

[We gratefully acknowledge support from the Simons Foundation and member institutions.](#)

## [export.arXiv.org](#) > [help](#)



([Help](#) | [Advanced search](#))

[Help Table of Contents](#) |

# Robots Beware

*Indiscriminate automated downloads from this site are not permitted.* We have limited server capacity and our first priority is to support interactive use by our many human users. Several interfaces designed to provide machine access to arXiv are provided. See our [OAI-PMH](#), [arXiv API](#) and [RSS](#) documentation. There are also facilities for [bulk data](#) download.

## Millions and billions of distinct URL's

The [arXiv.org website](#) is under all-too-frequent attack from [robots](#), spiders and accelerators that mindlessly download every link encountered, ultimately trying to access the entire database through the listings links. Obviously, large search engines offer an invaluable service to web users and we work with them to find efficient and effective ways to index arXiv content. In many cases, however, we are subject to accidental denial-of-service attacks by well-intentioned but thoughtless novices, ignorant of common sense [guidelines](#).

Following the de-facto [standard for robot exclusion](#), this site has maintained since early 1994 a file [/robots.txt](#) that specifies those URL's that are off-limits to robots (and this "Robots Beware" page was originally posted March 1994).

Mindlessly downloading all of the URLs on this site will return terabytes of data. This has very real cost to us in terms of bandwidth consumed, and in terms of the responsiveness of our service for our many tens of thousands of real users.

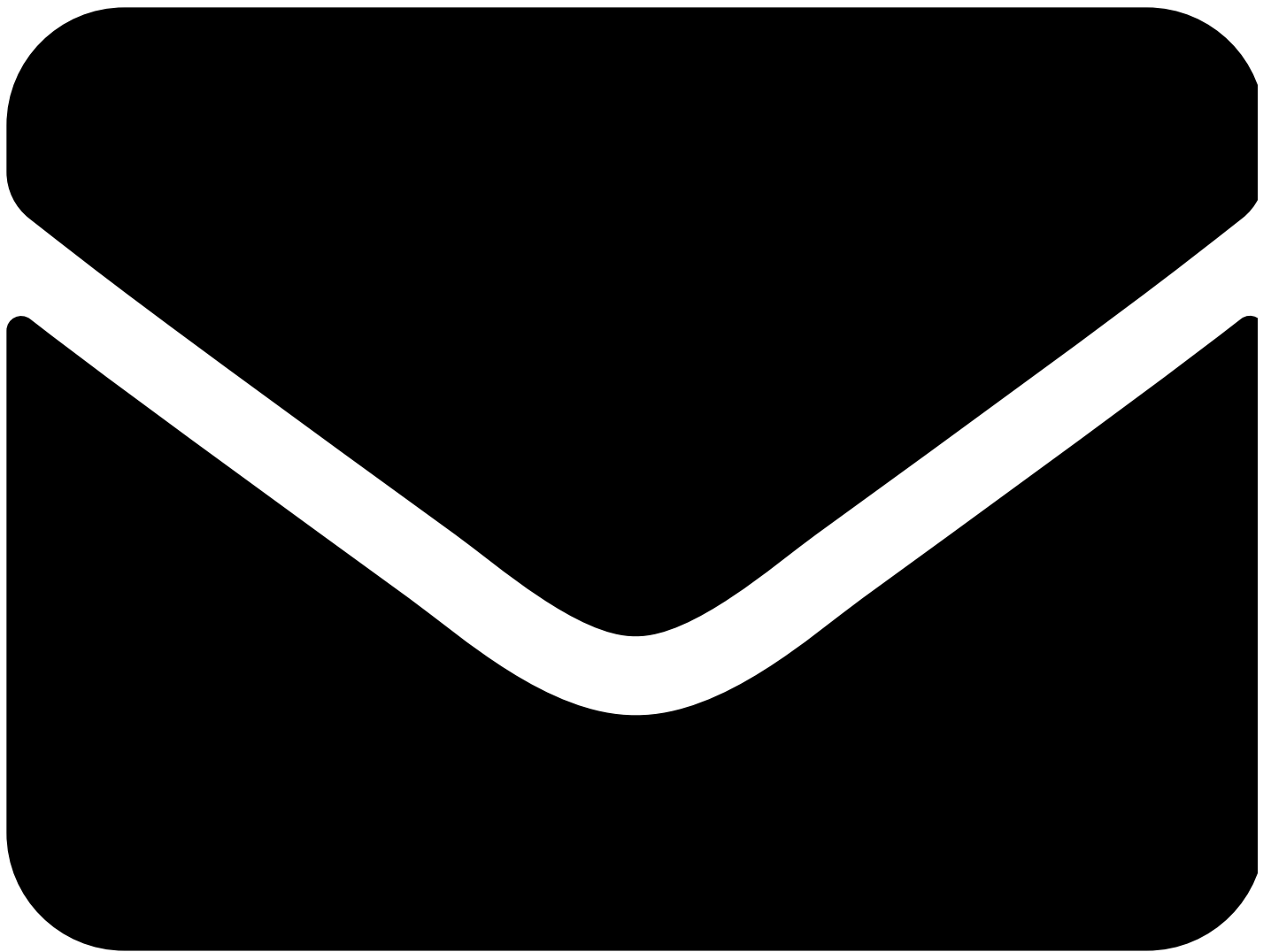
This server is configured to monitor activity and deny access to sites that violate the above guidelines. Continued rapid-fire requests from any site after access has been denied (i.e. with *403 Access denied* HTTP response) will be interpreted as an attack; and we will respond accordingly — without hesitation, and without further warning.

