesuchenden, die Anfragen stellen, sind die Änderungen jedoch weitgehend transparent.

BESONDERE HERAUSFORDERUNGEN FÜR DIE GWDG

Die grundlegende Konfiguration des Ticket-Systems wurde mit dem Start im Jahre 2004 an die damaligen Bedürfnisse der GWDG angepasst und über die Jahre kontinuierlich erweitert. Nach nun etwa 20 Jahren zeigt sich jedoch, dass ein neuer Weg beschritten werden muss.

Hierzu muss man bedenken, dass eingehende Tickets, die nicht sofort beantwortet können, in einem Ticket-System nach Themen geordnet und bestimmten Personen bzw. Gruppen zur weiteren Bearbeitung übergeben werden müssen.

Der traditionelle Begriff, mit dem OTRS und KIX bis Version 17 den Mechanismus zur Gruppierung von Tickets beschreiben, lautet „Queue“, zu Deutsch „Warteschlange“. Und tatsächlich ist der Begriff „Warteschlange“ ja treffend, denn Tickets, die nicht gleich beantwortet werden können, befinden sich in einem Wartezustand, bis sie schließlich doch bearbeitet werden können.

Mit der Einführung von OTRS bei der GWDG wurden entsprechend dem damaligen Dienstleistungsangebot passende thematische Queues eingeführt, für die dann die jeweils zuständigen Personen als Bearbeiter*innen hinzugefügt wurden. Es wurde darauf geachtet, dass jede Queue personell ausreichend „unterfüttert“ war, so dass auch bei Urlaub und Krankheit einzelner Personen möglichst immer jemand die Beantwortung von Tickets übernehmen konnte. Die Berechtigungen auf den Zugriff der Queues wurden sorgfältig vergeben, um das Risiko möglicher Fehlbedienungen zu reduzieren.

Natürlich wurde auch den Bedenken des Betriebsrats Rechnung getragen, der in Sorge um mögliche Nachteile für die GWDG-Bediensteten war.

Im Laufe der Jahre mit wachsendem Leistungsspektrum der GWDG entwickelte sich der Begriff Queue in der Wahrnehmung der Mitarbeiter*innen immer mehr zu einem reinen Ordnungsbegriff. Die Zahl der Queues wuchs immer weiter, und die Queues wurden immer spezieller. Infolge dessen haben viele der Mitarbeiter*innen heute eine große Zahl von Queues zugeordnet bekommen und zu beobachten, ohne dass die personelle Unterfütterung der Queues entsprechend mitgewachsen ist.

Das bringt betriebliche Nachteile mit sich: Tickets können z. B. übersehen oder falsch in Queues eingeordnet werden, weil Queues verwechselt werden. Sind Queues personell schwach besetzt, bleiben darin enthaltene Tickets tendenziell zu lange liegen. Auch die Rechtevergabe wird fehleranfälliger, weil Dinge übersehen werden können.

KIX 18 bringt eine erfreuliche Klarstellung mit sich, denn Queues heißen dort nicht mehr Queues, sondern Teams. Das bringt zum Ausdruck, dass das primäre Ziel der Gruppierung sein soll, Personengruppen zusammenzufassen, die gleiche oder verwandte Themen bearbeiten. Die neue Strukturierung der Teams in KIX 18 soll also die tatsächliche organisatorische Struktur der GWDG möglichst gut abbilden.

Damit geht eine erhebliche Reduzierung der Zahl der Queues bzw. Teams im Ticket-System einher. Die Namen der angebotenen Services sollen im Rahmen der Wandlung von KIX vom reinen Trouble-Ticket- zum Service-Management-System als zusätzliches Ordnungsmerkmal eingeführt werden, so dass Tickets nach Services klassifiziert werden können.

Mit dem Schritt der Reduzierung von Queues/Teams zugunsten einer besseren personellen Unterfütterung der neuen KIX-Teams soll auch eine insgesamt zuverlässigere Abarbeitung von Tickets erreicht werden.

Für die Migration vom jetzigen in das neue System bedeutet die Reduktion jedoch Mehraufwand. Da ja ein bestimmter Anteil der Tickets aus dem Altsystem in das neue System übernommen werden muss, müssen die neuen Teams als neue Queues im bisherigen KIX neu eingerichtet und die Tickets aus den bisherigen Queues dorthin verschoben werden, so dass sie dann später einfach migriert werden können.

Es wird also notwendig, die bisherigen Queues eindeutig auf die neuen Queues abzubilden. Ferner müssen die Servicennamen zur Ticket-Kategorisierung schon im jetzigen KIX eingeführt werden. Dafür müssen die Servicennamen ticketgerecht finalisiert werden.

ZEITPLAN

Folgender Zeitplan ist für die Migration auf KIX 18 vorgesehen:

| TERMIN (CA.) | ARBEITSSCHRITT |
|----------------|--|
| April 2024 | Festlegung der neuen Teams-Struktur |
| Mai 2024 | Einrichten der neuen Queues/Teams in KIX 17 |
| Mai 2024 | Festlegung der Abbildung alter auf neue Queues |
| Juni 2024 | Einführung der Ticket-Kategorien in KIX 17 |
| Juli 2024 | Verschiebung der Tickets in KIX 17 |
| August 2024 | Probetrieb mit den neuen Queues in KIX 17 |
| August 2024 | Schulung der Administrator*innen |
| September 2024 | Migration nach KIX 18 |
| September 2024 | Schulung der Mitarbeitet*innen |

Tabelle 1: Zeitplan für die Migration

FAZIT UND SCHLUSSWORT

Die Migration von KIX 17 nach KIX 18 ist insgesamt und insbesondere für die GWDG ein großer Schritt. Wir sind zuversichtlich, dass unsere Kund*innen und Nutzer*innen nach der vollzogenen Migration von den Vorteilen profitieren können, die wir erreichen möchten.

Gleichzeitig bitten wir schon jetzt um Verständnis dafür, dass die umfangreichen Umstellungsarbeiten vorübergehend vermutlich auch zu „Reibungsverlusten“ und längeren Antwortzeiten führen können.

REFERENZEN

- [1] „Neue E-Mail-Adresse zur Annahme von Problemmeldungen“, GWDG-Nachrichten 4/2005 (<https://gwdg.de/about-us/gwdg-news/archiv/gn0504.pdf#page=3>)
- [2] „Neue E-Mail-Adresse zur Annahme von Problemmeldungen“, GWDG-Nachrichten 9/2005 (<https://gwdg.de/about-us/gwdg-news/archiv/gn0509.pdf#page=3>)
- [3] „Trouble-Ticket-System OTRS im Probetrieb“, GWDG-Nachrichten 4/2006 (<https://gwdg.de/about-us/gwdg-news/archiv/gn0604.pdf#page=3>)
- [4] <https://de.wikipedia.org/wiki/OTRS>
- [5] <https://www.kixdesk.com>
- [6] <https://www.datacenter-insider.de/was-ist-aus-dem-otrs-fork-kix-geworden-a-559257>



GWDG Academy for Azubis

Text und Kontakt:

Nikolaj Kopp
nikolaj.kopp@gwdg.de
0551 39-30286

Seit Oktober 2023 bietet die GWDG im Rahmen des Kursprogramms ihrer Academy spezielle Kurse für Auszubildende an, die sich gezielt an angehende Fachinformatiker*innen, IT-Systemelektroniker*innen und Elektroniker*innen richten und alltägliche Themen der Ausbildungsgänge bei der GWDG behandeln. Das breite Dienstspektrum der GWDG als Rechen- und IT-Kompetenzzentrum wird dabei genutzt und die z. T. langjährige Erfahrung der Trainer*innen an die Auszubildenden weitergegeben. Als einrichtungsübergreifender Lernort für Auszubildende der Max-Planck-Gesellschaft, der Universität Göttingen und der anderen Wissenschaftseinrichtungen am Göttingen Campus fördert die *GWDG Academy for Azubis* zudem die Zusammenarbeit.

ZIELSETZUNG

Wie in der Max-Planck-Gesellschaft, der Universität Göttingen und der Universitätsmedizin Göttingen werden auch bei der GWDG schon seit Jahren Fachinformatiker*innen in unterschiedlichen Fachrichtungen ausgebildet. Beim alljährlichen MPG-Ausbilder*innentreffen der Elektro-, Elektronik- und IT-Berufe im April 2023 kam die Idee auf, Kurse gezielt für Auszubildende unserer Gesellschafter und Partner im Göttinger Rechenzentrum anzubieten. Auszubildende sollen so die Möglichkeit bekommen, von der breiten Expertise der GWDG-Mitarbeiter*innen zu profitieren und vertiefte Einblicke in die alltäglichen Themen eines Rechenzentrums zu bekommen, zu denen insbesondere Rechenzentrumstechnik, Netzwerkbetrieb und Anwendungsentwicklung gehören. Außerdem bietet solch ein zentrales Angebot die Gelegenheit, Auszubildende aus anderen Einrichtungen kennenzulernen und die Vernetzung untereinander zu stärken. Dabei sollen nicht nur die angehenden Fachinformatiker*innen angesprochen werden, sondern explizit auch (IT-System-)Elektroniker*innen.

ERSTE ERFAHRUNGEN

Die ersten Kurse für Auszubildende starteten im Herbst letzten Jahres. So fanden im Oktober ein zweitägiger Storage-Kurs zu den Grundlagen im Umgang mit unterschiedlichen Speichersystemen statt und im November ein ebenfalls zweitägiger Rechenzentrumstechnik-Kurs, in dem die Auszubildenden den Betrieb eines Rechenzentrums in der Theorie und in Form einer umfangreichen Führung auch in der Praxis kennengelernt haben. Beide Kurse waren bereits wenige Tage nach ihrer Ankündigung ausgebucht und werden bzw. wurden daher in diesem Jahr mehrmals wiederholt, um auch noch weiteren Auszubildenden die Möglichkeit einer Teilnahme zu bieten. Im Februar 2024 wurde außerdem ein dreitägiger Kurs zu Virtuellen Cloud-Infrastrukturen durchgeführt, in dem sich die Auszubildenden sowohl mit den Cloud-Angeboten der GWDG als auch mit externen Cloud-Anbietern wie AWS, Google und Microsoft beschäftigt haben. Alle Kurse der *GWDG Academy*

for Azubis legen ihren Fokus neben der notwendigen Theorie vor allem auf praktische Beispiele, um den Umgang mit den jeweiligen Systemen zu schulen. Hier kann ganz besonders davon profitiert werden, dass die Trainer*innen z. T. mehrjährige Erfahrung in ihrem Spezialgebiet mitbringen und diese in ihren Kursen teilen.

