Knight Georgetown Institute

November 2025

Better Access: Data for the Common Good

A Framework for Accessing High-Influence Public Platform Data

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 are now indispensable 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 restrict researcher access while public data is increasingly monetized for advertisers, data brokers, and training 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.

That is the reason for developing ***Better Access***, a baseline framework for **independent access to public platform data**: the content, data, and information posted to platforms that anyone can access.

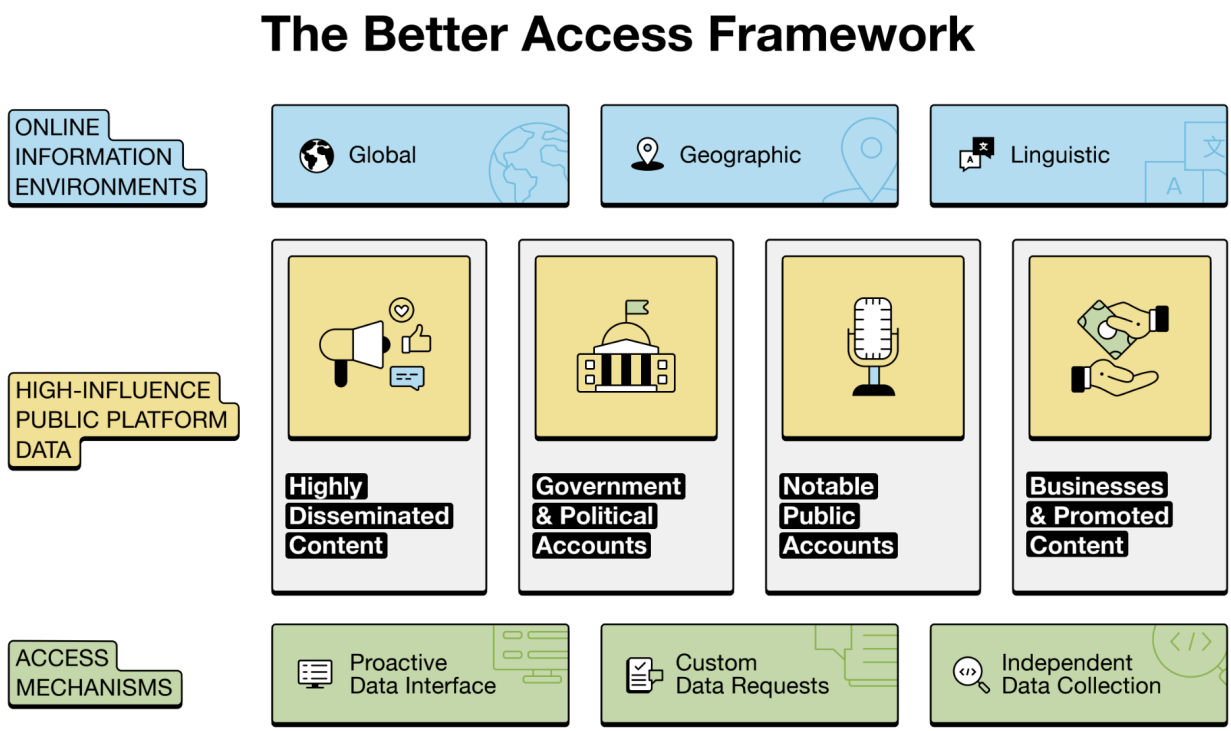
Published by the Knight-Georgetown Institute (KGI), this framework is the product of KGI's Expert Working Group (EWG) on Public Platform Data, a group of leaders from research, civil society, and journalism.

A Framework for High-Influence Public Platform Data

The 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.

Public platform data does not all carry the same weight. Some accounts and content have far more influence than others. The *Better Access* framework focuses on the 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 or non-public data 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.



Categories of High-Influence Public Platform Data

High-influence public platform data refers to 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. It 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 data:

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

Government and Political Accounts: Posts from accounts belonging to elected officials, candidates, political parties, and government institutions.

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

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

Defining Information Environments

What counts as influential varies dramatically across contexts. A viral post in the United States or India will tell us little about online discussion in Angola or Albania. For platform data access to be meaningful in different information environments, it must respond to local needs.

To account for these differences, the framework emphasizes three types of information environment:

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

Geographic: Users of a platform in a particular geography.

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

This approach ensures that vital research does not privilege only the largest countries or dominant languages.

How Access Should Work

To make public platform data meaningfully available, the framework defines three complementary data access mechanisms that platforms should support:

Proactive Data Interface: A structured, platform-supported access mechanism that provides ongoing, predictable access to high-influence public platform data. Examples include research Application Programming Interfaces (APIs), searchable archives, or downloadable datasets that are updated regularly without requiring researchers to make individual requests.

Custom Data Requests: Tailored datasets provided upon request to meet the specific local or thematic needs of researchers. 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 preserves researcher independence and enables validation of platform-provided data.

Together, these mechanisms can advance flexibility, relevance, and accountability across diverse research settings. Furthermore, this approach to high-influence public platform data can help to advance ethical research. The framework emphasizes access to public platform data categories where the public value of transparency is strongest. By focusing here, the framework sets clearer boundaries, lowers (but does not eliminate) privacy risks, and articulates a baseline expectation for consistent, responsible 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 Europe, the US, and beyond are already implementing or considering data access mandates. This framework offers a practical foundation: 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. This framework underscores that transparency is not a regulatory burden, but a democratic necessity. Adopting a uniform, cross-industry baseline would give the public the visibility it needs to see the digital public square clearly and to shape its future.