Knight Georgetown Institute

December 2025

Better Access: Data for the Common Good

A Guide for Platforms and EU Policymakers

Overview

Online platforms and services shape what we know, how we connect, and who gets heard. From elections and public health to commerce and conflict, platforms now function as essential infrastructure for civic life. Their influence is vast, and so is the need to understand them.

As critical conversations publicly unfold on digital platforms, the ability to study them at scale has steadily diminished. Platform interfaces and insight tools like CrowdTangle that once gave researchers, journalists, and civil society a window into public online discourse have disappeared. Platforms now restrict researcher access to public data while monetizing the same data for advertisers, data brokers, and training of artificial intelligence (AI) systems. This imbalance – where companies profit while independent researchers are left in the dark – undermines transparency, limits free expression, and weakens oversight.

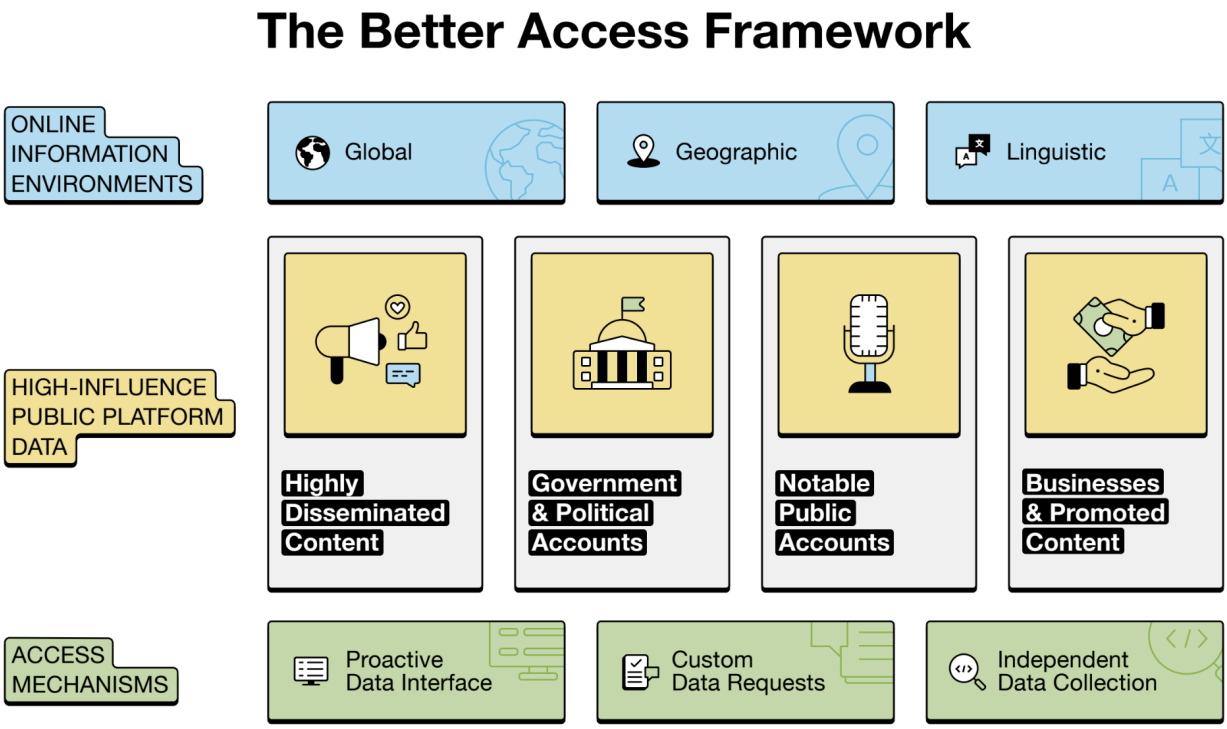
European policymakers have taken steps to expand access to public platform data through the Digital Services Act (DSA). Article 40.12 of the DSA requires platforms to provide researchers access to public platform data "including, where technically possible, to real-time data, provided that the data is publicly accessible in their online interface." Yet the law leaves several critical questions unanswered: What counts as publicly accessible platform data? How can data be accessed? And how should platforms operationalize DSA obligations in ways that support research and strengthen public oversight?

The Knight-Georgetown Institute (KGI)'s ***Better Access*** framework addresses these questions by offering a baseline model for **independent access to public platform data**: the content, data, and information posted to platforms that anyone can access. The framework defines the minimum forms of public platform data that platforms should make available to a wide range of researchers to be compliant with DSA Article 40.12, why this data matters, and how access should work in practice. The *Better Access* framework and definitions support effective implementation of the DSA's obligations.

A Framework for High-Influence Public Platform Data

The *Better Access* framework offers a roadmap for expanding access to high-influence public platform data: the narrow slice of public platform data that has the greatest impact on civic life due to its reach, source, or role in shaping what people see online. The framework focuses on the narrow subset of public posts, accounts, and interactions that matter most for shaping public discourse.