WEITERE PLANUNG

Aufgrund der positiven Resonanz der Teilnehmer*innen und der hohen Anmeldezahlen sind aktuell noch weitere Kurse für Auszubildende in Vorbereitung. So waren bereits für das vergangene Jahr ein Netzwerk-Kurs und ein Programmierung-Kurs geplant, die beide aufgrund des Sicherheitsvorfalls leider kurzfristig verschoben werden mussten und nun in diesem Jahr nachgeholt werden. Auch zu weiteren für Auszubildende der o. g. Fachrichtungen interessanten und wichtigen Themen sind neue Kursangebote denkbar und werden derzeit evaluiert. Alle in den Ausbildungskursen behandelten Themenfelder finden sich in den Ausbildungsrahmenplänen der jeweiligen Ausbildungsberufe wieder.

GWDG Academy for Azubis

The new *GWDG Academy for Azubis* courses offer a wide range of topics for trainees in the IT and electronics professions. Launched in October 2023, the first three courses on storage systems, datacenter technology and virtual cloud infrastructures have already taken place and met with great interest. For 2024, at least five courses in total are planned. Registrations are possible via the GWDG Academy Portal under <https://academy.gwdg.de> for all trainees in the IT and electronics professions of the Max Planck Society, University of Göttingen and all other scientific institutions on the Göttingen Campus in the required minimum year of training .

Termine der nächsten Kurse für Auszubildende

Folgende Termine sind aktuell für die nächsten Kurse für Auszubildende geplant:

- Rechenzentrumstechnik: 18.06. – 20.06.2024, 9:00 – 16:00 Uhr
- Grundlagen der Programmierung: 01.07. – 03.07.2024, 9:00 – 16:00 Uhr
- Virtuelle Cloud-Infrastrukturen: 08.10. – 10.10.2024, 9:00 – 16:00 Uhr
- Storage: 23.10. – 24.10.2024; 9:00 – 15:30 Uhr

Weitere Kurstermine sind in Planung und werden frühzeitig bekanntgegeben.

ANMELDUNG

Eine Anmeldung zu den Kursen ist über das Portal der GWDG Academy unter <https://academy.gwdg.de> möglich. Voraussetzungen für die Anmeldung sind ein Benutzeraccount und der Status als Auszubildende*r in dem jeweils angegebenen Mindestlehrjahr der betreffenden Ausbildungsgänge. Um in den Kursen bestmöglich auf alle Auszubildenden eingehen zu können, ist die Teilnehmer*innenzahl pro Kurs auf neun begrenzt. Wenn nicht anders angekündigt, finden alle Kurse als reine Präsenzkurse in Deutsch im Multimediaraum 01 des Göttinger Rechenzentrums, Burckhardtweg 4, statt. ●

Kurz & knapp

Zukunft von Exchange und SharePoint On-Premises bei der GWDG

Uns erreichen immer wieder Fragen zur Zukunft unserer Exchange- und SharePoint-Angebote. Microsoft drängt seit vielen Jahren seine Kund*innen in die Cloud und mindestens ebenso lange hält sich das Gerücht einer Abkündigung der lokal betriebenen On-Premises-Produkte. Dies ist aktuell jedoch nicht der Fall. Für Exchange hat Microsoft für nächstes Jahr eine weitere neue On-Premises-Version mit entsprechendem langfristigen Support angekündigt und für SharePoint gibt es mit der Subscription Edition bereits einen Nachfolger zu unserer eingesetzten On-Premises-Variante. Für beide Produkte planen wir, auf die neuen Versionen zu migrieren, und werden rechtzeitig dazu einen konkreten Zeitplan kommunizieren. Die GWDG beabsichtigt, beide Produkte bis auf Weiteres weiterhin lokal zu betreiben und anzubieten. Im E-Mail-Bereich haben wir zudem mit Open-Xchange eine mögliche Alternative zu Exchange im Angebot, die bei Bedarf längerfristig lokal betrieben werden kann. Für Fragen oder gerne auch bei Interesse an Open-Xchange schreiben Sie bitte eine E-Mail an support@gwdg.de.

Buck

Ablösung von Rocket.Chat durch Matrix

Die GWDG stellt mit dem Dienst „Rocket.Chat“ mit gleichnamiger Software seit 2017 eine Chat-Plattform für den Austausch ihrer Nutzer*innen mit- und untereinander sowie mit externen Partnern und Einrichtungen zur Verfügung.

Mit fortschreitender, technischer Entwicklung und breiter Annahme dieses Angebots wird es nun Zeit, das Angebot zu modernisieren. Im September 2024 wird die GWDG daher ihr bisheriges Chat-Angebot ändern und dafür dann als neue Plattform einen Dienst mit der Software „Matrix“ anbieten.

Der bisherige Dienst hat sich über Jahre in seinem offenen Betrieb und seiner Funktion als vielfältige Drehscheibe in der Kommunikation innerhalb der GWDG, der GWDG zu ihren

Nutzer*innen, der Nutzer*innen in Forschung und Lehre untereinander und mit externen Partnern und Einrichtungen bewährt.

Gestartet als „Testballon“ nach einer internen Evaluation für die Verbesserung der Kommunikation innerhalb der GWDG nahm die Beliebtheit des Dienstes schnell zu und wurde quasi über Nacht mit Beginn der Pandemie in 2020 zu einem zentralen Kommunikationsmittel für viele Nutzer*innen und Einrichtungen.

Die offene Natur und der einfache Zugang werden auch in dem neuen Angebot erhalten bleiben. Der neue Dienst wird nicht nur die weitere Versorgung von – in Spitzenzeiten – über 3.000 Nutzer*innen gleichzeitig weiterhin sicherstellen, sondern in der Umstellungsphase auch eine Möglichkeit anbieten, viele Inhalte aus dem jetzigen Dienst nach Matrix zu migrieren. Details zur Übergangsphase und Übernahme von Inhalten werden frühzeitig in einer gesonderten Ankündigung bekanntgeben. Es wird angestrebt, den über 30.000 Accounts mit über 30 Millionen Nachrichten eine Möglichkeit zu geben, die Kommunikation auf der neuen Plattform unter Matrix fortzusetzen, ohne auf wesentliche Inhalte verzichten zu müssen.

Die GWDG modernisiert ihr Chat-Angebot, um für Nutzer*innen aus Forschung und Lehre und deren Partner, mit der Erfahrung aus dem jetzigen Betrieb, weiterhin ein zentrales, einfaches und breit zugängliches Angebot für die Kommunikation anbieten zu können.

Wegmann

Aktualisierung von Chat AI (ehemals LLM-Service)

Chat AI (ehemals LLM-Service) [1] bietet seit Anfang Mai eine komplett neu gestaltete Benutzeroberfläche mit mehr Funktionen und vielen neuen Features. Sie können jetzt den System-Prompt für alle unsere intern gehosteten Modelle festlegen, was Ihnen uneingeschränkte Kontrolle über die Antworten des Modells gibt.

Darüber hinaus können Sie den Chat-Verlauf auf Ihrem Gerät speichern und später wieder laden, um sicherzustellen, dass Sie Ihre Arbeit nicht unerwartet verlieren. Wir sichern Ihnen zu, dass Ihre Daten niemals auf unseren Servern gespeichert werden und nur auf Ihrem Gerät gespeichert werden,

wenn Sie dies explizit anfordern.

Die neue Oberfläche basiert auf einem speziell entworfenen Theme, inspiriert durch die Partnerschaft zwischen der GWDG und KISSKI, und bietet eine Dark-Mode für unsere lichtempfindlichen Nutzer*innen. Die Sprache kann leicht auf Deutsch umgestellt werden.

Neben der neuen Benutzeroberfläche bietet Chat AI jetzt auch eine OpenAI-kompatible API für ausgewählte Forscher*innen auf Anfrage. Sie können sich für den API-Zugriff über die Chat AI-Dienstwebseite [2] bewerben, die über die KISSKI-Dienstliste [3] zugänglich ist.

[1] <https://chat-ai.academiccloud.de>

[2] <https://kisski.gwdg.de/leistungen/2-02-llm-service>

[3] <https://kisski.gwdg.de/leistungen>

Doost Hosseini

Änderungen an der Verfügbarkeit von Chat-AI-Modellen

Seit dem Start im Februar 2024 bietet unser Service „Chat AI“ unter <https://chat-ai.academiccloud.de> den Nutzer*innen der Academic Cloud den Zugriff auf State-of-the-Art Large Language Models (LLMs). Dies umfasst Open-Source-Modelle, die auf der GWDG/KISSKI-Infrastruktur gehostet werden, und externe Modelle wie GPT-4. Da der Dienst in Deutschland an Popularität gewonnen hat, haben wir neue Modelle und Funktionen hinzugefügt, um der steigenden Nachfrage gerecht zu werden.

Die folgende Änderung betrifft nicht unsere Nutzer*innen in Niedersachsen und Mitglieder von Max-Planck-Instituten. Während wir gerne weitere Ressourcen aus unserer eigenen Infrastruktur bereitstellen, um den Zugriff auf unsere Open-Source-Modelle sicherzustellen, ist unser Budget für die Nutzung von externen Dienstleistungen leider begrenzt. Daher können wir externe (OpenAI) Modelle unentgeltlich nur noch für Nutzer*innen in Niedersachsen und Mitglieder aller Max-Planck-Institute in Deutschland anbieten.

Für Nutzer*innen außerhalb Niedersachsens, die nicht einem Max-Planck-Institut angehören, beachten Sie bitte, dass wir seit dem 24. Mai 2024 keinen unentgeltlichen Zugriff mehr auf die Modelle „OpenAI GPT-3.5“ und „OpenAI GPT-4“ anbieten können. Wenn Sie Interesse haben, die OpenAI-Modelle weiter zu nutzen, können Ihre Einrichtung und die GWDG einen Nutzungsvertrag abschließen – wir geben die Kosten für die OpenAI-Modelle mit geringem Overhead weiter. Unsere Open-Source-Modelle wie bspw. Llama3 bleiben allerdings für Sie weiterhin kostenlos verfügbar! Diese Modelle erzielen ebenfalls eine sehr gute Leistung und werden vom Projekt des KI-Servicezentrums KISSKI (<https://kisski.gwdg.de>) gefördert.

Wir sind bemüht, unseren Dienst kontinuierlich zu verbessern und sicherzustellen, dass er allen unseren Nutzer*innen zugutekommt. Vielen Dank für Ihr Verständnis und Ihre Unterstützung

Doost Hosseini

Neuer Apple-Rahmenvertrag

Um der steigenden Verbreitung von Apple-Geräten im Forschungsumfeld Rechnung zu tragen, hatte die GWDG die Aufgabe übernommen, einen neuen Apple-Rahmenvertrag für

Wissenschaftseinrichtungen auszuschreiben. Dieser folgt für alle Einrichtungen, die ihre Teilnahme erklärt haben, nahtlos auf den derzeitigen Apple-Rahmenvertrag, der zum 31. März 2024 ausgelaufen ist.

