

Zentrale Auswertung von Nutzungsdaten für FID-Lizenzen als Service des KfL

12.12.2023

Seit 2022 stellt das KfL den Service der Zentralen Auswertung von Nutzungsdaten für FID-Lizenzen bereit: Nutzungsreports aller laufenden FID-Lizenzen, die vom KfL oder durch einzelne FID selbst verhandelt wurden, werden zentral eingeholt, geprüft, gespeichert, ausgewertet und an die lizenznehmenden FID weitergeleitet. Nach der Aufbereitung und Analyse der Nutzungsdaten durch das KfL erhalten die FID in einer strukturierten und einheitlichen Darstellungsform die Auswertungen ihrer Lizenzangebote („FID-Reporting“). Der Bereitstellungstermin der Auswertungen ist in der Regel Anfang April und orientiert sich damit an den Fristen für die FID-Antragstellung bei der DFG. Auf Anfrage können für Antragsverfahren, die später im Jahr stattfinden, Zwischenauswertungen für die erste Jahreshälfte erstellt werden.

Fragestellung

Nutzungsdaten und andere produktspezifische Kennzahlen können unter anderem Aufschluss darüber geben, ob die Produktauswahl im Lizenzangebot eines FID die zugehörige Fachcommunity erreicht, dem Bedarf entspricht und angenommen wird („Servicequalität“). Zusätzlich können Nutzungsdaten in Relation zu den Lizenzkosten das Preis-Leistungsverhältnis einer Lizenz sichtbar machen („Cost-per-Use-Analyse“). Mit der Auswertung der eingeholten Nutzungsstatistiken soll daher die folgende Fragestellung beantwortet werden: Wie viel kostet die Nutzung einer Ressource im festgelegten Auswertungszeitraum pro Zugriff? Die erhobenen Kennzahlen wie auch die Auswertungsergebnisse sollen den FID als Arbeitsinstrument und verlässliche Entscheidungsgrundlage für ihr weiteres Vorgehen dienen.

Auswertungskriterien

Der Service der Zentralen Auswertung von Nutzungsdaten orientiert sich stark an den Erfordernissen des DFG-Datenblattes und den dort definierten und erfragten Kennzahlen.¹ Um eine Lizenz möglichst umfassend zu beschreiben, werden Werte wie Produkttyp, Titelanzahl, Lizenzmodell (Community, Campus, National)², die Größe des kalkulierten Nutzerkreises und die eventuelle vertragliche Vereinbarung dauerhaften Zugriffs für eine Lizenz ermittelt und angegeben. Die genannten Werte liefern nicht nur Informationen zu der jeweiligen Lizenz, sondern bieten auch Anhaltspunkte für die produktbezogene Kontextualisierung der Auswertungsergebnisse – so lässt sich über diese Angaben beispielsweise in Erfahrung bringen, wie umfassend eine Ressource ist und von wie vielen Personen oder Einrichtungen sie potentiell genutzt werden kann. Als zusätzliche Information für die FID ermittelt das KfL – sofern möglich – die Anzahl der in einem Kalenderjahr tatsächlich für die Nutzung einer Lizenz registrierten Einzelnutzer*innen bzw. Institutionen (im Jahresdurchschnitt bzw. mit Stand Dezember des auszuwertenden Kalenderjahres oder aber bis Ende des Lizenzzeitraums³).

¹ DFG (2023): 12.103 - Datenblatt Fachinformationsdienste für die Wissenschaft. Tabellenblatt C2: Lizenzprodukte im Detail. Online verfügbar unter https://www.dfg.de/formulare/12_103/12_103.xlsx

² Begriffliche Definition im Kontext der Auswertung: „Lizenzmodell“ und „Nutzerkreismodell“ werden synonym verwendet.

³ Begriffliche Definition im Kontext der Auswertung: „Lizenzzeitraum“, „Vertragszeitraum“ und „vertragliche Laufzeit“ werden synonym verwendet und stehen für die vertraglich vereinbarte Dauer der Zugriffsrechte auf eine Lizenz. Ausdrücklich nicht gemeint ist in diesem Kontext die Zeitspanne eines „laufenden“ Vertrags, in der bei einer Lizenz mit dauerhaften Zugriffsrechten beispielsweise noch inhaltliche Updates geliefert werden.

