



Accessing platform data under the EU Digital Services Act

Opportunities for machine learning research

João Vinagre, Joana Gonçalves de Sá, Paula Fortuna, Sophia Dietrich, Paulo Almeida

19 September 2025

ECML PKDD 2025, Porto, Portugal

Outline

- I. The Digital Services Act
- II. Data Access
- III. Access to publicly available data
- IV. Access for vetted researchers
- V. Interactive session: research questions, systemic risks, data requests and safeguards

The Digital Services Act

What is the Digital Services Act (DSA)?

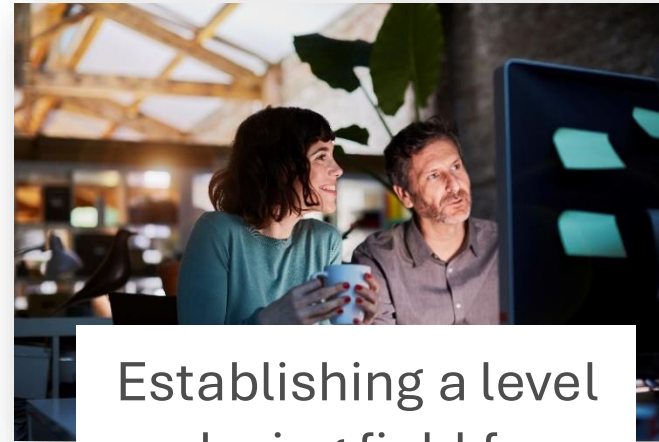
EU Regulation: common set of rules in all member states.



Creates a **safer digital space** and mitigates legal fragmentation.