Den Zuschlag hat, nach Ablauf aller erforderlichen Fristen, erneut die Firma Cancom [1] erhalten, die das wirtschaftlichste Angebot abgegeben hat.

Am Rahmenvertrag nehmen 53 Wissenschaftseinrichtungen teil und er umfasst sowohl den Kauf von Apple-Hardware als auch optionalen Wartung und First Level Support.

Außerdem kann auf Wunsch Guthaben für das Volume Purchase Program (VPP) [2] für Apps erworben werden. Darüber hinaus wird auch das Zubuchen des Automated Device Enrollments (ADE) [3] zur Nutzung mit der in der Einrichtung verwendeten MDM-Lösung über den durch Cancom bereitgestellten Webshop möglich sein.

Im Rahmen der optionalen Wartungsverlängerung steht für den neuen Rahmenvertrag auch die Option zur Verfügung, Apple Care for Enterprise [4] für die beschafften Geräte dazubuchen. Apple Care for Enterprise umfasst nicht nur einen weltweiten [5] Vor-Ort-Hardware-Support, sondern auch eine Supportnummer, die zu allen Apple-Hard- und -Software betreffenden Problemen kontaktiert werden kann und somit auch eine Entlastung für den Support in den Einrichtungen darstellt.

Der Rahmenvertrag hat zunächst eine Laufzeit von zwei Jahren und kann in gegenseitigem Einvernehmen jeweils um ein Jahr bis zum 31.03.2028 verlängert werden.

[1] <https://www.cancom.de>

[2] <https://support.apple.com/de-de/102073>

[3] <https://support.apple.com/de-de/102300>

[4] <https://www.apple.com/de/support/professional/enterprise>

[5] Liste der Länder, in denen ACE nutzbar ist:

<https://support.apple.com/choose-country-region/enterprise>

Heider

Kursprogramm der GWDG Academy für zweite Halbjahr 2024

Das Kursprogramm der GWDG Academy für das zweite Halbjahr 2024 wurde jetzt unter <https://academy.gwdg.de> veröffentlicht. Es enthält eine Vielzahl neuer Kurse, insbesondere aus den Bereichen Scientific Computing bzw. High-Performance Computing und KI-Dienste. Alle Kurse finden entweder in einem geeigneten Online-Format, das sich in den letzten Jahren bewährt hat, oder als Präsenzkurs statt. Nähere Informationen dazu finden Sie bei den jeweiligen Kursen.

Otto

Förderprojekt der MISSION KI zu Datenaustauschstandards auf dem FDO Summit in Berlin präsentiert

(News der MISSION KI vom 20.03.2024)

Eine wettbewerbsfähige Datenökonomie erfordert einen reibungslosen Datenaustausch. Darum hat MISSION KI [1], eine Initiative der Deutschen Akademie der Technikwissenschaften und des Bundesministeriums für Digitales und Verkehr, im

Januar 2024 fünf der renommiertesten deutschen Institute im Bereich der Datennormierung damit beauftragt, eine FDO-Infrastruktur für die sektorübergreifende Vernetzung von Datenräumen zu entwickeln. Zwei Monate später stellt das Forschungskonsortium nun auf dem International FAIR Digital Objects Implementation Summit [2], 20.-21. März 2024, Berlin, erstmals einen Demonstrator der FDO-Infrastruktur vor.

Das Projekt vereint die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), den Daten-Service-Anbieter IndiScale, die Fraunhofer-Gesellschaften Institut für Software- und Systemtechnik (ISST) und Institut für Angewandte Informationstechnik (FIT), den Lehrstuhl für Informations- und Automatisierungssysteme für die Prozess- und Werkstofftechnik (IAT) an der RWTH Aachen und das Deutsche Institut für Normung e.V. (DIN).

„Mit der Expertise aus dem Konsortium und der Vernetzung über die MISSION KI kann dieses Projekt erheblich zur Vernetzbarkeit unterschiedlicher Datenräume beitragen“, so Prof. Dr. Philipp Wieder von der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen.

„Die Herausforderung liegt darin, gemeinsame, flexible und sichere Standards zu etablieren, um die Vorteile vernetzter Datenräume nutzen zu können. FDOs bieten einen vielversprechenden Ansatz – eine einheitliche Methode zur Indizierung und Katalogisierung von Daten“, sagt Dr. Peyman Khodabakhsh, Projektleiter der MISSION KI.

Das Herzstück des Vorhabens ist die Entwicklung einer prototypischen Infrastruktur, die die Vereinheitlichung der Beschreibung von Daten über verschiedene Datenräume hinweg ermöglicht. Die Lösung basiert auf der Kombination des FDO Ansatzes mit dem Eclipse Dataspace Connector (EDC) – eine der führenden Datenraum-Technologien. Ein besonderer Fokus liegt auf der Demonstration der Kompatibilität alternativer Entwicklungen in der Datenraum-Architektur.

Als Grundlage für einen sicheren und effizienten Datenaustausch zwischen verschiedenen Organisationen, Branchen und Ländern werden Datenräume immer wichtiger. Sie fördern Innovationen, indem sie Unternehmen und Forschungseinrichtungen erlauben, gemeinsam an Daten zu arbeiten, Erkenntnisse zu teilen und auf dieser Basis neue Produkte, Dienstleistungen und Geschäftsmodelle zu entwickeln. Ein wesentliches Merkmal dieser Datenräume ist ihre verteilte, dezentralisierte Struktur. Die Vernetzung von Datenräumen wird zu einem kritischen Faktor für die wirtschaftliche Entwicklung, die Förderung der Wissenschaft und die Stärkung der digitalen Souveränität.

Generell weisen Datenräume eine große Heterogenität in Bezug auf technische und organisatorische Komponenten auf. Entscheidungen über Architektur und Standards können branchenbedingt stark variieren, mit unterschiedlichen Standards zur Beschreibung von Daten. Diese Vielfalt an Datenformaten und Standards führt beim Datenaustausch zwischen Teilnehmern aus verschiedenen Datenräumen zu Kompatibilitätsproblemen.

FAIR Digital Objects als universeller und unabhängiger Datenstandard

Fair Digital Objects (FDOs) sind eine Umsetzung der FAIR-Prinzipien gemäß des Data-Governance-Acts der EU. Sie zielen darauf ab, digitale Objekte auffindbar („Findable“), zugänglich („Accessible“), interoperabel („Interoperable“) und wiederverwendbar („Reusable“) zu gestalten. Indem sie Technologien für dauerhafte Verlinkung, umfangreiche Metadatenbeschreibungen und sichere Datenaustauschprotokolle nutzen, fördern FDOs ein transparentes, sicheres und nachhaltiges digitales Ökosystem.

[1] <https://mission-ki.de>

[2] <https://fairdo.org/fdof-summit-2024>

Otto



LLM Services and the AI Act of the EU

Text and Contact:

Mohammadtaher Azampour
mohammadtaher.azampour@gwdg.de

A Large Language Model (LLM) is a type of Artificial Intelligence (AI) with numerous capabilities that have found various applications in multiple domains. [1] The vast possible usage of this technology gave it a revolutionary characteristic in the world of technology and AI. [2] Considering the increasing usage of LLM in the AI world, this technology is here to stay and play its role as a critical foundation for further innovations. An important issue regarding this technology is related to legal challenges and, consequently, compliance with the EU legal system, particularly the AI Act (AIA). In continuation, we will talk about this technology and governing rules, then the details of the obligations under the AIA will be under discussion, and at the end, we will look at a practical case to check things out.

LLMS AND EU REGULATIONS

To have a clear view of the scope of the discussion, we should provide a definition for LLMs to see how they fit into the broader world of AI technology. However, regarding the related technological phrases, in most cases, definitions and categorizations are not acceptable to everyone, and most of them are used interchangeably. [3] By the way, we believe LLMs, or Large Language Models, belong to a group of foundational models. These foundational models, being extensive and versatile AI models, encompass any model trained on diverse data that can be customized for various subsequent tasks, serving as a basis for creating other applications. Therefore, LLMs, classified within this foundational model category, are engineered to comprehend and produce text akin to human capabilities. They undergo training on vast volumes of textual data, enabling them to generate natural language responses to input or queries. [4]

This technology has many uses in coding, advertising, product design, education, healthcare finance, law, and the public sector. However, these vast opportunities for usage are not free of risks. Researchers categorized the risks of using LLMs as an independent service or as a component of another model, system, or software. In practice, depending on the purpose and scope, using LLM services would face various difficulties, including cybersecurity concerns, the development of harmful cyber tools, potential failures in critical infrastructure, bias and discrimination, the absence of transparency, the emergence of hallucinatory content, and the need for robust data protection measures. [5]

The EU is trying to address these risks through proper rules that we can refer to under four categories:

Privacy

General Data Protection Regulation (GDPR) for the protection of personal data have much integration with LLM technology, like

the detection of personal data, data accuracy, and applying data subjects' rights in such a technology, which are exciting and challenging topics in this field.

Intellectual Property

Copyright-related directives and regulations regarding the use of protected materials in the training process, text and data mining, and the possibility of protecting the outputs of LLM services under this legal system are other challenges that the EU is trying to regulate appropriately (look at the directive (EU) 2019/790 of 17 April 2019).

Use Case Rules

Based on the utilization, regulations shall be under attention as well. For example, in the case of using LLMs in healthcare, industry-related regulations and obligations should be taken into account.

AIA

As we discussed at the beginning of this part, LLMs can be considered as AI, following the effort of the EU to regulate AI. One of the technologies that can be affected by this act would be LLM services, which we will discuss in more detail in the following pages.

LLMs und der AI Act der EU

In Anbetracht des zunehmenden Einsatzes von Large Language Models (LLM) in der Welt der Künstlichen Intelligenz (KI) wird diese Technologie auf Dauer Bestand haben und eine wichtige Grundlage für zukünftige Innovationen bilden. Ein wichtiges Thema im Zusammenhang mit dieser Technologie sind die rechtlichen Herausforderungen und folglich die Einhaltung des EU-Rechtssystems, insbesondere des KI-Gesetzes (AIA), die in diesem Artikel genauer beleuchtet werden.