Für die Cost-per-Use-Analyse werden darüber hinaus die folgenden Kennzahlen herangezogen:

- Gesamtlaufzeit einer Lizenz (in Monaten)
- Gesamtbruttokosten
- Zeitraum, in dem die Zugriffe gezählt wurden (in Monaten)
- Zahl der Zugriffe (unter ergänzender Angabe des Erhebungsstandards und der Metrik des vorliegenden Nutzungsreports)

Gegenstand der Auswertung

Nutzungsreports werden angefordert für alle Lizenzen, die in dem für die Auswertung festgelegten Auswertungszeitraum (entspricht standardmäßig dem vergangenen Kalenderjahr) aktiv waren und die durch das KfL oder durch einzelne FID selbst verhandelt wurden. Für nicht mehr existierende FID werden keine Nutzungsstatistiken eingeholt. Dauerhaft erworbene Lizenzen (Kauf) werden auch dauerhaft ausgewertet. Sollte bei Subskriptionen der Fall eintreten, dass der auszuwertende Zeitraum mehrere Vertragslaufzeiten umfasst (wie etwa bei Verlängerungsverträgen, die unterjährig starten), so betrachtet die Auswertung zunächst jede relevante Vertragslaufzeit inklusive der dazugehörigen Kennzahlen separat. Die Berechnung des Cost-per-Use-Betrages erfolgt sowohl jeweilig für die unterschiedlichen Lizenzzeiträume als auch übergreifend für den festgelegten Auswertungszeitraum unter Einbeziehung der entsprechenden anteiligen Gesamtbruttokosten. Dies bedeutet, dass abgelaufene Subskriptionen nur solange ausgewertet werden, wie sich der für die Auswertung festgelegte Zeitraum mit der vertraglichen Laufzeit der Subskription überschneidet. Für Open-Access-Lizenzen werden bislang keine Statistiken eingeholt.

Auswertungszeitraum

Die Auswertung bezieht sich auf das jeweils vergangene Kalenderjahr und errechnet den durchschnittlichen Cost-per-Use-Betrag für Zugriffe auf eine Lizenz in diesem Zeitraum. Die Auswertung nach Kalenderjahr trägt der jährlichen Berichterstattung Rechnung und kann beispielsweise dazu dienen, festzustellen, ob die Nutzung einer Ressource im Verlauf mehrerer Jahre steigt oder sinkt.

Ergänzend werden die Lizenzen nach vertraglicher Laufzeit ausgewertet. Diese Auswertung bezieht sich auf die gesamte bisherige Laufzeit des im vergangenen Kalenderjahr aktiven Vertrages oder gegebenenfalls der im vergangenen Kalenderjahr aktiven Verträge für eine Lizenz. Die Auswertung nach vertraglicher Laufzeit dient dazu, den Cost-per-Use-Betrag für die bisherige Gesamtnutzung einer Lizenz zu errechnen und berücksichtigt damit die Tatsache, dass die jährliche Zugriffszahl auf eine Ressource starken Schwankungen unterliegen kann. Bei Lizenzen mit dauerhaften Zugriffsrechten kann durch die Auswertung der gesamten bisherigen Laufzeit gut veranschaulicht werden, dass die Kosten in Relation zur Nutzung von Jahr zu Jahr sinken.

In einigen Fällen ist es nicht möglich, die Nutzung einer Lizenz seit Beginn des Vertragszeitraums abzubilden, da kein durchgehend einheitlicher Erhebungsstandard bei der Zählung der Nutzung vorliegt, was zu einer inkonsistenten und damit nicht vergleichbaren Datenlage führt. In diesen Fällen werden die Daten des aktuellen Standards rückwirkend so weit wie möglich eingeholt und der Cost-per-Use-Betrag für den Zeitraum errechnet, für den Reports in einem einheitlichen Standard zur Verfügung stehen.

Andere Auswertungszeiträume können auf Anfrage gerne berücksichtigt werden.

Einholung und Aufbereitung von Nutzungsreports

Nutzungsstatistiken für das vergangene Jahr werden von den meisten Anbietern frühestens ab Ende Januar zur Verfügung gestellt. Dies bedeutet, dass das Zeitfenster zur Einholung und Auswertung der Nutzungsdaten eng begrenzt ist. FID mit anstehenden Antragsverfahren bei der DFG werden aus diesem Grund bei der Bereitstellung der Auswertungsergebnisse vorangestellt.

Wege, auf denen die Reports eingeholt werden, sind unter anderem: E-Mail-Anfragen bei Anbietern, Download aus den Admin-Portalen der Anbieter, Abruf via COUNTER_SUSHI API (Standardized Usage Statistics Harvesting Initiative)⁴, Weiterleitung der Nutzungsreports von Nationallizenzen durch die zuständigen Nationallizenzen-Teams, nur in Einzelfällen Erstellung der KfL-HAN-Proxy-Statistik durch die SUB Göttingen.

Bei durch die FID selbst verhandelten Lizenzen werden die FID darum gebeten, dem KfL die Nutzungsdaten einmal jährlich zur Verfügung zu stellen oder aber das KfL formlos (z. B. über E-Mail) zu mandantieren, die Statistiken im Auftrag einholen zu dürfen.