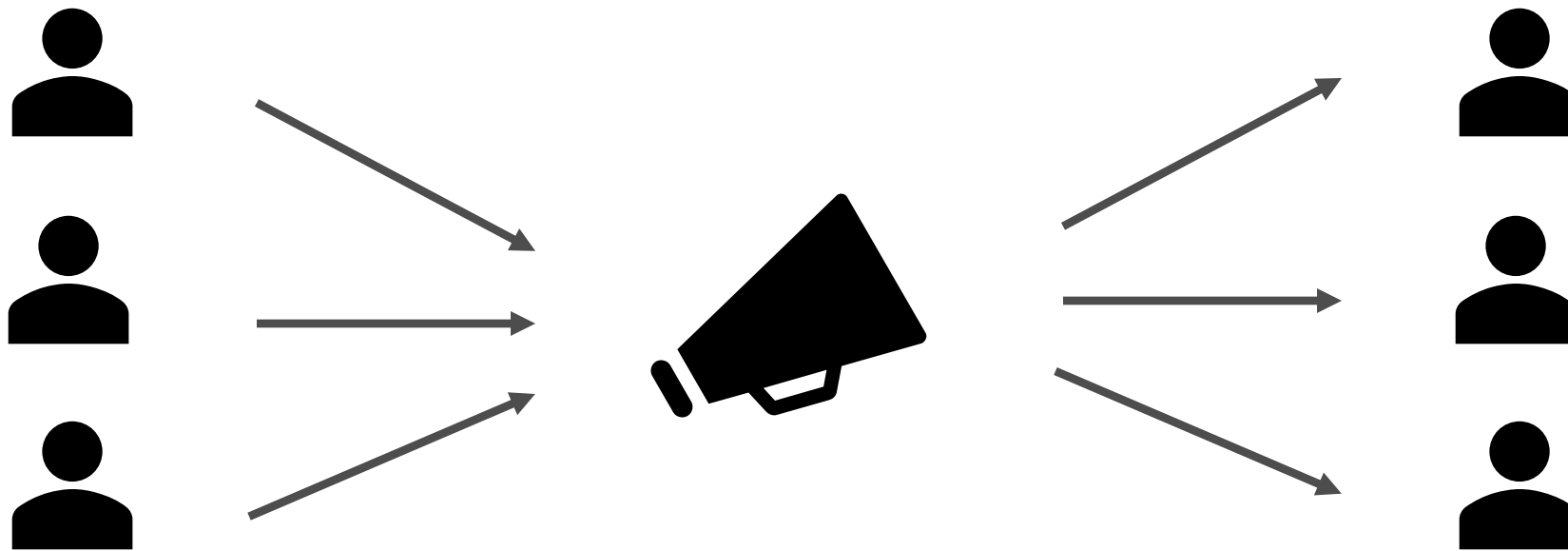


Protecting the
Fundamental Rights
of users



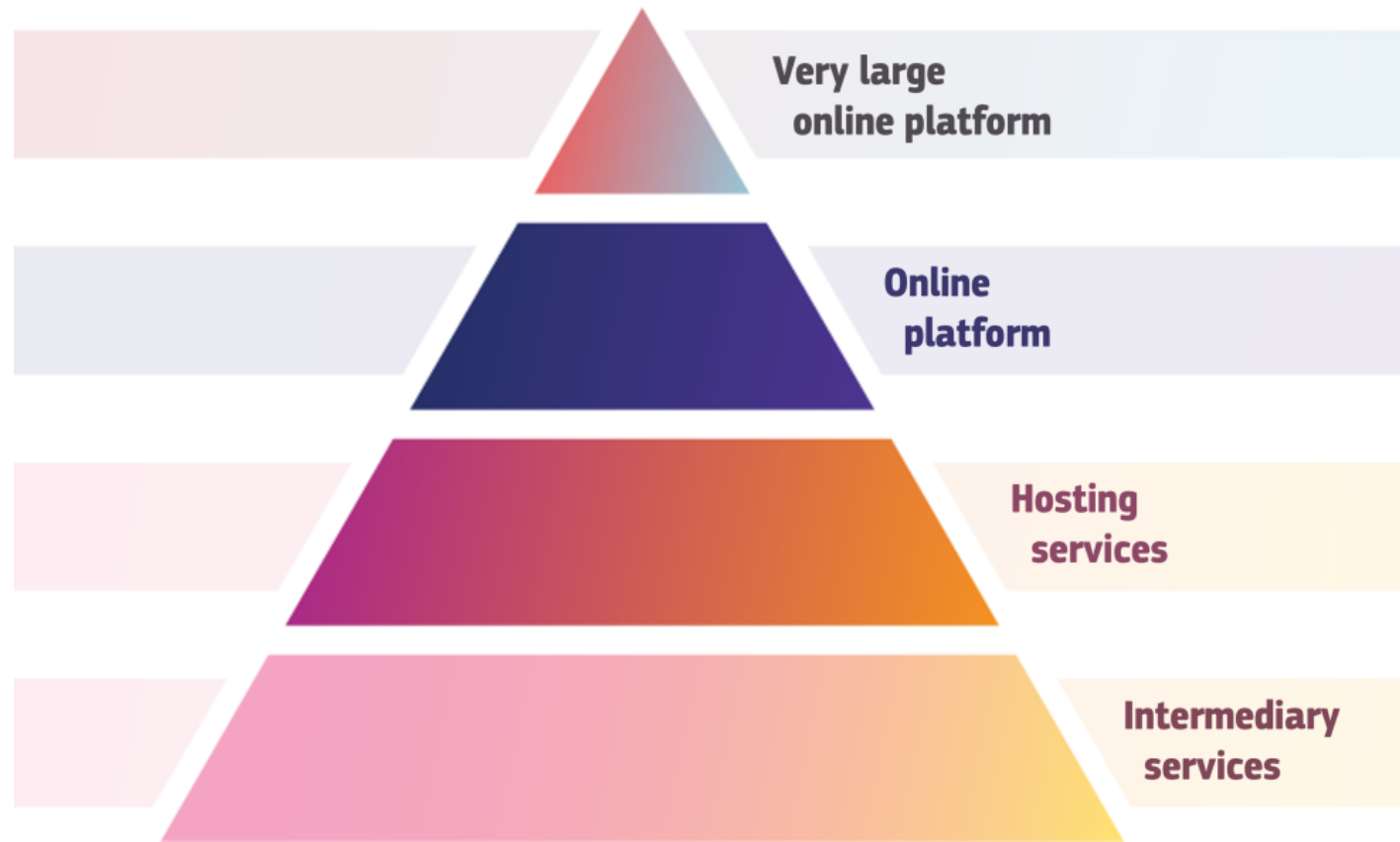
Establishing a level
playing field for
businesses