AIA AND LLMs

Following the digital transition, European leaders discussed the importance of issues around AI and asked the commission to focus on it from various perspectives (2020). The process of enactment followed, and in the end, the council and parliament of the EU agreed on the AIA (2023), and the parliament finally adopted that (2024). This act determined some new obligations for stakeholders of AI systems or models depending on the level of risk that they can impose. This act as a risk approached legal document tried to regulate AI based on the purpose of usage as the first comprehensive AI regulatory framework in the world. [6]

In the AIA, we cannot find any definition or direct reference to AI Models, foundation models, or LLMs. However, there is an undeniable similarity between the definition of foundation models in literature and the given definition for General-Purpose AI (GPAI) in the AIA, which can be used interchangeably (Look at Art. 3 (63) of the AIA). As we discussed before, LLMs can be considered as a class of foundation models. As a result, we can consider LLM as a class of GPAI. Finally, we can say, from an AIA perspective, LLM, as a class of AI, is a kind of GPAI model that has a general usage and can be used independently as a service or through software or other AI models or systems.

In the initial proposals for this act, there was no article about GPAI. Following the appearance of GPT by OpenAI and legislators' awareness regarding the potential risks, the EU tried to regulate this type of AI at the EU level through the AIA. In this regard, at the later steps, they added GPAI-related articles. [7]

Whether we use LLMs independently as a service or in another AI model or system, stakeholders (providers, representatives, and others) have to comply with the GPAI-related articles (TITLE VIII) and obligations in the act. Documentation, transparency, and respect for copyrights, with some exceptions for open source GPAI, are the primary defined obligations for providers and authoritative representatives of providers in this act.

Considering all the risks we referred to regarding the LLMs, it does not mean that LLMs, independently or as a part of a more extensive system or model, are high-risk AI systems by themselves. Making decisions in this regard depends on the purpose of use, its impact on the people, and the cumulative amount of computing used for training (floating point operations (FLOPs) greater than 10^{25}). Considering these particular conditions, a GPAI may be categorized as a model with systemic risk in the AIA (Art. 52a). Providers of such AI models have more developed and rigid obligations under Art. 52d.

COMPLIANCE FRAMEWORK

A provider of LLM services based on the AIA is obliged to follow the obligations in Art. 52a-52e. The compliance framework in these articles includes some steps. The first step is to determine whether the model has systemic risk or not to define the related obligations. Following that, considering the results of the first step, the provider should take steps toward compliance with the related obligation based on the given articles. In the last step, in addition to minimum legal obligations, general tools like contractual relationships, privacy policies, and others should be under attention to increase transparency and trust.

In the AIA, as we referred to, GPAI models are classified into

the categories of usual models and models with systemic risks. Deciding on the category of a LLM service and defining it as a GPAI with systemic risk needs a comprehensive technical analysis to evaluate its impact capabilities based on technical elements (Art. 52a (1) (a), (2)). In brief, the determined threshold regarding the computing power is high enough to include very few models like GPT.4. [8] Regarding other factors, being public or private with limited access and intended use case (limited use cases like particularly research, teaching, and learning or general usage) are effective in defining the capability of the impact of a model and considering it as a GPAI with systemic risk.

Following that step, if a service uses fewer FLOPs, has limited access for users, and has limited use cases, it probably can not be considered as a model with systemic risks and is only obliged to follow the general obligations of GPAI models, not the ones with systemic risks. These obligations include up-to-date technical documentation of the model, availability of information and documentation for the potential providers intended to use this model in their AI systems, placing a policy to protect EU copyright law, and publicizing training content summary (Art. 52c). A GPAI with higher FLOPs than the determined threshold, public access, general use, and vast impact would be considered as a model with systemic risk, and in addition to Art. 52c, they are obliged to follow Art. 52d, which includes obligations like model evaluation, systemic risk mitigation, reporting incidents, and ensuring cybersecurity.

Analyzing each of these obligations is necessary through the compliance journey, but it is not part of our topic in this brief paper. Among all of these obligations, the only one that has more details in the AIA is documentation. Regarding the documentation, the minimum content has been provided in Annex IXa and Annex IXb. The service can be used independently, taking some inputs, processing them, and giving outputs. In this case, Annex IXa determined the content of the documentation. On the other hand, regarding the use case, in some situations, it can be used in other AI projects for sub-providers who intend to integrate the model into their AI system. In this case, Annex IXb must be followed.

In the end, it is important to consider the responsibility of LLM providers for legal obligations beyond those covered by the AIA. For instance, providers of such models have to consider personal data protection in data processing, from training data to processing personal data during service provision. To build trust, we need a high level of transparency and accountability. To create these characters, being obliged only to the minimum legal obligations is not enough. Having clear and well-developed contractual terms, policies, guidelines, and pieces of training is highly advisable in this regard.

APPLYING THE COMPLIANCE FRAMEWORK IN PRACTICE

In this section, we will discuss one example of an LLM service (GPT.4) to see how this framework can be applied. Using this example and whatever we will refer to in this part does not mean any confirmation regarding the GPT.4 legal policy. The other important point is that we should consider that providers of GPAI models still have time to prepare themselves for full compliance with the AIA. Given the time of entry into force and application of the AIA (Art. 85), GPAI-related articles will apply from 12 months following the entry into force of this regulation. As a result, we can

not expect GPAI model providers to comply with the given acts in all aspects at the current time.

GPT.4, considering the cumulative amount of computing power used for training, shall be presumed to be a GPAI model with higher impact capabilities and, as a result, systemic risk. Its public access, general usage, and popularity will increase such a possibility as well. However, the provider can argue against such a belief following the predetermined procedure under Art. 52b.

Considering the GPT.4 as a high-risk GPAI model, this provider shall follow Art. 52c and 52d. Regarding Art. 52c, we could not find a separate document about copyright policy, but they referred to this issue when discussing the training process, which mentioned that the data used for training includes public data and licensed data. For publicizing a detailed summary of training data, they have just a brief section for the training process, which includes a link to an article for a technical explanation of the training, but we could not find any more details about the data that is used for training.

Regarding the documentation, the provider tried to explain the details of the research, use cases, improvements, technical issues, contributions, usage guidance, and others in various documents and pages to develop transparency and trust. In this brief, we can not discuss the content of the documentation and analyze whether it is entirely in compliance with the AIA.

Regarding Art. 52d, the provided documents regarding the GPT. 4 include some references to the model evaluation, risk detection, and mitigation. Although the documents are not very detailed, they try to cover the topic briefly.

Regarding safety and cybersecurity, they made some efforts in different parts to cover these topics to show what they did in this field and how they improved compared to the last versions. [9]

Apart from the documents that are prepared for GPT.4 specifically, and apart from the contents that are directly related to the AIA, by looking at the other documents and policies of the provider of GPT.4, we can understand the company by preparing and providing various terms and policies (four categories of policies and eight categories of terms) [10] in one hand is trying to increase the level of trust and transparency and in the other hand is decreasing the possibility of responsibility and the burden of proof.

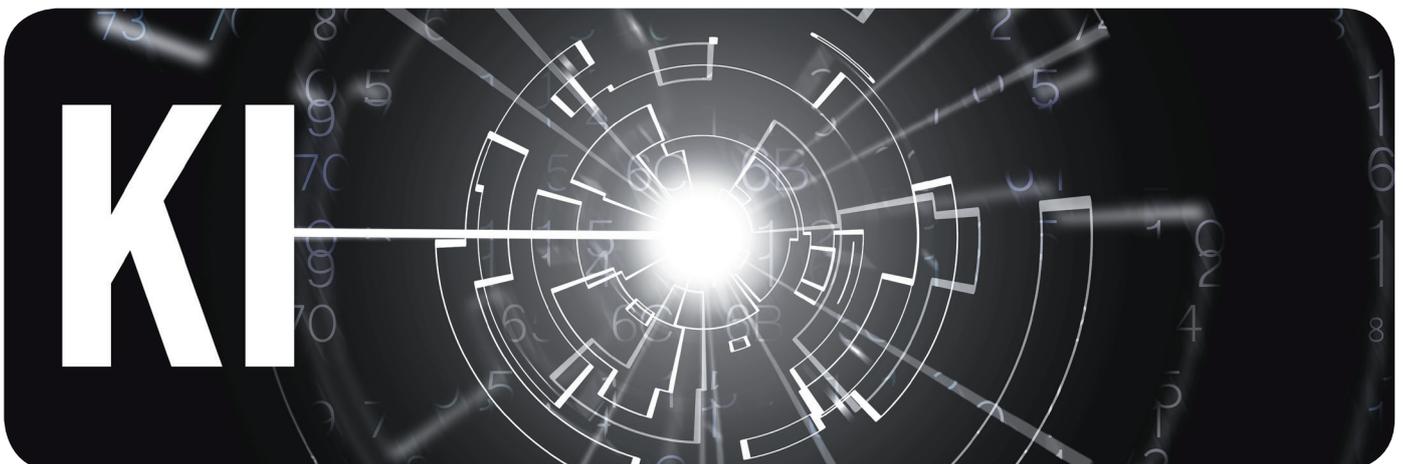
CONCLUSION

Developing LLMs as a service or as a part of a bigger picture is promising and necessary to be able to compete in the market,

but at the same time, it would lead to substantial legal obligations and responsibilities that the provider should comply with in the EU. Compliance involves understanding whether the AI being developed falls under the usual category or if it poses systemic risks. Once this is determined, providers need to prepare themselves for compliance with relevant obligations. Nowadays, all LLM service providers, including GWDC, are preparing to comply with these obligations through well-developed documentation, which the AIA significantly emphasizes. Additionally, preparing detailed policies and contractual tools is essential for mitigating legal risks and penalties. These tools can help outline responsibilities, establish protocols for addressing potential issues, and provide clarity on liability in case of any legal challenges. By taking proactive steps to comply with regulations and mitigate risks, service providers can operate within legal frameworks while fostering trust and accountability in their AI systems.

REFERENCES

- [1] Muhammad Usman Hadi , Qasem Al-Tashi , Rizwan Qureshi, et al.: A Survey on Large Language Models: Applications, Challenges, Limitations, and Practical Usage. TechRxiv. July 10, 2023. DOI: 10.36227/techrxiv.23589741.v1 , p. 5
- [2] Iqbal H. Sarker: LLM Potentiality and Awareness: A Position Paper from the Perspective of Trustworthy and Responsible AI Modeling. TechRxiv. February 27, 2024. DOI: 10.36227/techrxiv.170905626.67078570/v1, p. 5
- [3] <https://www.adalovelaceinstitute.org/evidence-review/foundation-models-public-sector/>
- [4] <https://www.ibm.com/topics/large-language-models>
- [5] <https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/54.pdf>
- [6] <https://www.europarl.europa.eu/topics/en/article/20230601/STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- [7] <https://www.stibbe.com/publications-and-insights/eu-artificial-intelligence-act-and-generative-ai-an-update>
- [8] EU AI Act Compliance Analysis: General-Purpose AI Models in Focus (<https://thefuturesociety.org/wp-content/uploads/2023/12/EU-AI-Act-Compliance-Analysis.pdf>)
- [9] <https://cdn.openai.com/papers/gpt-4-system-card.pdf>, p. 13
<https://openai.com/index/gpt-4-research>
- [10] <https://openai.com/de-DE/policies> ■



QA Catalogue – A Quality Assessment Tool for Library Catalogues

Text and Contact:
Ph.D. Péter Király
peter.kiraly@gwdg.de

QA catalogue is a quality assessment tool for library catalogues. It was created to support answering research questions about the quality of catalogues, and is still used in research and education activities of the GWDG. It analyses some aspects of the metadata records: whether they follow the rule of the standards, how completely they use the data elements available in the standards, what is the share of records having subject indexing and authority names, what vocabularies these terms came from, and what auxiliary properties are available that put the subject terms into context. Currently about a dozen libraries use this open source tool, among them the national libraries of the UK and Belgium. Several Computer Science students solved related scientific or technical questions during the history of the tool resulting with thesis or project reports. QA catalogue is used as “virtual research environment” for research papers published in German, English and Hungarian.