If some specific application requires relaxation of the above guidelines, [contact the arXiv administrators](#) *in advance* of any attempted download.

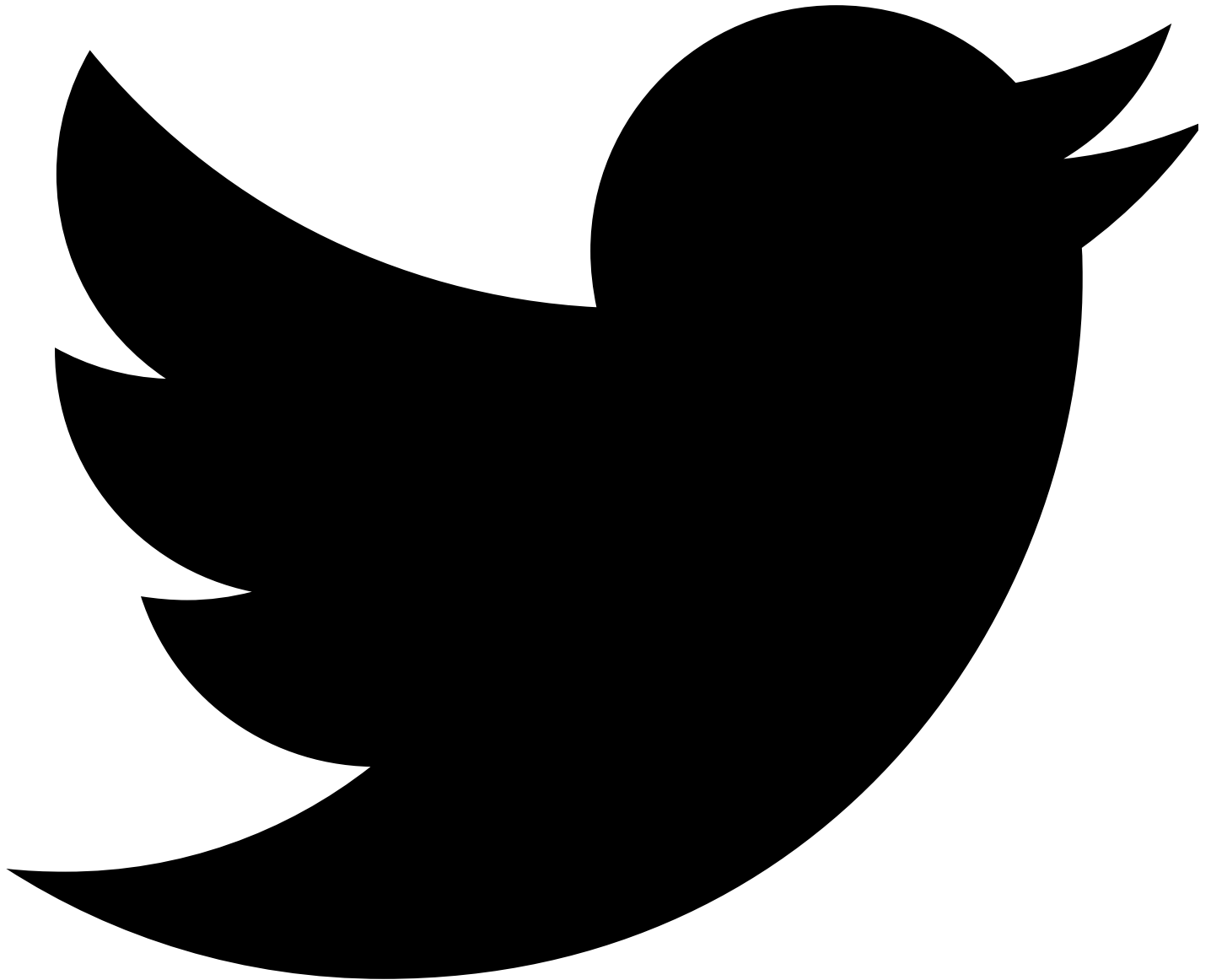


[About arXiv](#)

[Leadership Team](#)



[Contact](#)



[Follow us on Twitter](#)

[Help](#)

[Privacy Policy](#)

[Blog](#)

[Subscribe](#)

---

arXiv® is a registered trademark of Cornell University.

If you have a disability and are having trouble accessing information on this website or need materials in an alternate format, contact [web-accessibility@cornell.edu](mailto:web-accessibility@cornell.edu) for assistance.