Digital Services



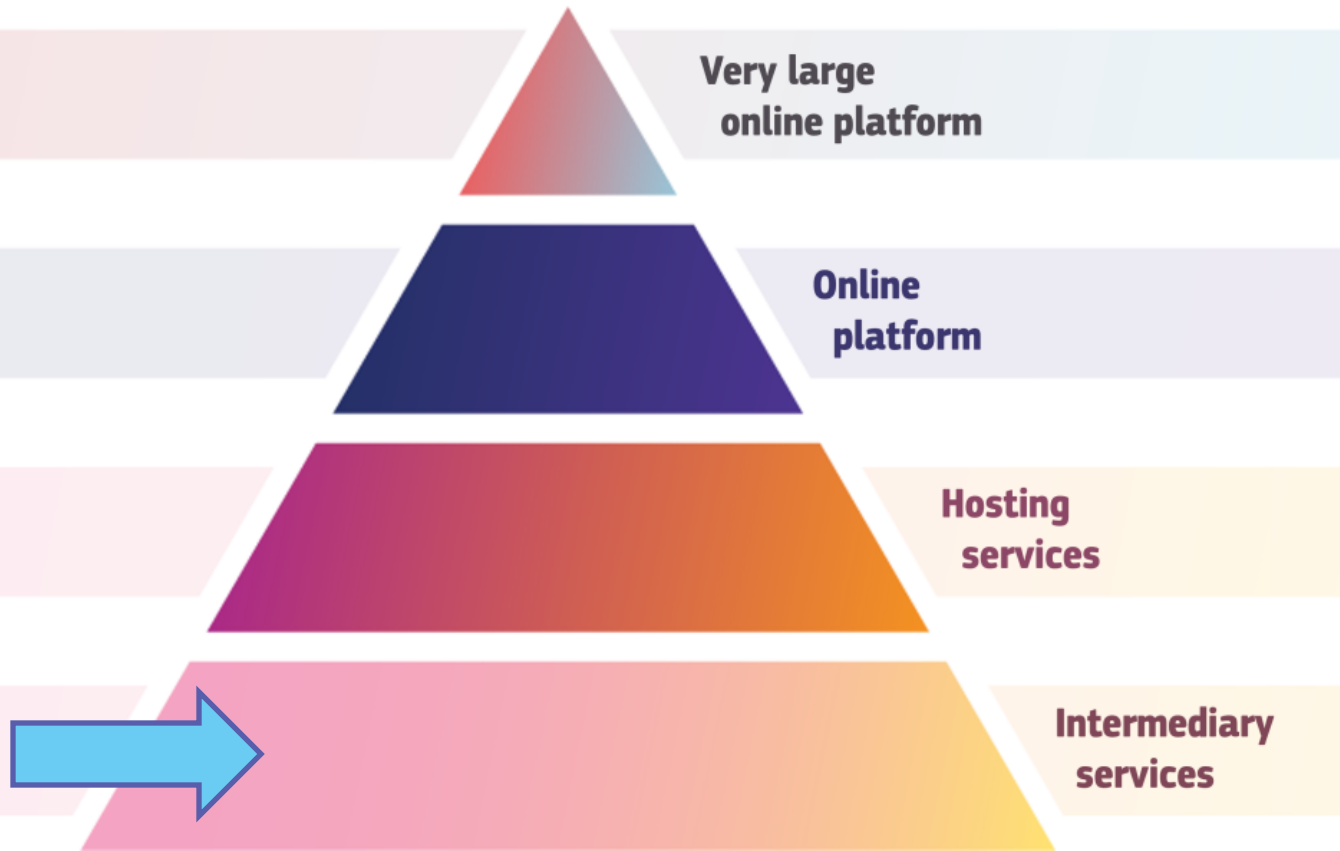
Intermediary Services

DSA layers



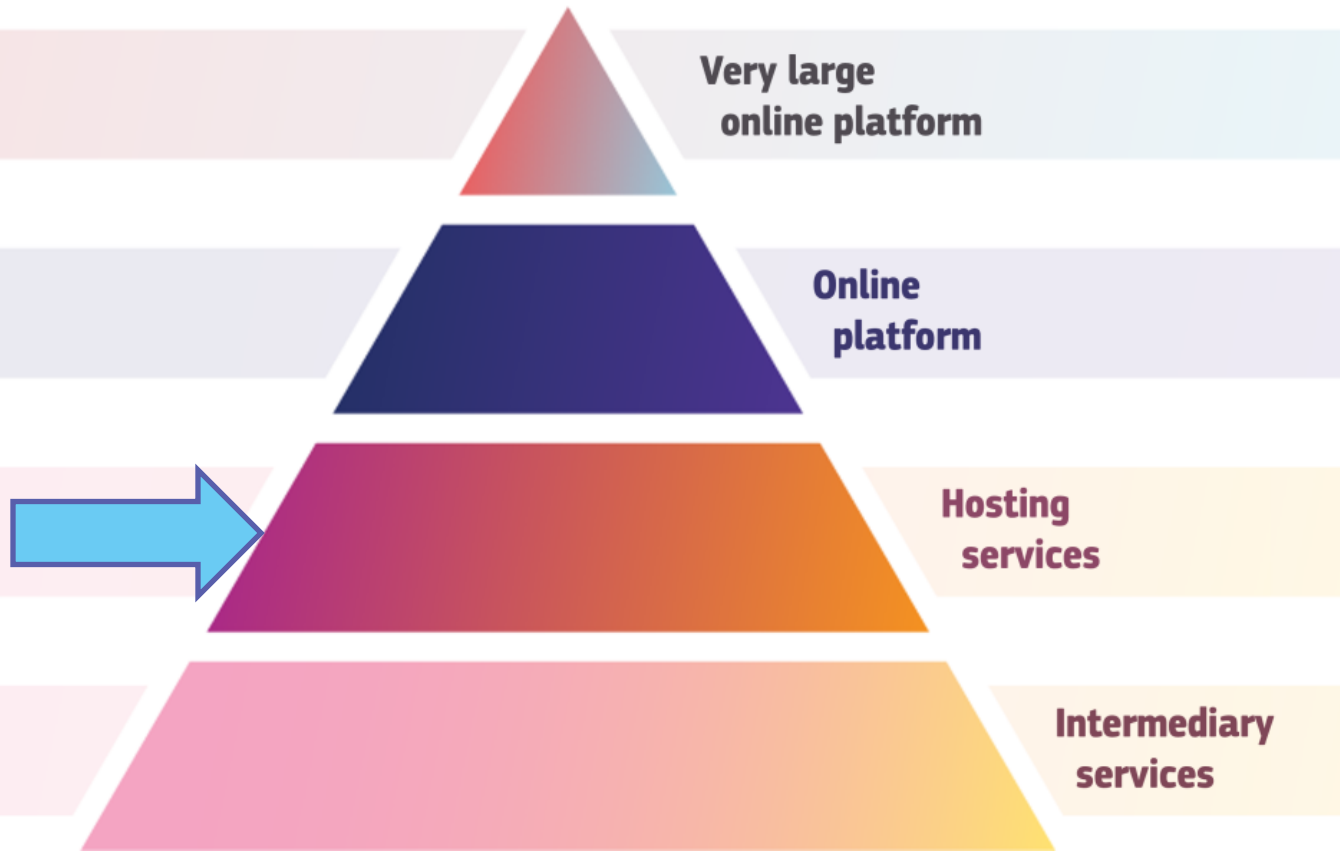
Source: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

To whom does the DSA apply?