INTRODUCTION

A preliminary caution: In the IT field, the term “library” is ambiguous. It often refers to a software package, but in this article it refers to the organisation that collects and organises books and other materials, making them accessible to its patrons, the readers. The article is about a tool made for use by librarians and researchers who would need help in understanding the specificities of data stored in and created by these organisations.

How should a book be described properly in a library catalogue? This question has a long past (and even a future) with several proposed methods which evolved over time. In the current epoch in the history of cataloguing or in other words bibliographic control we see the end of a period, and the start of a new one. In the past decades library practitioners and researchers called attention to different drawbacks of Machine Readable Cataloguing, shortly MARC (see famous witty phrases such as “MARC must die”, or “Stockholm syndrome of MARC”). There is intensive research on the next generation of proper bibliographic metadata schemas and formats, but MARC and similar traditional metadata schemas are not only still with us, they also make the overwhelming majority of catalogue records.

When in the 1960s Henriette Avram, an engineer at the Library of Congress, and her colleagues invented MARC, the available information storage space was much less than it is nowadays, so the information should be compressed. The record itself contains abbreviated forms, while the manual of the standard (such as a code book) describes the abbreviated terms in detail.

This makes the human understanding of bibliographic records difficult in its native form, but makes the machine readability and thus

QA Catalogue – ein Instrument für die Qualitätsbewertung von Bibliothekskatalogen

Der QA catalogue ist ein Qualitätsbewertungsinstrument für Bibliothekskataloge. Es wurde entwickelt, um die Beantwortung von Forschungsfragen zur Qualität von Katalogen zu unterstützen, und wird immer noch bei Forschungs- und Bildungsaktivitäten der GWDG eingesetzt. Es analysiert einige Aspekte der Metadatensätze: Ob sie den Regeln der Standards entsprechen, wie vollständig sie die in den Standards verfügbaren Datenelemente verwenden, wie hoch der Anteil der Datensätze ist, die über Sach- und Autoritätsnamen verfügen, aus welchen Vokabularen diese Begriffe stammen und welche Hilfseigenschaften verfügbar sind, die die Sachbegriffe in einen Kontext stellen. Derzeit nutzt etwa ein Dutzend Bibliotheken dieses Open-Source-Tool, darunter die Nationalbibliotheken von Großbritannien und Belgien. Mehrere Informatikstudent*innen haben im Laufe der Geschichte des Tools verwandte wissenschaftliche oder technische Fragen gelöst, was zu Diplomarbeiten oder Projektberichten führte. Der QA catalogue dient als „virtuelle Forschungsumgebung“ für Forschungsarbeiten, die auf Deutsch, Englisch und Ungarisch veröffentlicht wurden.

validation easy. Theoretically at least. The problem is that during the decades while the basic structure of these schemas remained the same, they continued to grow into giant standards, with a number of dictionaries – small ones or large ones –, which sometimes are externally developed and maintained by other organisations, such as the content classification schemes. This made validation difficult, and as preliminary experiences show even the most famous library catalogues contain a large share of records having at least one invalid data element.

The purpose of the current research is to estimate the quality of the records by measuring structural features, and find those records which could be improved. With statistical analysis librarians and those who would like to work with or to understand data for different purposes can get an overview of the catalogue. This overview might be helpful as well in the transition to the new bibliographic schemas and formats. Knowing the data is always necessary, and it is independent of the format. The research is manifested as an open source software called QA catalogue [1]. The software also works as an educational tool: Computer Science and other students who are interested in cultural heritage data can and do participate in the development with their creative minds, and thus far a couple of BA and MA students made important achievements.

The ideal form of metadata assurance workflow follows a particular cycle:

1. Ingest/transfer data from the catalogue to the measuring tool
2. Measuring individual records
3. Aggregation and statistical analysis of the record-level metrics
4. Reporting
5. Evaluation with domain experts
6. Improving the records within the catalogue (if necessary)

QA catalogue covers steps 2-4. Compared with other open source MARC tools, the main characteristic of this one, is handling the semantics of the data. It contains a Java model of the MARC21 standard, so it keeps track the following information about data elements: name, URL of the definition, acceptable codes and their meaning, value constraints, indexing rules, corresponding functions defined in the Functional Requirements for Bibliographic Records (FRBR) model, historical codes, applicable dictionaries, BIBFRAME name (if there are) and other rules. It is implemented to enable extensions, so the ruleset covers not only MARC21, but several MARC variations, such as the Deutsche Nationalbibliothek's MARC version, OCLC MARC (only parts available in partner libraries' records), and locally defined fields in different libraries (Gent, Szeged, Finnish National Library), moreover it could handle other bibliographic metadata schemas such as PICA+ and UNIMARC. When the tool reads a record, it creates a triplet for each data element: location (field, subfield etc.), value, and the definition of the data element, thus it can validate all data against its definition.

The tool analyses a couple of different aspects of the catalogue. Their results (stored as CSV files, database tables and Solr index) are rendered by a lightweight web interface. The user interface is intended for the librarian, and it also contains a search interface for the catalogue, in order to connect the results of the analyses and the records.

ANALYSES

Validation

Validation iterates all parts of the records and checks if the data element and its value fits the content rules, e.g. it is a recognised structural element, the value is a proper key from the allowed dictionaries, the date encoding represents a real date, or the ISBN/ISSN numbers are valid. It groups the errors according to their location (record level, control field, subfield etc.). A common problem in almost every catalogue is that they use data elements which are locally defined but without publishing their documentation. In those cases, it is not possible to decide if the value of such an element follows the rules or not, thus the user interface provides two statistics: errors including these problems and clear errors (see Figure 1).

Completeness

Completeness checks how field groups, fields and subfields are distributed in the collection. Field groups are the large categories under which the MARC standard lists individual fields, such as Numbers and Code (fields 01X-09X), or Edition, Imprint (25X-28X). The resulting tables show how many records contain a particular data element (both in count and percentage), show the occurrences, its central tendencies (mean, standard deviation, minimum, maximum). The tool also enables browsing terms within any data element, and the user can check which records contain a particular term.

Weighted Completeness

The simple completeness takes the same importance for every field. In the literature, there are suggested solutions to add weights of fields to emphasise importance. Thompson and Traill [2] set up a scoring calculation for 19 data elements giving points by occurrences (in the range of 0 to theoretically unlimited). Their purpose was to select the best from MARC records vendors provided alongside ebooks. Carlstone [3] followed a similar strategy but she was interested in MARC records of electronic serials (here a record might have negative numbers if it misses particular data elements). Booth [4] surveyed which data elements are considered in UK libraries as the most important factors in making a record "shelf-ready". The tool implements these analyses, and provides visualisations for the score histograms (both for the total score and the individual factors).

Functional Analysis

The FRBR model defines 12 functions (such as search, identify, sort, display) which should be supported by bibliographic records. Delsey [5] initialised a map between these functions and the individual data elements. The analysis at this stage checks how much portions of the data elements belonging to a function are available in each record, then visualises the histograms of the support of functions. Since a usual record contains about 200 data elements, while the map handles more than 2,000 elements which support functions (MARC21 defines about 3,000 elements), even the "richest" records can not support more than 10-15% of the functions, the histograms display relative support. Its intended use is to compare the support of individual functions, so the librarian can conclude that a catalogue has more data supporting the function of selection than of searching.

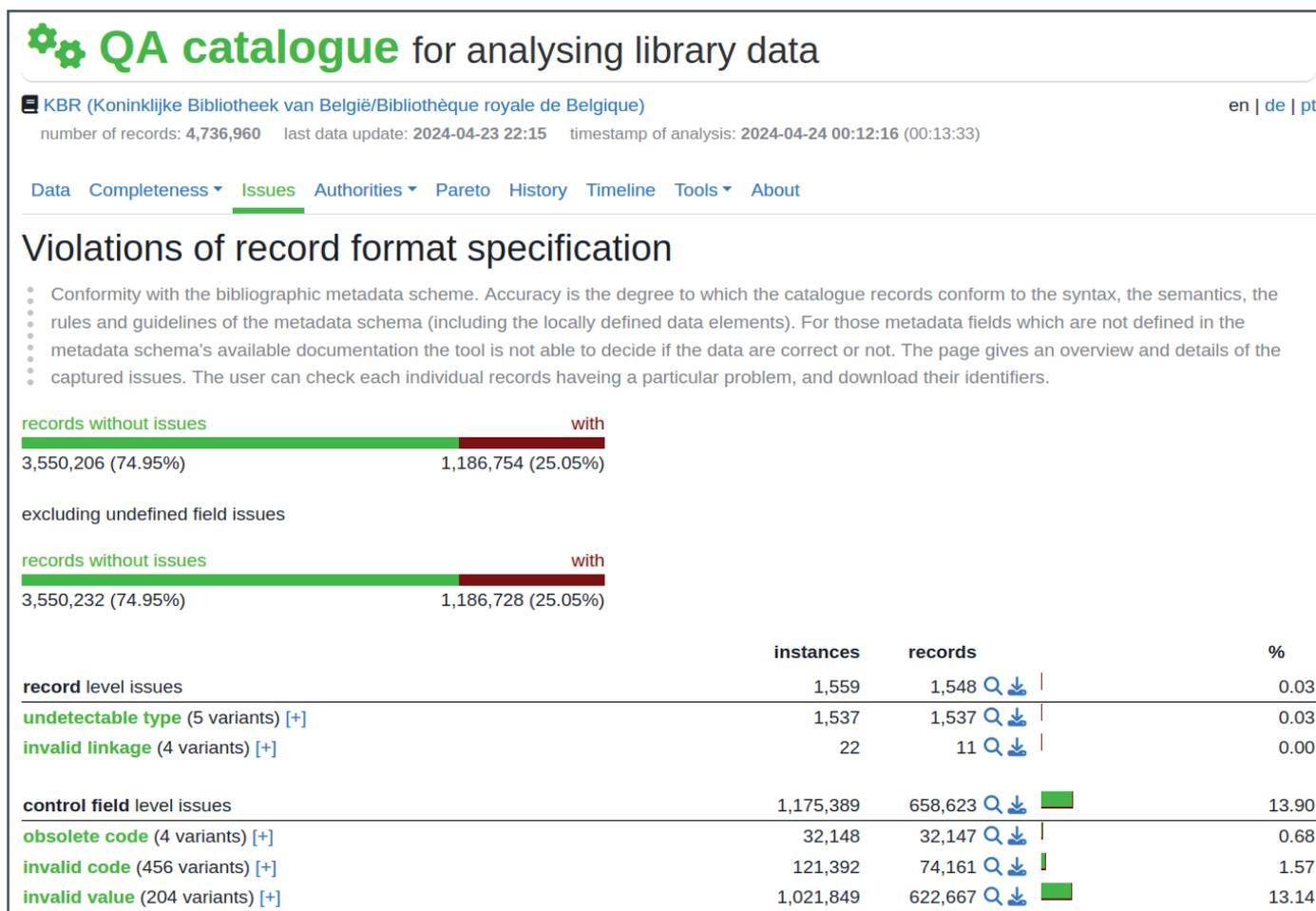


Figure 1: The validation result of the catalogue of the Belgian national library. 25% of the records contain some invalid values. The issues are groups into categories and subcategories, each is searchable. The librarian can download the list of record identifiers for future improvement.

