# Security Whitepaper

## plusmeta-Plattform 1.9+

Stand: 28.07.2020

*Als plusmeta-Plattform wird im Folgenden die Gesamtanwendung beschrieben, die sich aus verschiedenen Systemkomponenten zusammensetzt.*

## Überblick

Endanwender greifen über eine Webanwendung auf die plusmeta-Plattform zu. Diese kommuniziert im Hintergrund mit einem Applikationsserver und Cloud-Diensten zur Authentifizierung und Datenhaltung. Zur Sicherstellung der Verfügbarkeit werden Laufzeitfehler an einen Diagnoseserver übermittelt. Alle Daten werden verschlüsselt zwischen den Anwendungskomponenten über HTTPS (TLS 1.2) übertragen.

## Umgebungen

Die plusmeta-Plattform wird redundant in drei verschiedenen Umgebungen gehostet: Produktiv, Staging und Development. Die Produktiv- und Staging-Umgebungen können von Kunden verwendet werden, die Development-Umgebung wird intern zu Testzwecken verwendet. Durch den Einsatz von Orchestrierungsdiensten zur Verwaltung der Virtualisierungsumgebungen ist eine maximale Verfügbarkeit gewährleistet.

## Funktionsprinzip

Hauptanwendungszweck der plusmeta-Plattform ist die Datenanalyse zum Zweck der Metadatengenerierung. Dazu werden vom Benutzer hochgeladene Dateien einmalig analysiert und ausgewertet. Der für die Analyse notwendige Textinhalt wird in diesem Prozess aus den Dateien extrahiert und zusammen mit Dateimetadaten getrennt von der Quelldatei gespeichert. In den meisten Nutzungsszenarien wird die Quelldatei danach nicht mehr von der Anwendung benötigt und entsprechend nicht übertragen (Ausnahmen: Erweiterte Vorschau von PDF-Dateien oder expliziter Download der Quelldatei).

## Datenschutz

Benutzer stimmen bei der ersten Anmeldung am System den von plusmeta vorgegebenen Nutzungsbedingungen und Dataschutzbestimmungen zu. Diese sind unter der Adresse https://help.plusmeta.de/usage-agreement/ einsehbar.

Nur Enterprise-Kunden: Organisation können eigene Datenschutzhinweise oder Sicherheits-Richtlinien definieren und diese von Benutzern bestätigen lassen, bevor die Anwendung verwendet werden kann.

## Verschlüsselung

Um eine größtmögliche Sicherheit zu gewährleisten werden alle Daten sowohl bei der Übertragung als auch im Ruhezustand verschlüsselt. *Encryption-in-Transit* erfolgt dabei über eine Verschlüsselung mit TLS 1.2. *Encryption-at-Rest* erfolgt auf Dateispeicher- und Datenbankebene mit AES256.

## Datensicherung

Backups der Datenbank werden täglich (3 Uhr) automatisch ausgeführt, für 7 Tage aufbewahrt und danach automatisch gelöscht. Alle Backups der Datenbank sind verschlüsselt.

Alle Objekte, Eigenschaften und Metadaten können vom Benutzer jederzeit über die Oberfläche in offenen Exportformaten (JSON, CSV, ZIP) heruntergeladen werden.

## Mandantenfähigkeit

Pro Organisation (Mandant) wird ein eigener AWS Cognito-Identitätenpool angelegt und jeder Benutzer einem Identitätenpool zugewiesen (Cognito-Verbundidentität). Ein dort authentifizierter Benutzer hat Zugriffsrechte auf seine privaten Dateien, innerhalb des Pools geteilte Dateien und innerhalb des Pools veröffentlichte Dateien. Ein nicht authentifizierter Benutzer (sog. Öffentlicher Nutzer durch geteilten Projektlink) hat nur auf die innerhalb des Pools veröffentlichte Dateien Zugriffsrechte.

## Authentifizierung

Bei der ersten Anmeldung wird der Benutzer einmalig und direkt über die von AWS bereitgestellte Cognito-API authentifiziert. Bei allen folgenden Anfragen wird der Benutzer durch von Cognito ausgestellte JSON Web Tokens authentifiziert. Authentifizierungs-Tokens haben eine zeitlich beschränkte Gültigkeit (1h) und werden daher in regelmäßigen Abständen geprüft und (über Refresh-Tokens) erneuert. Die Authentifikation am Applikationsserver erfolgt über die gleichen Tokens (validiert über JSON Web Keys). Wir speichern keine Benutzerdaten in unserer Datenbank bzw. unserem Applikationsserver.

Nur Enterprise-Kunden: Alternative SSO-Authentifizierungsmöglichkeiten (SAML 2.0 oder OpenID Connect) können auf Anfrage über Cognito bereitgestellt werden. Eine Multi-Faktor-Authentifizierung (per SMS oder TOTP) ist pro Benutzer freischaltbar.

## Autorisierung

Innerhalb eines Mandanten können Organisationen ihr eigenes Rechtekonzept abbilden. Frei definierbaren Rollen können verschiedene Berechtigungen zugewiesen werden, die Aktions- und Sichtbarkeitsrechte innerhalb der Anwendung steuern. Die Sichtbarkeit von Objekten (Dateien) und Projekten kann vom Nutzer festgelegt oder von der Organisation vorgegeben werden.