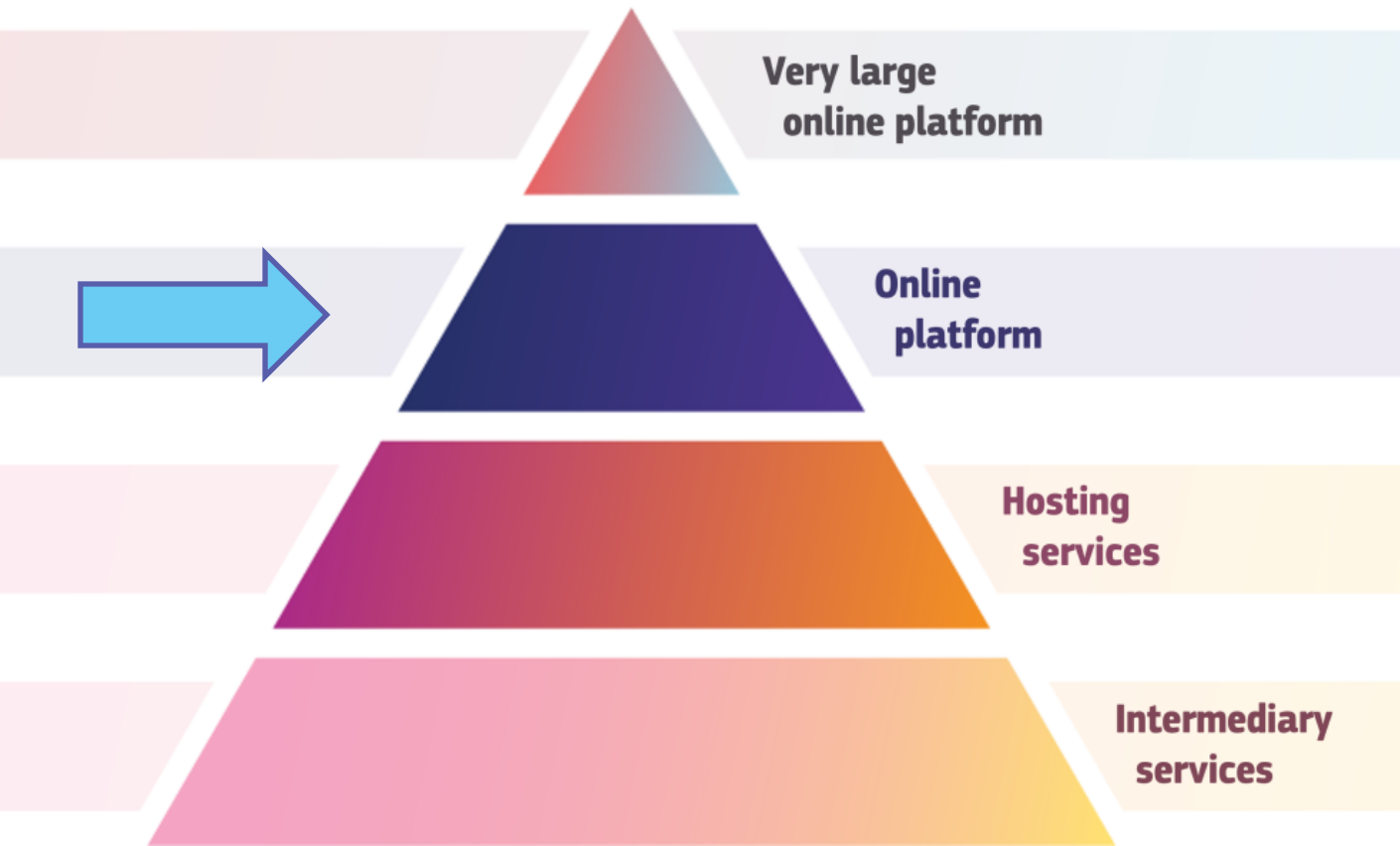
- “*Mere conduit*”: providers of **communication services**.
- *Example*: services offering network infrastructure, such as Internet access providers and domain name registrars.

To whom does the DSA apply?