Die anbieterseitige Bereitstellung der Nutzungsreports erfolgt in diversen Dateiformaten wie zum Beispiel XLSX, CSV/TSV, HTML, PDF oder auch Grafiken. Dies zieht ebenso wie der Umstand, dass viele der Nutzungsreports nachträglich nach den auszuwertenden Lizenzen, ggf. sogar Einzeltiteln und/oder zugriffsberechtigten Institutionen gefiltert werden müssen, zum Teil sehr aufwendige Aufbereitungsschritte nach sich. Erst diese Schritte ermöglichen es, Nutzungsreports zur Verfügung zu stellen, die mit wenigen Ausnahmen einheitlich als Excel-Dateien formatiert sowie produktspezifisch vorgefiltert sind.

Erhebungsstandards von Nutzungsdaten

Die heterogene Zusammensetzung des Datenmaterials beschränkt sich nicht allein auf die gelieferten Dateiformate. Auch die Erhebungsmethoden von Nutzungszahlen variieren erheblich bis hin zu Statistiken, deren Methoden bei der Erfassung der Daten nicht nachvollziehbar sind und bei denen trotz KfL-seitiger Nachfrage bei den Anbietern unklar bleibt, wie genau Zugriffe definiert und gezählt werden.

Nur standardisierte Erhebungsmethoden führen zu aussagekräftigen, glaubwürdigen und konsistenten Statistiken. Daher wird grundsätzlich bei der Einholung von Nutzungsreports durch das KfL angestrebt, Statistiken im aktuellsten verfügbaren COUNTER-Standard - auch als „Release“ bezeichnet - zu erhalten. COUNTER (Counting Online Usage of Networked Electronic Resources) bezeichnet eine der ersten Normungsorganisationen für moderne Informationsumgebungen, die in einer Kollaboration mit Verlegern und Bibliotheken einen Standard, den sogenannten Code of Practice, für die Zählung der Nutzung elektronischer Ressourcen entwickelt hat und diesen fortlaufend aktualisiert.⁵ Jedoch ist selbst bei Anbietern, die COUNTER-Reports liefern, der aktuellste Release nicht immer verfügbar, sodass sich auch hier Unterschiede und - im Falle einer Umstellung auf den neuesten Release - ggf. auch inkompatible Zählweisen zwischen den Versionen ergeben. Dies ist unvermeidbar, da die laufende Weiterentwicklung und Anpassung des Standards an die Bedarfe der COUNTER-Community Veränderungen bei der Art und Weise der Zugriffszählung mit sich bringt.⁶

Der Erhebungsstandard, nach dem die Zugriffe auf eine Lizenz erfasst werden, und die Metrik, die die Zugriffsart definiert, spielen eine nicht unerhebliche Rolle beim Verständnis und der Einordnung der Daten. Auch diese Angaben finden sich daher in der Auswertung.

⁴ National Information Standards Organization (2023): Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol (ANSI/NISO Z39.93-2014). Online verfügbar unter <https://www.niso.org/standards-committees/sushi>

⁵ vgl. Mellins-Cohen, Tasha (2021): Release 5.0.2. The friendly guide for librarians. S. 3f. Online verfügbar unter <https://media-library.projectcounter.org/file/The-Friendly-Guide-for-Librarians>

⁶ vgl. dieser Abschnitt COUNTER (2023): COUNTER Code of Practice Release 5.1. Introduction. Online verfügbar unter <https://cop5.projectcounter.org/en/5.1/01-introduction/index.html>

Qualität der Daten

Eine gewisse Fehlertoleranz und Näherungswerte (wie etwa bei tageskursabhängigen Umrechnungen der Gesamtbruttopreise) müssen den in der Auswertung angegebenen Daten zugestanden werden. Obwohl weitgehend verlässlich, können in Einzelfällen auch anbieterseitig fehlerhafte oder unvollständige Nutzungszahlen geliefert werden - beispielsweise aufgrund technischer Schwierigkeiten. Die eingeholten Nutzungsreports werden durch das KfL im Hinblick auf Kriterien wie Daten für die richtige Lizenz, zum Produkttyp passende Metrik, Anzahl der teilnehmenden Institutionen (bei Campus- oder Nationallizenzen), korrekter Erhebungszeitraum, Erfassung für alle Zugriffswege etc. geprüft und ggf. reklamiert. Nach Durchlaufen unserer Prüfroutinen gehen wir zunächst grundsätzlich davon aus, dass die gelieferten Daten korrekt sind. Falls Sie einzelne Zahlen für nicht plausibel halten sollten, so fragen Sie bitte nach.

Auch bei allen anderen Rückfragen oder bei weiterem Informationsbedarf steht das KfL gerne als Ansprechpartner zur Verfügung.

Claudia Christophersen, SUB Göttingen