## Verfügbarkeit

Die maximale Verfügbarkeit der Anwendung wird an Werktagen (Deutschland, Baden-Württemberg) zwischen 8:00 Uhr und 18:00 Uhr sichergestellt. Außerhalb dieser Zeit können geplante Wartungsarbeiten an der Anwendung stattfinden. Die Reaktionszeit des Supports beträgt während der Betriebszeiten maximal 24h.

Nur Enterprise-Kunden: Die garantierte Verfügbarkeit der Produktiv-Anwendung und die maximalen Reaktionszeiten des Supports werden über ein Service-Level-Agreement (SLA) vertraglich festgehalten. Der Kunde hat in diesem Fall Anspruch auf Erstattungen, wenn die Verfügbarkeit nicht eingehalten werden kann.

## Monitoring

Laufzeitfehler, ungewöhnliche API-Anfrage und fehlgeschlagene Anmeldeversuche werden überwacht und in Echtzeit an den plusmeta Support gemeldet. Damit kann sichergestellt werden, dass ein Angriffsversuch frühzeitig erkannt wird und bei Bedarf Gegenmaßnahmen eingeleitet werden können.

## Maßnahmen zur Vermeidung von Sicherheitslücken

- Automatischer Abgleich eingebundener Fremdsoftware mit Datenbanken bekannter Sicherheitslücken über den Dienst “GitHub Security Alerts”
- Regelmäßige manuelle Kontrolle und Aktualisierung verwendeter Fremdsoftware
- Regelmäßige Erneuerung von Sicherheitsschlüsseln und Zertifikaten
- Keine lokale IT-Infrastruktur zum Betrieb der Anwendung. Physischer Zugriff durch Mitarbeiter oder Eindringlinge nicht möglich.
- Interne Verwendung von Passwortmanagern und (wenn möglich) Zwei-Faktor-Authentifizierung für die Absicherung kritischer Infrastrukturzugänge
- Continuous Integration/Continuous Deployment: Geprüfte Aktualisierungen am Quellcode der Anwendung können innerhalb weniger Minuten vollautomatisch geprüft, ausgerollt und bereitgestellt werden.
- Agile Softwareentwicklung: Stabile freigegeben Versionen werden alle drei Wochen auf dem Produktivsystem bereitgestellt. Sicherheitskritische Aktualisierung sofort.
- Evergreen Updates: Alle Benutzer arbeiten immer automatisch mit der neusten freigegebenen Version der Software. Dadurch werden Sicherheitslücken durch veraltete Software ausgeschlossen.
- Keine Speicherung von Benutzerdaten auf unseren Servern. Diese werden ausschließlich innerhalb der Cognito-Identitätenpools verwaltet.
- Überwachung nicht-autorisierter Zugriffe zur frühzeitigen Erkennung von Angriffen.

## Endpunkte

### URLs zum Datenaustausch

*Produktiv:*

- https://app.plusmeta.de (Webanwendung)
- https://api.plusmeta.de (Applikationsserver)

*Staging:*

- https://dev.plusmeta.de (Webanwendung)
- https://api-dev.plusmeta.de (Applikationsserver)

*Diagnose:*

- https://diagnosis.plusmeta.de (Sentry)

*Authentifizierung*

- https://cognito-identity.eu-central-1.amazonaws.com
- https://cognito-idp.eu-central-1.amazonaws.com

### Öffentliche IP-Adressen

- 52.59.75.7 (Produktiv)
- 3.124.23.166 (Staging)
- 3.123.75.112 (Testing)
- 5.10.169.177 (Datenbank)
- 52.28.213.52 (Diagnose)

## Cloud-Infrastruktur

Die zum Betrieb notwendige Cloud-Infrastruktur wird vom Anbieter Amazon Web Services (AWS) in Frankfurt (Region: eu-central-1) betrieben. Für die Sicherheitsarchitektur innerhalb der AWS-Dienste existiert ein eigenes Whitepaper.[1] Die unmittelbar am Betrieb der Software beteiligten Systeme sind im Folgenden aufgeführt:

**AWS Cognito:**

- Authentifizierung, Autorisierung und Benutzerverwaltung
- Versendung von Transaktionsmails

**AWS ElasticBeanstalk:**

- Orchestrierung Applikationsserver (Virtualisierung und Lastverteilung)
- Plattform: Java 8 auf 64bit Amazon Linux 2.10

**AWS RDS**

- Bereitstellung und Backups der Datenbanken

**AWS Lightsail:**

- Virtualisierung Diagnoseserver
- Plattform: Java 8 auf 64bit Debian Linux

**AWS S3:**

- Speicherung von Quelldateien und Vorschaubildern
- Hosting der Web-Applikation

**AWS CloudFront:**

- Content Delivery Network. SSL-Endpunkt.

## Verwendete Software von Drittanbietern

Eine aktuelle Liste von eingebundener Fremdanbieter-Software und zugehöriger Lizenzen kann innerhalb der Webanwendung unter dem Menüpunkt „Lizenzen" abgerufen werden.

[1] https://d0.awsstatic.com/whitepapers/Security/AWS_Security_Whitepaper.pdf