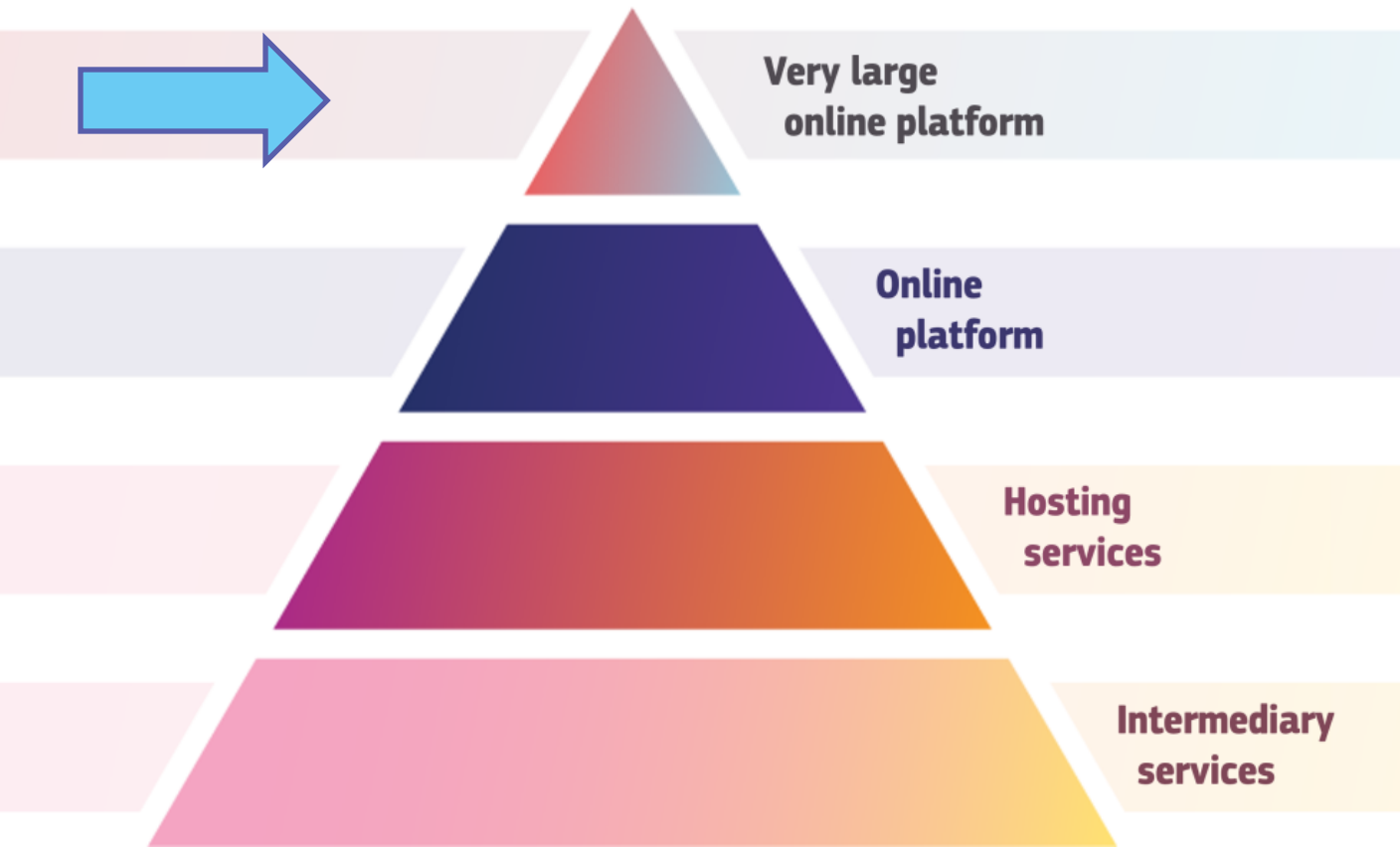
- Services that **store** information at the recipient's request.
- *Example:* cloud and web hosting services (also including online platforms).

To whom does the DSA apply?



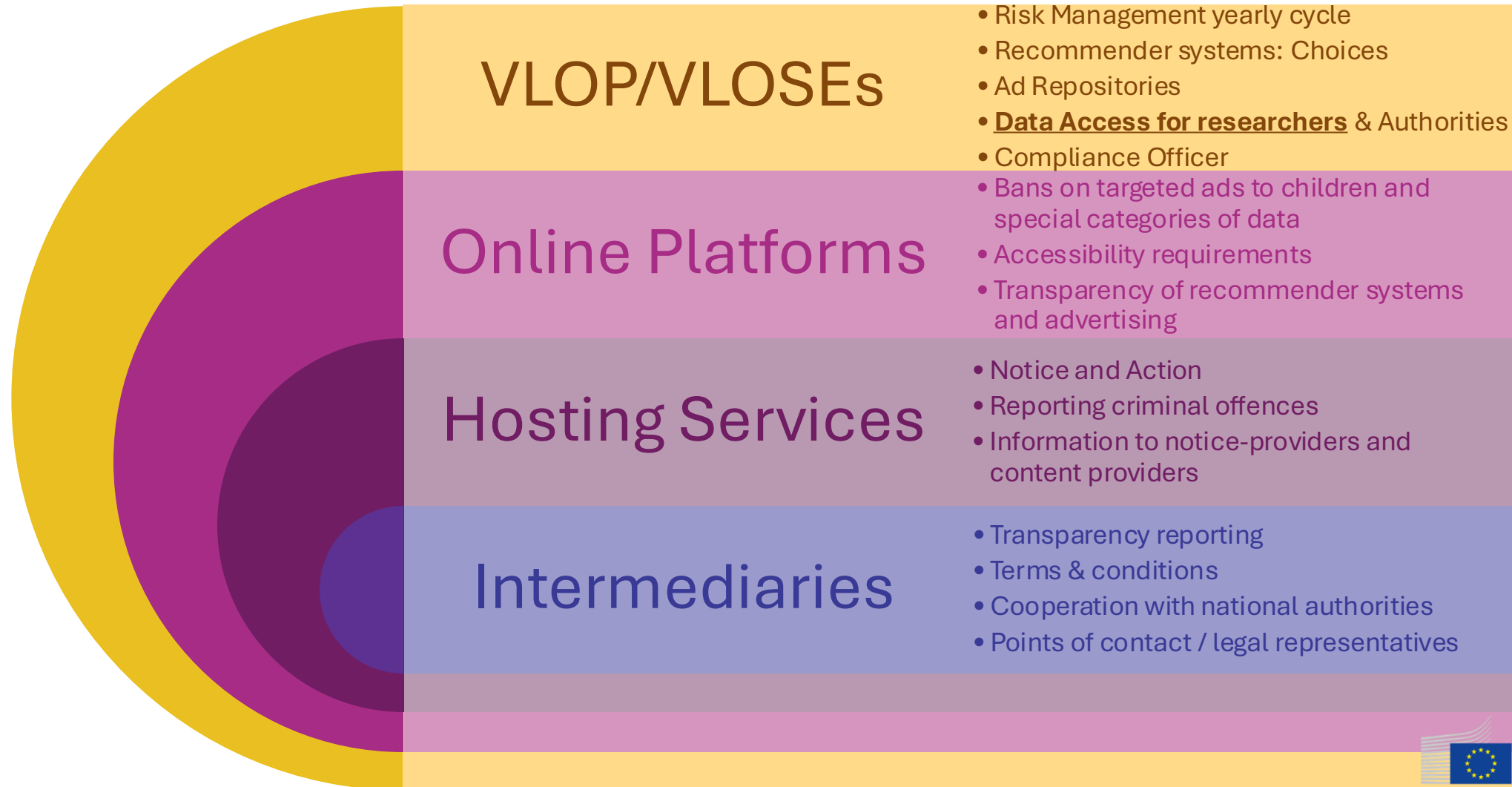
- Services that **store** and **disseminate** information at the recipient's request.
- *Example:* online platforms that bring together sellers and consumers (such as online marketplaces, app stores, collaborative economy platforms and social media platforms).

