Zalando SE Data Catalogue for Vetted Researchers (Article 40 DSA)

Introduction and Scope

In compliance with Article 40(4) of the EU Digital Services Act (DSA), Zalando SE is committed to providing vetted researchers with access to data that is necessary for studying systemic risks in the European Union.

This Data Catalogue lists data categories, data structures, and metadata that are potentially available for research purposes, focusing on the systemic risks identified in our annual risk assessment, conducted in accordance with Article 34 of the DSA and subject to regulatory oversight.

In line with the DSA, this inventory focuses primarily on data relevant to the systemic risks identified in the Zalando SE DSA Risk Report. We do not provide an exhaustive catalogue of all data held by the company.

Key Documents and Resources

For context on the systemic risks and our platform governance, please refer to the following documents, which form the foundation of this Data Catalogue:

- **DSA Risk Report:** Zalando Transparency Hub.
- **DSA Transparency Hub:** For information on content moderation and governance, please consult the Zalando Transparency Hub.

The Request and Access Process

Vetted researchers must follow the official procedure set out by the Digital Services Coordinator (DSC) of the Member State of establishment (Germany), or by the European Commission in the DSA Data Access Portal (linked here).

Zalando SE will only grant access to data upon receiving a Reasoned Request from the relevant DSC, confirming that the researcher has been vetted and the request is necessary and proportionate to the study of systemic risks in the EU.

Protection of Trade Secrets and Confidentiality as well as Privacy

In accordance with Article 40(5) of the DSA, Zalando SE reserves the right to request an amendment to a Reasoned Request where access to data would lead to significant vulnerabilities in the security of our service or the protection of confidential information, in particular trade secrets, or personal data of customers, users and/or employees. This includes, but is not limited to, proprietary algorithmic logic, detailed commercial pricing strategies, and granular internal performance metrics that, if disclosed, would cause competitive harm. In such

cases, we commit to proposing alternative means of access, such as providing aggregated, synthetic, or privacy-preserving differentially private data.

Access Modalities

The final access modality is determined by the DSC after assessment of the request, taking into account data security, confidentiality, and the protection of personal data. Available access modalities may include:

1. **Secure Processing Environments (SPE):** Providing access to the data within a secure, remote computing environment managed by Zalando SE, which prevents direct data download.
2. **Secure Data Transfer:** Providing the data via secure transfer mechanisms (e.g., SFTP or cloud transfer) after appropriate pseudonymization or aggregation, as required by the nature of the data and the research scope.

Data Catalogue by Systemic Risk

The data assets listed below are directly relevant to the three systemic risks identified in our annual risk assessment.

Systemic Risk	Relevant Data Categories	Data Asset Buckets (Specific Data Products)	Description
1. Dissemination of Illegal Content	Content Moderation and Governance	**Public Transparency Report**	Aggregated and historical metrics on content moderation decisions, take-down volumes, and governance processes. **Source:** Transparency Hub.
		Automated Content Moderation / Content Moderation	Data on algorithmic systems used for content pre-publication filtering (e.g., product listing checks) as well as post publication Notice & Action proceedings). **Source:** Internal Moderation /Anti-Abuse Systems.

	Product Data (Prices, Quantities, Characteristics)	**Curated Product Information on Partner Level**	Full offer data (product, price, stock, availability status, per merchant) for third-party partners. **Source:** Offering/Inventory Data Domains.
		Curated Product Information	Curated product-related information on Config and Simple SKU granularity (attributes, characteristics). **Source:** Product Data Domain.
	Merchant Data	**Curated Information on Merchants**	Curated and aggregated information on onboarded merchants (partners), including history of updates. **Source:** Partner Data Domain (Merchant Profile Service).
2. Negative Effects on Fundamental Rights	User Data (Exposure & Engagement)	**Customer Onsite Behaviour Data**	Pseudonymized, event-level data on key customer activities: PDP views, Cart/Wishlist operations, and Sales orders placed. **Source:** Sales/Onsite Activity Data Domains.
		Customer Onsite Behaviour Data	Frontend onsite data providing product card view events. **Source:** Onsite Activity Data Domain.

	Ad Targeting and Profiling (Pricing & Metadata)	**Events on Ad Targeting**	Transaction-level tracking events (clicks, views, impressions) for ZMS ad campaigns, including metadata and pricing data (Cost per click data). **Source:** Marketing Data Domain (ZMS Ad Platform).
	Testing of New Features	**A/B Test Results Input Data**	Experiment-level metadata (descriptions, hypotheses) and anonymized input metrics used in experiments. Note: Raw results, proprietary methodologies, or internal performance forecasts will be protected as trade secrets under Article 40(5) of the DSA. **Source:** Experimentation Platforms.
	Account Deletion	**Deletion Events**	Event tracking customer requests for account deletion through the Privacy Portal. **Source:** Customer Data Domain.
3. Negative Effects on Physical and Mental Well-being	Content Moderation and Governance	**Content Production & Moderation Policies**	Documentation on policies, community guidelines, and content steering data that relate to well-being risks. **Source:** Internal Documentation/Community Guidelines.