Subject Indexing

In MARC, there are fields to support different methods of content description. The analysis checks all these fields and explores how they are used: how many records have at least one content description, how many terms are attached to the record, what kinds of classification schemes or Knowledge Organisation Systems (KOS) are used, what are the additional fields alongside the term (such as the source of the term, the version of the KOS etc.) The user interface provides links to the list of individual terms within a KOS, where the terms' links trigger searches.

Authority Name Analysis

Authority names are more similar than subjects as it provides normalised names for persons, meetings, corporates and even titles, which may or may not contain additional information, such as identifiers to vocabularies, dates and titles associated with the name, nature of contribution in the current record. This analysis follows the structure of the subject analysis (see Figure 2).

PICA AND K10PLUS CATALOGUE

In Germany most of the libraries work in networks when it comes to cataloguing. The largest of such networks is called Gemeinsamer Bibliotheksverbund (GBV). It maintains the union catalogue K10plus having 70+ million bibliographic records (the largest in Germany, one of the largest in the world). The records follow a metadata schema called PICA+ (used mainly in Germany,

France and The Netherlands). Its structure shares common features with MARC21, but the semantics are different. These schemas contain a fixed set of data elements, but libraries can define their own extra elements. The metadata schemas are usually available in documents created for human readers, but these documents, even if they are available in a digital format (e.g. in PDF or HTML) are not processable so that a software could use them as one of the inputs for the analysis. For this purpose a JSON representation of the MARC21 schema have been created, then Jakob Voß (Verbundzentrale des GBV) went further and created Avram schema, [6] which is a special JSON schema describing the properties that a bibliographic metadata schema might have in a machine processable way (named after Henriette Avram). Jakob Voß also maintains an Avram schema for PICA+.

To adapt QA catalogue to PICA+ we had to be able to read Avram schema and to use it as an input for model building against which the analyses can be run. For handling MARC21 record QA catalogue have Java classes to store the semantics of the data elements. QA catalogue could export MARC21 – including the local versions –, to Avram schema, so other tools can and do reuse it. For PICA and UNIMARC (see below) we used Avram files describing the details of the metadata schema. These Avram files are used during validation and other analyses. Due to the structural and semantic differences between PICA and MARC21 we had to modify the underlying business logic to make analyses “pluggable” to different schemas.

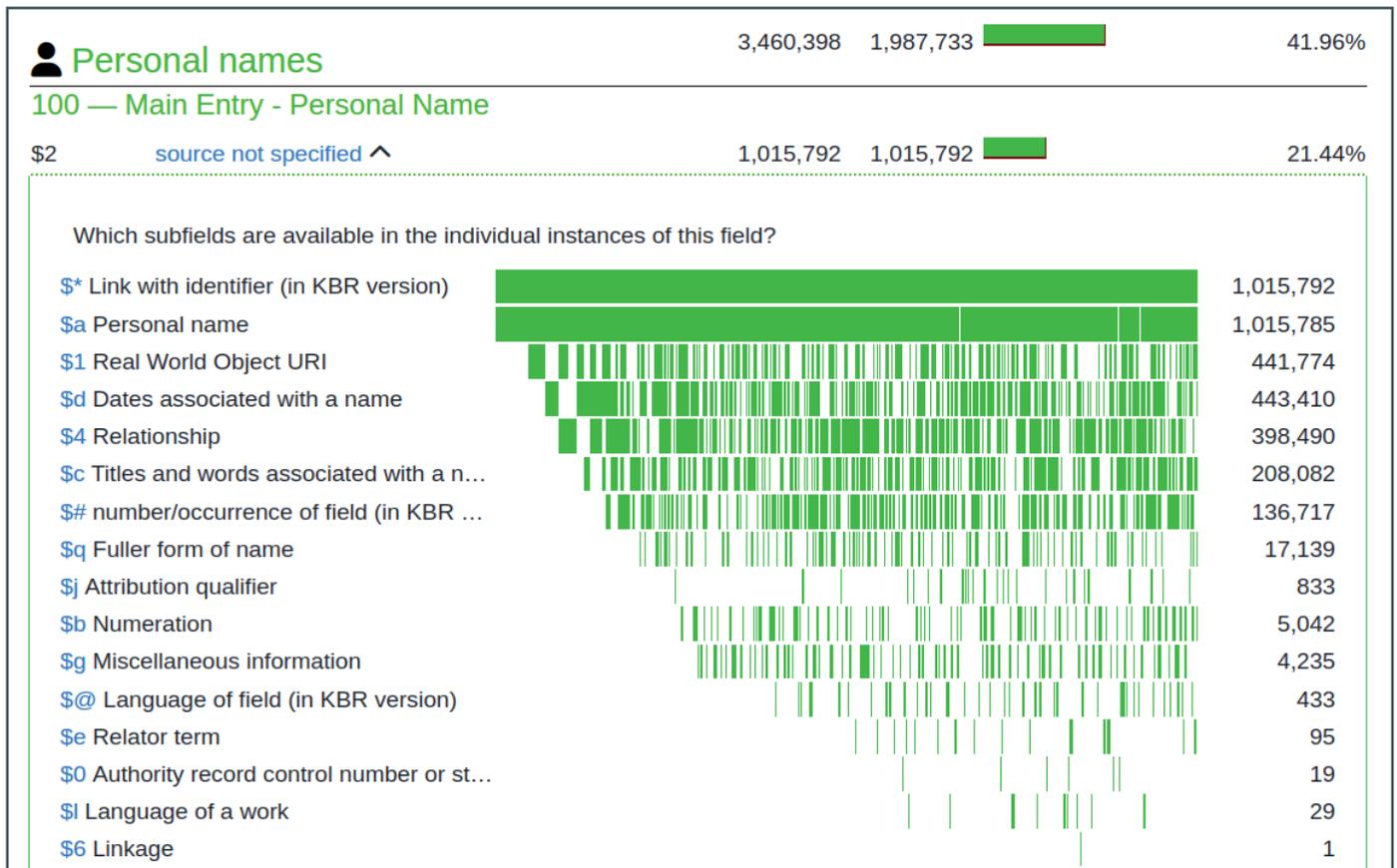


Figure 2: The authority name analysis reveals how much record has a particular data element (here: information about the author), what properties are used and in what combinations.

Regarding K10plus there are two extra challenges. One of them is the number of records and thus the size of output. In Göttingen at the GWDG it would have been possible to use extra hardware settings and big data analysis techniques, but we wanted to keep the tool available for institutions with commodity hardware, so we spent energy on optimising the algorithm and the external toolset instead of finding hardware based solutions. Another challenge comes from K10plus being a union catalogue of about a thousand libraries: the different quality related aspects (completeness, validity, subject indexing, name authority) should be calculated not only for the whole catalogue, but also for individual libraries.

This feature was co-developed by Jakob Voß, and we received important feedback from the catalogue’s community.

UNIMARC

Soon after MARC was invented national libraries started to develop their own customised versions. It is not surprising: MARC was the first machine readable bibliographic data standard, and before that there were no methods to uniform machine based cataloguing practices, thus libraries tried to implement their existing customs and traditions to MARC. A number of national versions emerged, such as USMARC, UKMARC, HUNMARC etc. From the late 70’s there were approaches to consolidate the situation by creating a „lingua franca” of MARC (a commonly used format to facilitate data exchange between libraries): such as the cooperation originally formed between Anglo Saxon countries, that later became MARC21, and UNIMARC, which was initiated by the International Federation of Library Associations and Institutions (IFLA).

Although MARC21 is used by the majority of the countries, UNIMARC is used in France, Portugal, Slovenia, Romania, Russia and elsewhere, and – similar to MARC21 – there are locally defined extensions in different libraries. MARC21 and UNIMARC share a common serialization format (having positional control fields, three letter long numeric encoded data fields with indicators and one letter long encoded subfields), but the semantics of the data elements are quite different.

The adaptation of QA catalogue to UNIMARC follows a similar approach than in the PICA use case. The first step of this process was to create the Avram schema out of the PDF manuals. Once it was done we needed to adapt the existing code to work with the structural and semantic differences between MARC21 and UNIMARC – involving mapping data elements in the two schemas that aim similar purposes (for example describe subjects or authorised format of names). The final task is the adaptation of the user interface. This work is in progress mainly developed by Computer Science student Haris Gegić as his master project. It will be published in the next release of the software.