To whom does the DSA apply?



- Online platforms with more than **45 million monthly active recipients in the EU**, so-called Very Large Online Platforms (**VLOPs**) and Very Large Search Engines (**VLOSEs**).
- Supervised at EU level by the EC
- **Linked to data access**

What obligations for digital services in the DSA?



Very Large Platforms and Search Engines*

- AliExpress
- Amazon Store
- Apple App Store
- Booking.com
- Google Search
- Google Play
- Google Maps
- Google Shopping
- LinkedIn
- Facebook
- Instagram
- Bing
- XNXX
- Pinterest
- Pornhub
- Snapchat
- Shein
- Stripchat
- TikTok
- X
- Temu
- XVideos
- Wikipedia
- Youtube
- Zalando

* **As of 27th May 2025.** Source: <https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses>

Who enforces the DSA?



Digital Services Coordinator in EU Member States

- Independent authorities
- Direct supervision and enforcement of platforms with less than 45 million monthly active recipients in the EU
- Coordination and exchanges with other national authorities



European Board for Digital Services

- Ad-hoc independent advisory group
- Composed by national DSCs
- Chaired by the EC
- Advises DSCs and EC, issues recommendations
- Ensures consistent application of the DSA



European Commission

- Direct enforcement of the rules for VLOPs and VLOSEs
Advises on cross border disputes
- Intervenes following DSC requests

Transparency features of the DSA

Transparency Reports



Bi(annual) statistics on content moderation, incl. accuracy, speed & human resources

Transparency Database



Near real time content moderation seismometer

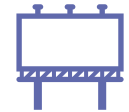
Terms & Conditions



 **OPEN TERMS ARCHIVE**

Clear and Transparent Language of the T&C

Ad Library



Repository of the ads hosted by VLOPs & VLOSEs

Risk Assessments



Analysis of algorithmic risk factors

Independent Audits



Test of algorithmic systems

Data Access



Study of systemic risks

Whistleblower tool



Employee and stakeholders can anonymously report bad practices and infringements

Systemic risks

- Dissemination of illegal content
- Negative effects in fundamental rights*
 - Human dignity
 - Respect for private and family life
 - Protection of personal data
 - Freedom of expression and information
 - Non-discrimination
 - Respect for the rights of the child
- Negative effects on civic discourse and electoral processes, and public security
- Negative effects on:
 - Gender-based violence
 - Protection of public health and minors
 - Serious negative consequences to personal physical and mental well-being

* Charter of the Fundamental Rights of the European Union

Data Access

Three types of data access in Article 40

EC and DSC of
establishment
(40.1)

VLOPSEs to provide **data for monitoring and assessment of compliance** with DSA obligations.

Vetted
researchers
(40.4)

Reserved to **researchers fulfilling specific conditions**, gives access to **non-public data**, only for research projects that contribute to the understanding of systemic risks in the EU.

'Qualified' Researchers
(40.12)

For researchers and CSOs, provides access to tools such APIs with **public data**, and protects those who use automated access means (e.g. scraping) to study EU systemic risks.

Access to publicly
available data

Data access for researchers

DSA Article 40(12)

Providers of **very large online platforms or of very large online search engines** shall give access **without undue delay** to data, including, where technically possible, **to real-time data**, provided that the data is **publicly accessible in their online interface** by researchers, including those affiliated to **not for profit bodies, organisations and associations**, who comply with the **conditions set out in paragraph 8, points (b), (c), (d) and (e)**, and who use the data **solely** for performing research that contributes to the **detection, identification and understanding of systemic risks** in the Union pursuant to Article 34(1).

Next slide

Conditions to access data

Access to publicly available data (40.12)

- Independent from commercial interests
- Disclose their funding
- Demonstrate capacity to protect personal data, data security and confidentiality requirements
- Demonstrate the proportionality & necessity of the request and the contribution to the general objective
- Use the data for the sole purpose of contributing to the understanding of EU systemic risks.

Actors

Researchers

**VLOPs &
VLOSEs**

Accessing data

Request data directly to the platform

- VLOPs and VLOSEs have the obligation to give access to data “without undue delay” and “where technically possible, to real-time data”

Access data independently

- Independent data access techniques, such as scraping, crawling and some types of user data donations cannot be prohibited for eligible researchers.