	Interaction Data / Customer Care	**Customer Care Events**	Unique row per Customer Care (CuCa) case/shipment combination. **Source:** Customer Care Data Domain.
	Content Recommendations (Inputs)	**Interaction Events**	This dataset provides the core interaction data (PDP views, Cart/Wishlist actions) that fuels recommender systems. **Source:** Sales/Onsite Activity Data Domains.

Contact Information

For all inquiries related to this DSA Article 40 Data Catalogue, or to follow up on a pending request from a Digital Services Coordinator, please contact us: **authorities-dsa@zalando.de**

***

Zalando SE Datenkatalog für zugelassene Forschende (Artikel 40 DSA)

Einleitung und Geltungsbereich

Gemäß Artikel 40 Absatz 4 des Gesetzes über digitale Dienste (DSA) verpflichtet sich Zalando SE, zugelassenen Forschenden Zugang zu Daten zu gewähren, die für die Untersuchung systemischer Risiken in der Europäischen Union notwendig sind.

Dieser Datenkatalog listet Datenkategorien, Datenstrukturen und Metadaten auf, die potenziell für Forschungszwecke zur Verfügung stehen, wobei der Schwerpunkt auf den systemischen Risiken liegt, die in unserer jährlichen Risikobewertung, die gemäß Artikel 34 des DSA durchgeführt wird und der behördlichen Aufsicht unterliegt, identifiziert wurden.

In Übereinstimmung mit dem DSA konzentriert sich dieses Inventar primär auf Daten, die für die systemischen Risiken relevant sind, die in der DSA-Risikobewertung der Zalando SE identifiziert wurden. Wir stellen keinen erschöpfenden Katalog aller vom Unternehmen gespeicherten Daten bereit.

Wichtige Dokumente und Ressourcen

Für den Kontext zu den systemischen Risiken und unserer Plattform-Governance verweisen wir auf die folgenden Dokumente, die die Grundlage dieses Datenkatalogs bilden:

- **DSA-Risikobewertung**: Zalando Transparency Hub.
- **DSA Transparency Hub**: Informationen zur Inhaltsmoderation und Governance finden Sie im Zalando Transparency Hub.

Datenzugangsantrag und Datenzugangsprozess

Zugelassene Forschende müssen das offizielle Verfahren befolgen, das vom Koordinator für digitale Dienste (DSC) des Mitgliedstaats der Niederlassung (Deutschland), oder der Europäischen Kommission im DSA-Datenzugangsportal (hier verlinkt) festgelegt wurde.

Zalando SE wird Zugang zu Daten nur dann gewähren, wenn der zuständige DSC ein begründetes Verlangen übermittelt, das die Zulassung des Forschers sowie die Notwendigkeit und Verhältnismäßigkeit der Anforderung für die Untersuchung systemischer Risiken in der EU bestätigt.

Schutz von Geschäftsgeheimnissen und Vertraulichkeit sowie Datenschutz

In Übereinstimmung mit Artikel 40 Absatz 5 des DSA behält sich Zalando SE das Recht vor, ein Änderungsersuchen bezüglich eines begründeten Verlangens zu stellen, wenn der Zugang zu Daten zu erheblichen Sicherheitslücken in unserem Dienst oder dem Schutz vertraulicher Informationen, insbesondere Geschäftsgeheimnissen, oder personenbezogener Daten von Kunden, Nutzern und/oder Mitarbeitern führen würde. Dies umfasst, ist aber nicht beschränkt auf proprietäre algorithmische Logik, detaillierte kommerzielle Preisstrategien und granulare interne Leistungskennzahlen, deren Offenlegung einen Wettbewerbsnachteil verursachen würde. In solchen Fällen verpflichten wir uns, alternative Zugangsmöglichkeiten vorzuschlagen, wie die Bereitstellung von aggregierten, synthetischen oder datenschutzfreundlichen Daten, wie z.B. durch Differential Privacy.

Zugangsmodalitäten

Die endgültige Zugangsmodalität wird vom DSC nach Bewertung der Anforderung festgelegt, wobei Datensicherheit, Vertraulichkeit und der Schutz personenbezogener Daten berücksichtigt werden. Verfügbare Zugangsmodalitäten können umfassen:

- **Sichere Verarbeitungsumgebungen (SPE)**: Bereitstellung des Datenzugangs innerhalb einer sicheren, von Zalando SE verwalteten Rechenumgebung, die das direkte Herunterladen von Daten verhindert.
- **Sicherer Datentransfer**: Bereitstellung der Daten über sichere Transfermechanismen (z. B. SFTP oder Cloud-Transfer) nach entsprechender Pseudonymisierung oder Aggregation, wie es die Art der Daten und der Forschungsumfang erfordern.