SHACL4BIB

SHACL (Shape Constraints Language) is a domain specific language created under the umbrella of the World Wide Web Consortium for validating Resource Description Framework (RDF) statements. RDF follows the linked data and open world assumption paradigms, so in contrast to relational databases an object might have any kinds of properties, that sets aside the concept of “records”. It has lots of good consequences, but makes validation quite hard. SHACL (and a parallel approach Shape Expressions,

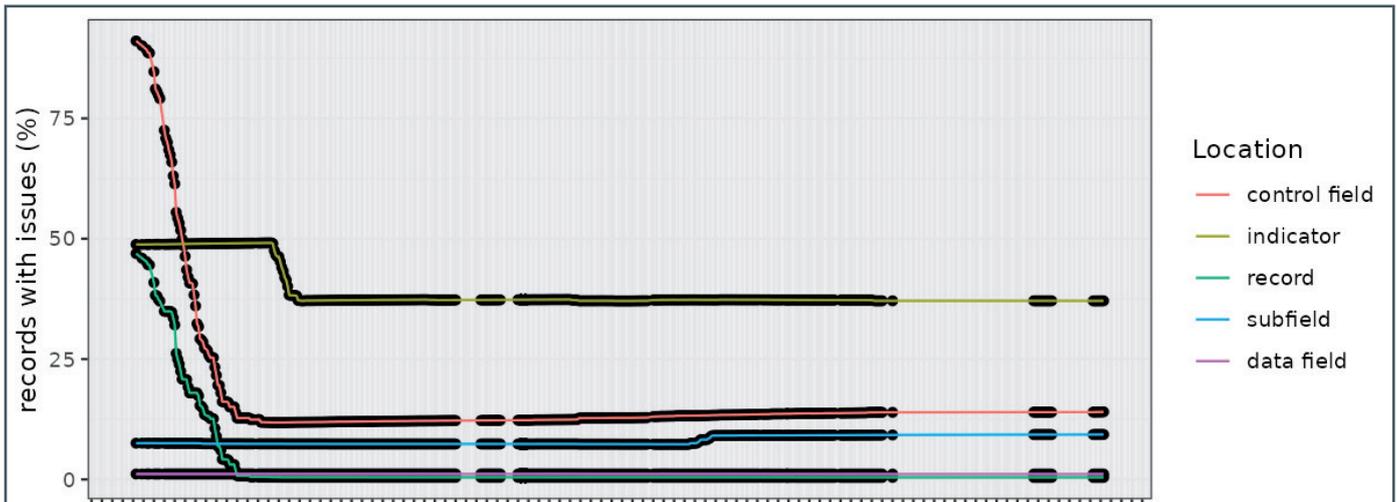


Figure 3: The effect of continuous quality assessment: changes of validation results during a 2 year period at KBR. It highlights the effect of data cleaning campaigns. After these campaigns the quality level of the catalogue is rather stable. For some types of errors fixing would not worth the efforts required, so they remain uncleared.

ShEx) tries to solve this situation by introducing an (also RDF based) language that enables data managers to set rules against RDF datasets. Some parts of this language are closely bound to the semantic web technologies, but there is an interesting part, a vocabulary of constraints that could be reused in other technological contexts. This vocabulary contains terms such as *minCount* (to set a minimal number of property instances), *maxLength* (the maximal length of a string value), *lessThan* (for comparing the values of two properties), or logical operators to combine such criteria. In SHACL these terms are used as predicates in statements where the subject is the entity we would like to validate, the object is either another entity or a scalar value. In order to use this vocabulary we just find a way to address these entities and extract their values. To achieve these purposes different ‘addressing languages’ are available, such as XPath for XML format, JsonPath for JSON, SQL selectors for tabular data. With the help of these languages, the Metadata Quality Assurance Framework (of which QA catalogue is an extension) provides methods to apply SHACL constraints to CSV, XML and JSON records.

In the bibliographic universe we have two relevant languages: MARCSpec for MARC based records (MARC21, UNIMARC), and PICA Path for PICA records. These languages specify how one can find a particular data element in the bibliographic record. During the development it turned out that for our quality assessment purposes we should extend SHACL vocabulary with some terms, such as “contentType” (to check the content type of a referenced resource, e.g. an image), “unique” (ensuring that the value, such as and identifier is unique in the dataset), or properties for scoring anti patterns or good practices. The user of the system could formulate the requirements as configuration files (in YAML or JSON format), or via the Java API. Listening 1 shows a simple ruleset.

It informs us that we apply this configuration to MARC records, thus we use MARCSpec as the addressing language. We check only one data element, the subfield *a* of field *040* (that contains the encoded form of the original cataloguing agency that created the record at first place), and we expect that the followings hold: 1) the element exists at least in one instance in each record, and 2) its value starts with the string *BE-KBR00* (the agency is a unit of the Royal Library of Belgium). The output of such analysis is a CSV file containing at least the record identifier and the result

```
format: MARC
fields:
  - name: 040$a
    path: 040$a
  rules:
    - id: 040$a.minCount
      minCount: 1
    - id: 040$a.pattern
      pattern: ^BE-KBR00
```

Listening 1: An example configuration inYAML format

of the validation checks. The result might have a status information (passed, failed, n/a), a score or both. Depending on the configuration it might contain additional information. The user interface picks the configuration and the analysis output and generates an HTML report. The implementation of this feature has been partly done by Computer Science student Jean Michel Nzi Mba as his Bachelor thesis [7].

A VISUAL DASHBOARD

One of the hardest things to handle in quality assessment is the abundance of information. The tool collects information about a lot of details – for the whole catalogue, individual features, individual data elements, subsets of the catalogue and the combinations of these. On the other hand we wanted a single visual dashboard that displays the most important quality features represented by half a dozen numbers and plots including the top level result of validation, completeness, subject and name authority analysis and one of the weighted completeness. This project has been done by Computer Science student David Nelles as his master project [8]. His project involved a survey about the prototype of the dashboard, which gave us some important feedback (although due to the low number of responses the result was not significant). The dashboard itself proved to be an important feedback about the reusability of the UI components and the output format of the analyses.

SOME TECHNICAL DETAILS

The tool is open source licensed under GPL v3.0. The backend of the tool is written in Java and R, the frontend is in PHP and JavaScript. For each version there are downloadable release packages. The Java backend is available as a library in Maven repositories, so others can use it in library data related software projects as a Java API. To make usage easier we maintain a Docker image that contains both the backend and frontend with all necessary components and packages. With it, the usage of the tool could be simplified to two commands at the command line on any operating system to have a working dashboard from zero. Some features of QA catalogue are available as web services – enabling the tool to embed into cataloguing workflows. We use tools such as GitHub Action pipelines, SonarCloud, Coveralls and others to monitor and improve code quality and consistency. More than a dozen developers from Germany, Hungary, Czech Republic, Norway, Switzerland, Finland, USA, Serbia and Brazil contributed to the software with code, documentation or translation. Many thanks for them, specially for the students who selected this project as their research topic.

FUTURE PLANS

QA catalogue is knowingly used by a dozen libraries (including national libraries such as the British Library, and KBR, the Royal Library of Belgium) as an element of continuous quality assessment (see Figure 3), but the author would like to increase this number. Each of them provides important feedback and ideas for further developments.

In a DFG supported project Aqinda (Agile Qualitätssicherung von Metadaten zu kulturellen Objekten im Kontext von Datenintegrationsprozessen), a joint research and development of the GWDG, Göttingen State and University Library and Marburg University, we are working on a user interface to make the creation of quality criteria easier.

There are lots of open research questions regarding catalogue quality assessment. To mention some of them: a) If the record refers to a hierarchical Knowledge Organisation System, whether it uses generic or specific terms (located close to or far away from the top-level terms of the hierarchy). b) If we see the catalogue as a graph in which the records represent the nodes and subjects represents edges which connects records together, what are the nature of this network? What conclusions could be drawn from these features in the domain of historical sociology of culture? c) Given the abundance of properties of a catalogue record, how can

we detect “grey zone” duplications where more metadata records describe the same library resource but there are differences in the metadata?

Based on the results of QA catalogue theoretically it would be possible to initialise a new GWDG service: continuous quality assessment for libraries, it is an open question if GWDG would like to enhance its current portfolio into this direction.

We are looking for additional cooperation with libraries, so if you are interested in this research, please contact the author.

NOTES AND REFERENCES

- [1] The tool’s source code is available in the following repositories: <https://github.com/pkiraly/qa-catalogue> (backend with a detailed user guide), <https://github.com/pkiraly/qa-catalogue-web> (web based user interface). A demo site with the catalogue of Universiteitsbibliotheek Gent (Belgium) is available at <http://gent.qa-catalogue.eu/>.
- [2] Kelly Thompson and Stacie Trill (2017): „Implementation of the scoring algorithm described in Leveraging Python to improve ebook metadata selection, ingest, and management“, Code4Lib Journal, Issue 38, 2017-10-18, <http://journal.code4lib.org/articles/12828>
- [3] Jamie Carlstone (2017): „Scoring the Quality of E-Serials MARC Records Using Java“, *Serials Review*, 43:3-4, pp. 271-277, DOI: 10.1080/00987913.2017.1350525
- [4] Emma Booth (2020): „Quality of Shelf-Ready Metadata. Analysis of survey responses and recommendations for suppliers“, Pontefract (UK): National Acquisitions Group, 2020. p 31, https://nag.org.uk/wp-content/uploads/2022/03/NAG-Quality-of-Shelf-Ready-Metadata-Survey-Analysis-and-Recommendations_2021Corrected.pdf
- [5] Tom Delsey (2002): „Functional analysis of the MARC 21 bibliographic and holdings formats“, Tech. report, Library of Congress, 2002. Prepared for the Network Development and MARC Standards Office Library of Congress. Second Revision: September 17, 2003, https://www.loc.gov/marc/marc-functional-analysis/original_source/analysis.pdf
- [6] <https://format.gbv.de/schema/avram/specification>
- [7] Jean Michel Nzi Mba (2023): „SHACL4MARC – Validating MARC records against locally defined ruleset“, Bachelor’s thesis, Georg-August-Universität Göttingen
- [8] David Nelles (2023): „Development of a Dashboard for a Metadata Quality Assurance framework“, Master project report, Georg-August-Universität Göttingen