Enforcement actions



On the basis of the RFIs and regular monitoring of compliance, **formal proceedings** are ongoing for **X, TikTok, AliExpress and Meta, Temu** including grievances on DSA Article 40.12



Investigations are ongoing to assess compliance of these platforms with DSA Article 40.12, **feedback from researchers** is important to feed into the Commission's work



Opening decisions will be available in the [Commission webpage](#)
Preliminary findings regarding X sent on 12 July 2024 ([press release](#)) &
AliExpress commitments adopted on 18 June 2025 ([public summary](#))

Access for vetted
researchers

Data access for researchers

DSA Article 40(4)

Upon a reasoned request from the Digital Services Coordinator of establishment, providers of **very large online platforms** or of **very large online search engines** shall, within a **reasonable period**, as specified in the request, **provide access to data to vetted researchers** who meet the requirements in paragraph 8 of this Article, for the **sole purpose of conducting research that contributes to the detection, identification and understanding of systemic risks in the Union**, as set out pursuant to Article 34(1), and to the **assessment of the adequacy, efficiency and impacts of the risk mitigation measures** pursuant to Article 35.

Conditions for researchers

Vetted Access (40.4)

- **Affiliated to research organisations (cf. Copyright Directive)**
- Independent from commercial interests
- Disclose their funding
- Demonstrate capacity to protect personal data, data security and confidentiality requirements
- Demonstrate the proportionality & necessity of the request and the contribution to the general objective
- **Commit to making the research results publicly available**
- Use the data for the sole purpose of contributing to the understanding of EU systemic risks, **and to the assessment of the adequacy, efficiency and impacts of the risk mitigation measures**

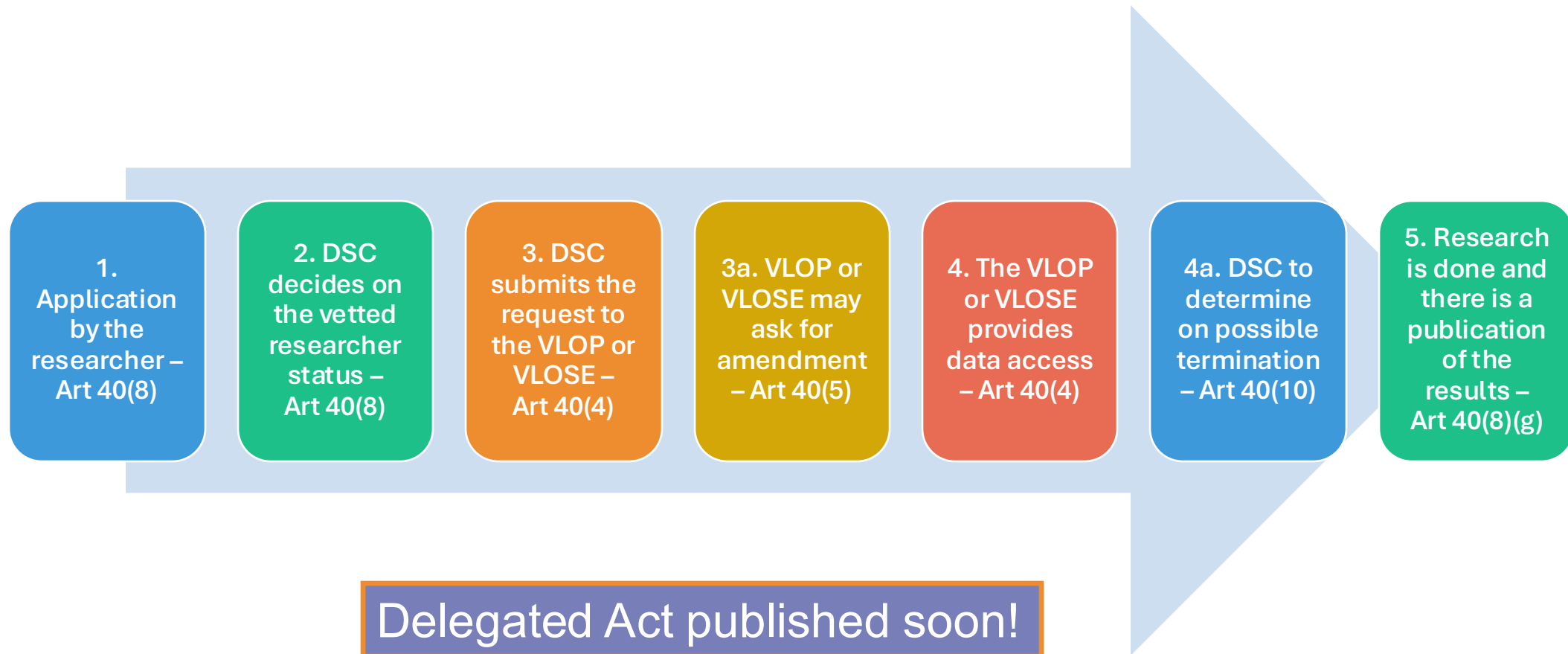
Actors

Researchers

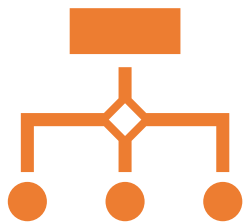
**Digital
Services
Coordinators**

**VLOPs &
VLOSEs**

Data access for vetted researchers



Application by researchers



One application for each project and each VLOP/VLOSE



All team members (1 main applicant + team) must fulfil all conditions:

How to fulfil data protection, confidentiality, security requirements?

What are the concerns?

Protection of personal data*

Protection of confidential information, including trade secrets

Security of the service

What safeguards?



Technical: clean rooms/data vaults, differential privacy, activity logging etc.