Better Access defines the kinds of public platform data that researchers should be able to ethically use in their work. This definition sets a floor, not a ceiling – research with other public platform data as well as non-public data, as envisioned under Article 40.4 of the DSA, is also important. The framework also sets out a model for data access mechanisms that platforms should implement in order to enable research, while acknowledging that permissionless, researcher-driven access to public platform data will remain necessary.



Applying *Better Access* to the Digital Services Act

Article 40.12 of the DSA requires in-scope companies to provide researchers with timely access to publicly accessible data to enable research that contributes to the detection, identification, and understanding of systemic risks in the EU. It also requires access to be proportionate to the research purpose. The *Better Access* framework establishes the minimal information environments, data access mechanisms, and data categories that platforms need to support to comply with DSA Article 40.12.

Defining Information Environments in the EU

Public platform data is present across the EU, but its influence can vary. A viral post in Austria could tell us little about popular trends in Sweden. For platform data access to be meaningful in different EU member states, it must respond to local researcher needs.

To account for differences across contexts, the *Better Access* framework emphasizes three types of information environments:

Global: A platform's total user base across languages and geographies.

Geographic: Users of a platform in a particular geography, for example the EU, a member state, or region of a member state.

Linguistic: Users of a platform engaging in a specific language or languages.

Article 40 should enable research across information environments in the EU. Data needs will vary depending on research questions and information environments. DSA data access should ensure that researchers studying systemic risks are able to access data relevant and proportionate to their specific research questions in specific information environments.

DSA Article 40 Access Mechanisms

To ensure effective access, the *Better Access* framework defines three complementary data access mechanisms that platforms should support. Each of these mechanisms should provide access to public platform data under Article 40.12.

Proactive Data Interface: A structured, platform-supported access mechanism that provides ongoing, predictable access to high-influence public platform data in the EU, as well as other necessary public platform data under Article 40.12. Examples include research Application Programming Interfaces (APIs), low-code/no code interfaces, searchable archives, or downloadable datasets that are updated regularly.

Custom Data Requests: Tailored datasets provided upon request to meet the specific local or thematic needs of researchers studying the detection, identification, and understanding of systemic risks in the EU. These requests may be fulfilled through bespoke datasets, tailored APIs, archives, or other methods.

Independent Data Collection: Researcher-initiated access to high-influence public platform data, typically through automated collection such as scraping or crawling. This mode of access under the DSA preserves researcher independence and enables validation of platform-provided data relevant to the detection, identification, and understanding of systemic risks in the EU.

These three access mechanisms broadly align with the European Commission's agreements with AliExpress to enable researcher access to public platform data. Multiple access mechanisms are necessary to ensure flexibility, relevance, and accountability across EU research environments and should be required for all platforms. Platforms and the European Commission should ensure access to each of these data modalities, including through effective platform interfaces, accessible channels for custom data requests, and specific updates to platform terms of service to clarify the permissibility of independent data collection.

Specifying Publicly Available Platform Data Under the DSA

The DSA requires platforms to provide researchers with access to data that is publicly accessible in their online interface. However, the law does not define what "publicly accessible" means, and platforms interpret the scope of that term inconsistently.

The *Better Access* framework takes a pragmatic approach to these issues by focusing on high-influence public platform data: the subset of public platform data that is most consequential, either because of its reach, its source, or its role in shaping what people see online. This is the data that tells us who holds influence, what content spreads, and how platforms themselves amplify certain voices.

Public posts vary in their significance. A family photo shared to a handful of followers is not the same, and should not be treated the same, as a head of state announcing a new policy or a viral video

reaching millions. The *Better Access* framework focuses narrowly on the types of high-influence public platform data that shape civic discourse and access to information.

The framework identifies four categories of high-influence public platform data:

Highly Disseminated Content: Posts or videos that achieve exceptional reach or engagement, shaping the public agenda in EU member states.

Government and Political Accounts: Posts from accounts belonging to elected officials, candidates, political parties, and government institutions in information environments relevant to the EU context.

Notable Public Accounts: Content from accounts belonging to celebrities, journalists, civic leaders, or other public figures whose reach gives them outsized influence in the EU.

Business Accounts and Promoted Content: Advertising and commercial messaging in the EU, which can sway consumer behavior, public health, or public trust.

While high-influence public platform data will represent a small subset of the data that is publicly accessible in a platform's online interface as defined in the DSA, it should be understood as a baseline expectation for research access. This narrower category of public platform data sets clearer boundaries for access and lowers (but does not eliminate) privacy and ethical risks associated with access.

The Way Forward

The stakes could not be clearer: when independent access to high-influence public platform data disappears, so does society's ability to understand and protect itself.

Regulators in the EU have taken important steps to lay the foundation for access to public platform data. The *Better Access* framework offers a practical foundation for implementation of the DSA: begin with the four categories of data that matter most for the public, ensure consistent and ethical access through multiple access mechanisms, and give researchers and the public the tools they need to understand the dynamics shaping public discourse in their own information environment.

High-influence public platform data is the minimum, most essential tier of data for accountability. The *Better Access* framework underscores that transparency is not a regulatory burden, but a democratic necessity. By adopting a uniform, cross-industry baseline for such data, the Commission can ensure the public has the visibility needed to see the digital public square clearly, and to help shape its future.