Datenkatalog nach Systemischem Risiko

Die unten aufgeführten Datenbestände sind unmittelbar relevant für die drei systemischen Risiken, die in unserer jährlichen Risikobewertung identifiziert wurden.

Systemisches Risiko	**Relevante Datenkategorien**	**Spezifische Datenprodukte**	**Beschreibung**
1. Verbreitung Illegaler Inhalte	Inhaltsmoderation und Governance	**Öffentlicher Transparenzbericht**	Aggregierte und historische Kennzahlen zu Entscheidungen über Inhaltsmoderation, Take-Down-Volumen und Governance-Prozessen. **Quelle**: Transparency Hub.
		Automatisierte Inhaltsmoderation / Inhaltsmoderation	Daten über algorithmische Systeme, die zur Filterung von Inhalten vor der Veröffentlichung (z.B. Prüfungen von Produktlistings) sowie über Mitteilungs- und Abhilfeverfahren nach der Veröffentlichung verwendet werden. **Quelle**: Interne Moderations-/Anti-Abuse-Systeme.
	Produktdaten (Preise, Mengen, Merkmale)	**Kuratierte Produktinformationen auf Partnerebene**	Vollständige Angebotsdaten (Produkt, Preis, Bestand, Verfügbarkeitsstatus, pro Händler) für Drittpartner. **Quelle:** Offering/Inventory Data Domains.
		Kuratierte Produktinformationen	Kuratierte produktbezogene Informationen auf Config- und Simple-SKU-Granularität (Attribute, Merkmale). **Quelle**: Product Data Domain.
	Händlerdaten	**Kuratierte Informationen über Händler**	Kuratierte und aggregierte Informationen über eingebundene Händler (Partner), einschließlich des Aktualisierungsverlaufs.

			Quelle: Partner Data Domain (Merchant Profile Service).
2. Negative Auswirkungen auf Grundrechte	Nutzerdaten (Exposition & Engagement)	**Onsite-Verhaltensdaten von Kunden**	Pseudonymisierte, ereignisbasierte Daten zu wichtigen Kundenaktivitäten: PDP-Ansichten, Warenkorb-/Wunschlisten-Vorgänge und aufgegebene Bestellungen. **Quelle**: Sales/Onsite Activity Data Domains.
		Onsite-Verhaltensdaten von Kunden	Frontend-Onsite-Daten, die Ereignisse zur Produktkartenansicht (*Product Card View*) bereitstellen. **Quelle**: Onsite Activity Data Domain.
	Ad-Targeting und Profiling (Preise & Metadaten)	**Ereignisdaten zum Ad-Targeting**	Transaktionsbezogene Tracking-Ereignisse (Klicks, Ansichten, Impressionen) für ZMS-Anzeigenkampagnen, einschließlich Metadaten und Preisdaten (Kosten pro Klick-Daten). **Quelle**: Marketing Data Domain (ZMS Ad Platform).
	Testen neuer Funktionen	**Input-Daten für A/B-Testergebnisse**	Experiment-Metadaten (Beschreibungen, Hypothesen) und anonymisierte Input-Kenndaten, die in Experimenten verwendet werden. Hinweis: Rohdaten, proprietäre Methodologien oder interne Leistungsprognosen werden als Geschäftsgeheimnisse gemäß Artikel 40 Absatz 5 des DSA geschützt. **Quelle**: Experimentation Platforms.

	Kontolöschung	**Ereignisdaten zur Datenlöschung**	Ereignis-Tracking von Kundenanfragen zur Kontolöschung über das Privacy Portal. **Quelle**: Customer Data Domain.
3. Negative Auswirkungen auf das körperliche und geistige Wohlbefinden	Inhaltsmoderation und Governance	**Regeln zur Inhaltserstellung und -kontrolle**	Dokumentation von Richtlinien, Gemeinschaftsstandards und Content-Steuerungsdaten, die sich auf Risiken hinsichtlich des Wohlbefindens beziehen. **Quelle**: Interne Dokumentation/Gemeinschaftsstandards.
	Interaktionsdaten / Kundenservice	**Ereignisdaten zum Kundenservice**	Jeder Datensatz repräsentiert eine einmalige Kombination aus Kundenservice-Fall und zugehöriger Sendung. **Quelle**: Customer Care Data Domain.
	Inhaltsempfehlungen (Inputs)	**Interaktionsdaten**	Dieser Datensatz liefert die wesentlichen Interaktionsdaten (PDP-Ansichten, Warenkorb-/Wunschlisten-Aktionen), die Empfehlungssysteme speisen. **Quelle**: Sales/Onsite Activity Data Domains.

Kontaktinformationen

Für sämtliche Anfragen im Zusammenhang mit diesem Datenkatalog nach Artikel 40 DSA oder zur Statusabfrage einer ausstehenden Anfrage eines Koordinators für digitale Dienste wenden Sie sich bitte an uns: **authorities-dsa@zalando.de**