Organizational: restricting access to few individuals, no right to subcontract/add people to the project, oversight by [Data Protection Officers/ethics board/other]



Legal: case-by-case Non-Disclosure Agreements between provider and researchers

* If personal data is requested, a **data protection impact assessment** should be submitted

Data transfers and processing need to **comply with GDPR**

Submission and assessment by DSCs

Application can be submitted either to:

- DSC of establishment of VLOP/VLOSE, or
- DSC of organisation of researcher

DSC of organisation of researcher will forward to DSC of establishment of VLOP/VLOSE after initial assessment

If all conditions are met, DSC of establishment grants the vetted status to applicants and sends a reasoned request to the VLOP or VLOSE to grant access to the vetted researchers

Reasoned request

Within 80 days of the submission, DSCs should either:

formulate the reasoned request, submit it to the data provider and inform principal researcher

inform why the reasoned request could not be formulated



In case more time is needed by the DSC, principal researcher must be informed of the the new date and reasons for the delay

Amendment requests

Within 15 days, VLOP/VLOSE may request amendments (suggesting alternatives if possible), in two cases:

Provider has no access to the data

Sharing data leads to significant vulnerabilities to security of the system or confidential information

DSC has 15 days to decide on the amendment request

If VLOP/VLOSE disagrees, a mediation procedure may be requested

Provision of access, termination and publication of results



Providers have to facilitate access to data during the time and under the modalities specified in the reasoned request



DSC may determine access termination if conditions for vetting are no longer met

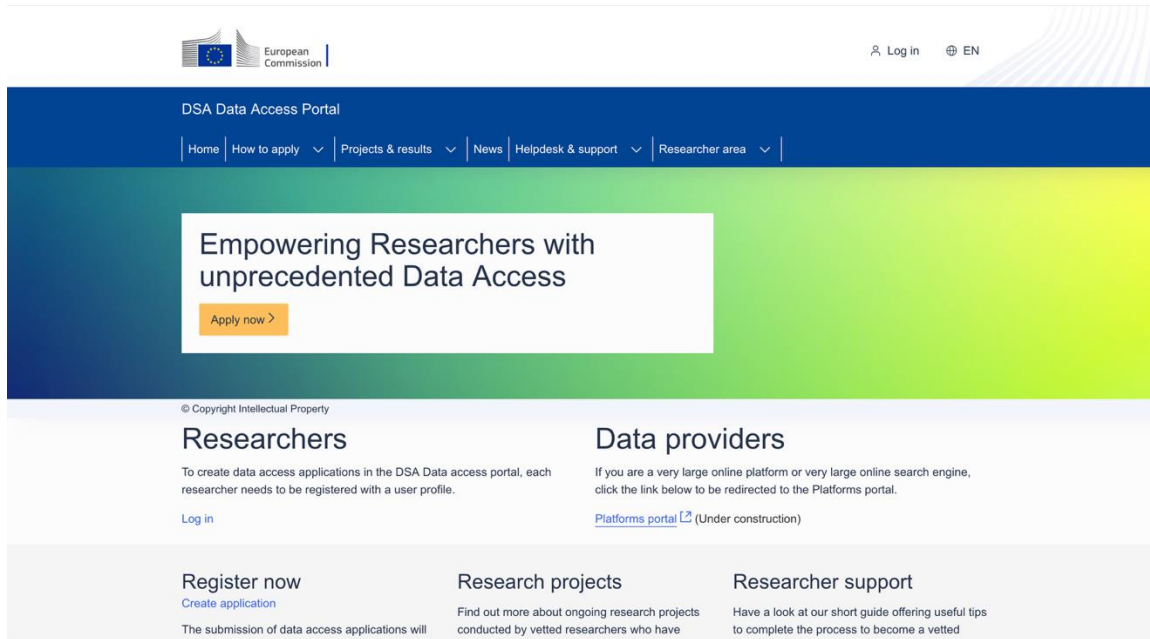


Researchers are required to publish results free of charge when project is over

Delegated Act on Data Access for researchers

- A **delegated act** further specifies the conditions under which sharing of data should take place and relevant procedures, taking into account the rights and interests of the actors involved.
- A draft of this delegated act underwent public consultation from 29 October to 10 December 2024 (**109 inputs received** from researchers, CSOs and businesses).
 - Adoption of a revised text **on July 2nd 2025**
 - Data access portal already online
 - Entry into force on **early November** (3 months after adoption)

DSA Data Access Portal



Developed and hosted by the Commission

One-stop-shop for information, contact points and procedures on access to data for researchers and data providers

Standardisation of data access applications and exchanges with DSCs and data providers

List of reasoned requests to data providers in the DSA data access portal public interface



Application form already online

Points of contact and data catalogues

- All DSCs must publish their points of contact (also on data access portal)
- **Data providers** must make easily findable on their interfaces:
 - Details of their **point of contact**;
 - Link to data access portal;
 - The DSA **data catalogue** (data assets, along with structure and metadata);
 - Suggested access modalities, according to level of sensitivity.



Examples of data that can be requested

Recital (13):

(...) Current examples of such data include data related to users of the services, such as profile information, relationship networks, individual-level content exposure and engagement histories; interaction data such as comments or other engagements; data related to content recommendations, including data used to personalise recommendations; data related to the targeting of advertisements and profiling, including cost per click data and other measures of advertising prices; data related to the testing of new features prior to their deployment, including the results of A/B tests; data related to content moderation and governance, such as data on algorithmic or other content moderation systems and processes, including changelogs, archives or repositories documenting moderated content, including accounts as well as data related to prices, quantities and characteristics of goods or services provided or intermediated by the data provider.

Interactive session:

- Filling in an application