INFORMATIONEN:
support@gwdg.de
0551 39-30000

Juni bis
Dezember 2024

Academy



| KURS | DOZENT*IN | TERMIN | ANMELDEN BIS | AE |
|--|-------------------|--|--------------|----|
| DEEP LEARNING BOOTCAMP: BUILDING AND DEPLOYING AI MODELS | Lewis | 03.06. – 04.06.2024 14:30 – 16:30 Uhr | 27.05.2024 | 3 |
| USING JUPYTER NOTEBOOKS ON HPC | Khuziyakhmetov | 04.06.2024 9:00 – 12:00 Uhr | 28.05.2024 | 2 |
| GRUNDLAGEN DER BILDBEARBEITUNG MIT AFFINITY PHOTO | Töpfer | 05.06. – 06.06.2024 9:30 – 16:00 Uhr | 29.05.2024 | 8 |
| DATA MANAGEMENT CONCEPTS FOR EFFICIENT AND USER-FRIENDLY HPC | Nolte | 06.06.2024 9:00 – 12:00 Uhr | 30.05.2024 | 2 |
| BIOINFORMATIC SERVICES AT THE GWGD | Paleico | 11.06.2024 9:00 – 12:00 und 13:00 – 16:00 Uhr | 04.06.2024 | 4 |
| WORKING WITH GRO.DATA | Dr. Király | 11.06.2024 10:00 – 11:30 Uhr | 04.06.2024 | 0 |
| HIGH PERFORMANCE DATA ANALYTICS – PART II | Dr. Ogaja, Meisel | 12.06. – 13.06.2024 9:30 – 16:00 Uhr | 05.06.2024 | 8 |
| USING THE GWGD SCIENTIFIC COMPUTE CLUSTER – AN INTRODUCTION | Eulert | 17.06.2024 9:30 – 16:00 Uhr | 10.06.2024 | 4 |
| RECHENZENTRUMSTECHNIK – KURS FÜR AUSZUBILDENDE | Schumann | 18.06. – 20.06.2024 9:00 – 16:00 Uhr | 11.06.2024 | 12 |
| ANSYS ON CLUSTER AND POST-PROCESSING OF SIMULATION RESULTS | Höhn, Dr. Kanning | 18.06.2024 9:00 – 12:00 und 13:00 – 16:00 Uhr | 11.06.2024 | 8 |
| QUICKSTARTING R: EINE ANWENDUNGSORIENTIERTE EINFÜHRUNG IN DAS STATISTIKPAKET R | Cordes | 19.06. – 20.06.2024 9:00 – 12:00 und 13:00 – 15:30 Uhr | 12.06.2024 | 8 |

| KURS | DOZENT*IN | TERMIN | ANMELDEN BIS | AE |
|--|-------------------------------------|--|--------------|----|
| GRUNDLAGEN DER PROGRAMMIERUNG – KURS FÜR AUSZUBILDENDE | Boers, von Rekowski | 01.07. – 03.07.2024 9:00 – 16:00 Uhr | 24.06.2024 | 12 |
| QUANTUM COMPUTING WITH SIMULATORS ON HPC | Dr. Boehme, Meisel | 02.07.2024 13:00 – 16:00 Uhr | 25.06.2024 | 2 |
| GETTING STARTED WITH LINUX BASH | Dr. Lüdemann, Eulert | 08.07.2024 9:00 – 12:00 Uhr | 01.07.2024 | 2 |
| HOW TO KISSKI | Dr. Lüdemann, Eulert | 08.07.2024 13:00 – 16:00 Uhr | 01.07.2024 | 2 |
| STATISTIK MIT R FÜR TEILNEHMER*INNEN MIT VORKENNTNISSEN – VON DER ANALYSE ZUM BERICHT | Cordes | 24.07. – 25.07.2024 9:00 – 12:00 und 13:00 – 15:30 Uhr | 17.07.2024 | 8 |
| POSTGRESQL – GRUNDKURS | Groh | 07.08.2024 9:30 – 16:00 Uhr | 31.07.2024 | 4 |
| POSTGRESQL FÜR FORTGESCHRITTENE | Groh | 08.08.2024 9:30 – 16:00 Uhr | 01.08.2024 | 4 |
| DEEP LEARNING WITH GPU CORES | Kirchner, Biniaz, Doost Hosseini | 22.08.2024 9:30 – 13:00 Uhr | 15.08.2024 | 2 |
| GRUNDLAGEN DER BILDBEARBEITUNG MIT PHOTOSHOP | Töpfer | 28.08. – 29.08.2024 9:30 – 16:00 Uhr | 21.08.2024 | 8 |
| INDESIGN GRUNDKURS – SCHWERPUNKT POSTERGESTALTUNG | Töpfer | 10.09. – 11.09.2024 9:30 – 16:00 Uhr | 03.09.2024 | 8 |
| INDESIGN – AUFBAUKURS | Töpfer | 25.09. – 26.09.2024 9:30 – 16:00 Uhr | 18.09.2024 | 8 |
| INTRODUCTION TO PERFORMANCE ENGINEERING | Dr. Ogaja | 26.09.2024 9:30 – 16:00 Uhr | 19.09.2024 | 4 |
| SUPERCOMPUTING FOR EVERY SCIENTIST | Eulert, Dr. Lüdemann | 02.10.2024 9:00 – 12:00 und 13:00 – 16:00 Uhr | 25.09.2024 | 4 |
| PRACTICAL: HIGH-PERFORMANCE COMPUTING SYSTEM ADMINISTRATION | Prof. Kunkel | 07.10. – 11.10.2024 9:00 – 18:00 Uhr | 30.09.2024 | 20 |
| VIRTUELLE CLOUD-INFRASTRUKTUREN – KURS FÜR AUSZUBILDENDE | Kopp | 08.10. – 10.10.2024 9:00 – 16:00 Uhr | 01.10.2024 | 12 |
| PERFORMANCE ANALYSIS OF AI AND HPC WORKLOADS | Dr. Ogaja, Dr. Lüdemann | 22.10.2024 9:00 – 12:00 und 13:00 – 17:00 Uhr | 15.10.2024 | 4 |
| STORAGE – KURS FÜR AUSZUBILDENDE | Quentin | 23.10. – 24.10.2024 9:00 – 12:30 und 13:30 – 15:30 Uhr | 16.10.2024 | 8 |
| INTRODUCTION TO ALPHAFOLD | Lux, Mühlhausen, Paleico | 29.10.2024 9:00 – 13:00 Uhr | 22.10.2024 | 2 |
| AFFINITY PUBLISHER – SCHNUPPERKURS FÜR EINSTEIGER*INNEN | Töpfer | 29.10.2024 10:30 – 12:30 und 13:30 – 15:30 Uhr | 22.10.2024 | 3 |

| KURS | DOZENT*IN | TERMIN | ANMELDEN BIS | AE |
|--|---------------------|--|--------------|----|
| DEEP LEARNING BOOTCAMP: BUILDING AND DEPLOYING AI MODELS | Lewis | 04.11. – 05.11.2024 14:30 – 16:30 Uhr | 28.10.2024 | 3 |
| ADMINISTRATION IM ACTIVE DIRECTORY | Quentin, Hast, Kopp | 06.11.2024 9:00 – 12:30 und 13:30 – 15:30 Uhr | 30.10.2024 | 4 |
| MONITORING WITH TIG STACK IN THE GWDG | Merz | 07.11.2024 9:00 – 10:00 Uhr | 31.10.2024 | 1 |
| PARALLEL PROGRAMMING WITH MPI | Prof. Haan | 12.11. – 13.11.2024 9:15 – 16:00 Uhr | 05.11.2024 | 8 |
| AFFINITY DESIGNER – SCHNUPPERKURS FÜR EINSTEIGER*INNEN | Töpfer | 12.11.2024 10:30 – 12:30 und 13:30 – 15:30 Uhr | 05.11.2024 | 3 |
| SECURE HPC – PARALLEL COMPUTING WITH HIGHEST SECURITY | Tabougua | 18.11.2024 10:00 – 11:30 Uhr | 11.11.2024 | 1 |
| GPU PROGRAMMING WITH CUDA – AN INTRODUCTION | Prof. Haan | 19.11.2024 9:15 – 16:00 Uhr | 12.11.2024 | 4 |
| EINFÜHRUNG IN DIE STATISTISCHE DATENANALYSE MIT SPSS | Cordes | 20.11. – 21.11.2024 9:00 – 12:00 und 13:00 – 15:30 Uhr | 13.11.2024 | 8 |
| GRUNDLAGEN DER BILDBEARBEITUNG MIT AFFINITY PHOTO | Töpfer | 26.11. – 27.11.2024 9:30 – 16:00 Uhr | 19.11.2024 | 8 |
| ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER*INNEN MIT VORKENNTNISSEN | Cordes | 04.12. – 05.12.2024 9:00 – 12:00 und 13:00 – 15:30 Uhr | 27.11.2024 | 8 |
| DEEP LEARNING BOOTCAMP: BUILDING AND DEPLOYING AI MODELS | Lewis | 09.12. – 10.12.2024 14:30 – 16:30 Uhr | 02.12.2024 | 3 |

Teilnehmerkreis

Das Angebot der GWDG Academy richtet sich an die Beschäftigten aller Einrichtungen der Universität Göttingen, der Max-Planck-Gesellschaft sowie aus wissenschaftlichen Einrichtungen, die zum erweiterten Kreis der Nutzer*innen der GWDG gehören. Studierende am Göttingen Campus zählen ebenfalls hierzu. Für manche Kurse werden spezielle Kenntnisse vorausgesetzt, die in den jeweiligen Kursbeschreibungen genannt werden.

Anmeldung

Für die Anmeldung zu einem Kurs müssen Sie sich zunächst mit Ihrem Benutzernamen und Passwort in der GWDG Academy (<https://academy.gwdg.de>) einloggen. Wenn Sie zum Kreis der berechtigten Nutzer*innen der GWDG gehören, erhalten Sie anschließend automatisch Zugang zu unserem Kursprogramm. Sollten Sie noch keinen Account besitzen, können Sie sich unter <https://id.academiccloud.de> registrieren und müssen ggf. auf Anfrage für die Anmeldung zu unseren Kursen freigeschaltet werden. Bei Online-Kursen kann das Anmeldeverfahren abweichen. Genauere Informationen dazu finden Sie in der jeweiligen Kursbeschreibung. Einige Online-Angebote stehen Ihnen jederzeit und ohne Anmeldung zur Verfügung.

Absage

Absagen können bis zu sieben Tagen vor Kursbeginn erfolgen. Bei kurzfristigeren Absagen werden allerdings die für den Kurs angesetzten Arbeitseinheiten (AE) vom AE-Kontingent der jeweiligen

Einrichtung abgezogen.

Kursorte

Die Kurse finden entweder in einem geeigneten Online-Format oder als Präsenzkurs statt. Nähere Informationen dazu finden Sie bei den jeweiligen Kursen. Auf Wunsch und bei ausreichendem Interesse führen wir auch Kurse vor Ort in einem Institut durch, sofern dort ein geeigneter Raum mit entsprechender Ausstattung zur Verfügung gestellt wird.

Kosten bzw. Gebühren

Die Academy-Kurse sind – wie die meisten anderen Leistungen der GWDG – in das interne Kosten- und Leistungsrechnungssystem der GWDG einbezogen. Die den Kursen zugrundeliegenden AE werden vom AE-Kontingent der jeweiligen Einrichtung abgezogen. Für alle Einrichtungen der Universität Göttingen und der Max-Planck-Gesellschaft sowie die meisten der wissenschaftlichen Einrichtungen, die zum erweiterten Kreis der Nutzer*innen der GWDG gehören, erfolgt keine Abrechnung in EUR. Dies gilt auch für die Studierenden am Göttingen Campus.

Kontakt und Information

Wenn Sie Fragen zum aktuellen Academy-Kursangebot, zur Kursplanung oder Wünsche nach weiteren Kursthemen haben, schicken Sie bitte eine E-Mail an support@gwdg.de. Falls bei einer ausreichend großen Gruppe Interesse besteht, könnten u. U. auch Kurse angeboten werden, die nicht im aktuellen Kursprogramm enthalten sind.



Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen