



Cornell University
Library

Repository sustainability: arXiv business model experience and implications

Simeon Warner and Oya Yildirim Rieger

DLIT, Cornell University Library
(simeon.warner@cornell.edu, rieger@cornell.edu)



All papers

Go!

Open access to 576,488 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

Subject search and browse: 17 Nov 2009: [2009 holiday schedule](#)8 Apr 2009: [Added public author identifiers, Facebook](#)See cumulative ["What's New"](#) pages**Robots Beware:** [indiscriminate automated downloading](#)

All papers

Go!

Physics

- [Astrophysics](#) (**astro-ph** [new](#), [recent](#), [find](#))
includes: [Cosmology and Extragalactic Astr](#)
[Energy Astrophysical Phenomena](#); [Instrum](#)
- [Condensed Matter](#) (**cond-mat** [new](#), [recent](#),
includes: [Disordered Systems and Neural N](#)
[Condensed Matter](#); [Quantum Gases](#); [Soft C](#)
[Superconductivity](#)

The Persistence of Cool Galactic Winds in High Stellar Mass Galaxies Between $z \sim 1.4$ and ~ 1

Kate H. R. Rubin (1), Benjamin J. Weiner (2), David C. Koo (1), Crystal L. Martin (3), J. Xavier Prochaska (1), Alison L. Coil (4), Jeffrey A. Newman (5)
(1) UCO/Lick Observatory, (2) Steward Observatory, (3) UCSB, (4) UCSD, (5) Univ. of Pittsburgh)

(Submitted on 13 Dec 2009)

We present an analysis of the MgII 2796, 2803 and FeII 2586, 2600 absorption line profiles in coadded spectra of 468 galaxies at $0.7 < z < 1.5$. The galaxy sample, drawn from the Team Keck Treasury Redshift Survey of the GOODS-N field, has a range in stellar mass (M_*) comparable to that of the sample at $z \sim 1.4$ analyzed in a similar manner by Weiner et al. (2009; W09), but extends to lower redshifts and has specific star formation rates which are lower by ~ 0.6 dex. We identify outflows of cool gas from the Doppler shift of the MgII absorption lines and find that the equivalent width (EW) of absorption due to outflowing gas increases on average with M_* and star formation rate (SFR). We attribute the large EWs measured in spectra of the more massive, higher-SFR galaxies to optically thick absorbing clouds having large velocity widths. The outflows have hydrogen column densities $N(\text{H}) \sim 10^{21} - 2 \times 10^{22} \text{ cm}^{-2}$ and extend to velocities of

Download:

- [PostScript](#)
- [PDF](#)
- [Other formats](#)

Current browse context:

astro-ph.CO[< prev](#) | [next >](#)[new](#) | [recent](#) | [0912](#)

Change to browse by:

[astro-ph](#)

References & Citations

- [SLAC-SPIRES HEP](#)
([refers to](#) | [cited by](#))
- [NASA ADS](#)

Bookmark [\(what is this?\)](#)

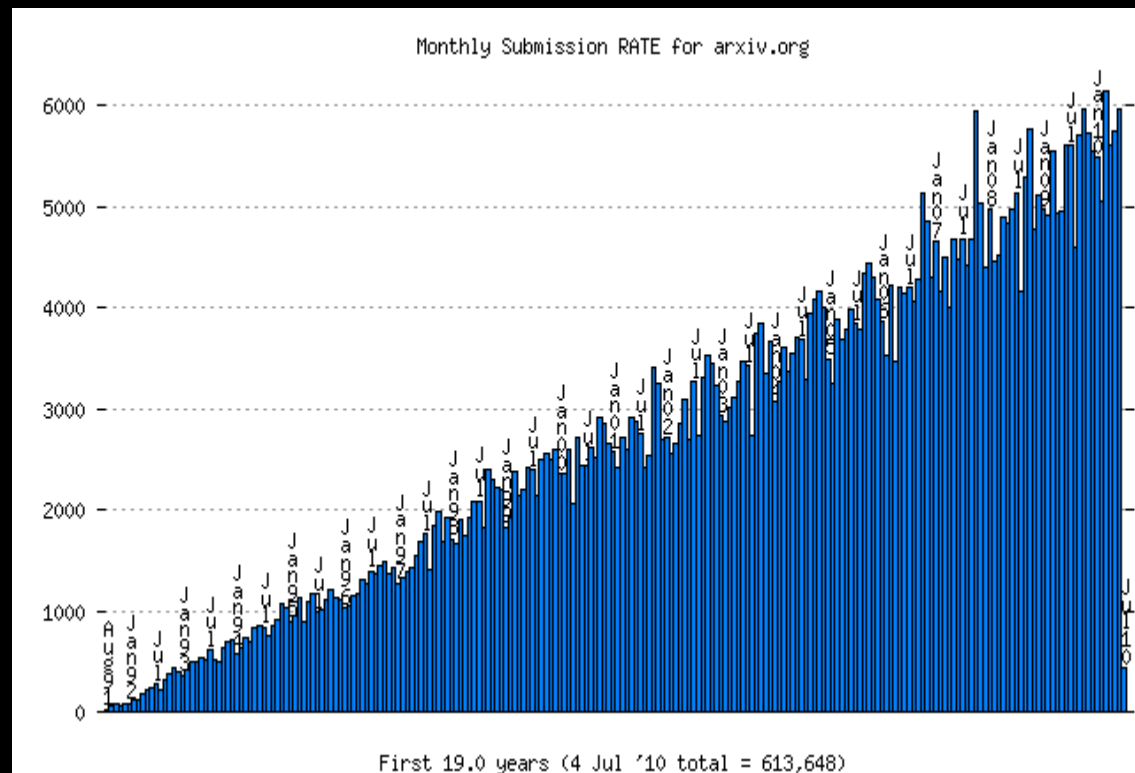
History of a disciplinary repository

1991 hep-th @ LANL

- gradual subject expansion: + alg-geom, cond-mat, hep-*... math, nlin, cs
- technology evolution: email -> ftp -> web, tex -> ps -> pdf
- mirrors, Front, SPIRES
- OAI-PMH

2001 Cornell Library

- q-bio, stat, q-fin
- logins, RSS, docx, pdflatex
- ADS collab, eprintweb.org
- new submission system
- SWORD submission



Statistics

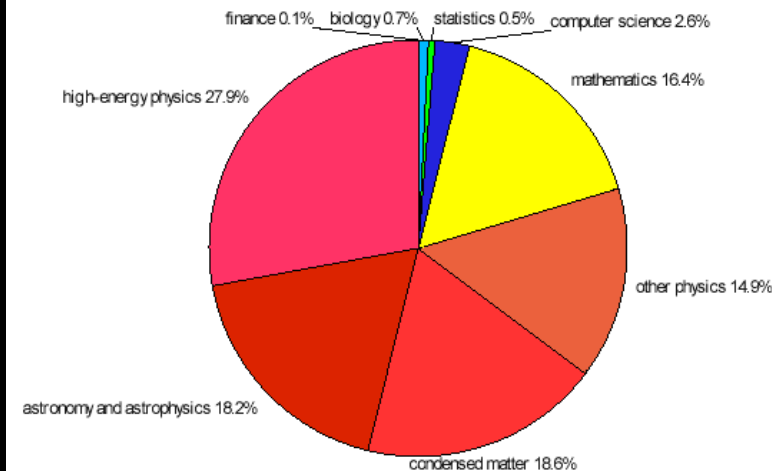
input

- ~613k articles, average ~1.5 versions/article
- ~60k new articles per year (~230/day)
- ~100k user accounts with ≥ 1 article saturated in HEP; growing math, cs, astro-ph, cond-mat

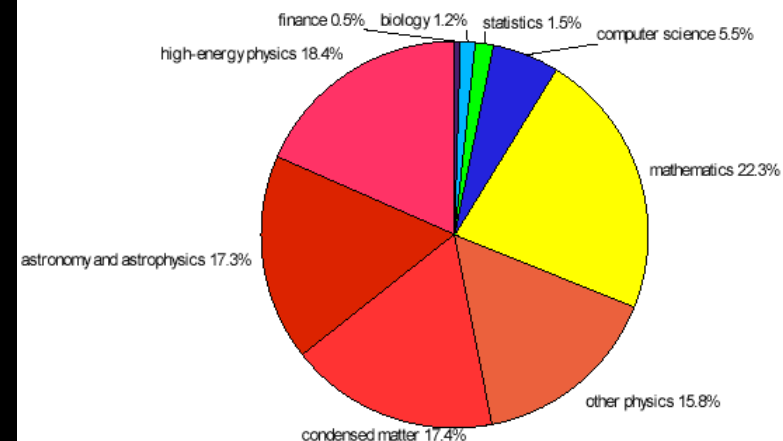
output

- ~30M full-text downloads (2009)
- ~8M “institutional downloads” (2009)
- ~20k email alert subscriptions
- feed many other discovery services (SPIRES, Google, RePEc, etc. ...)

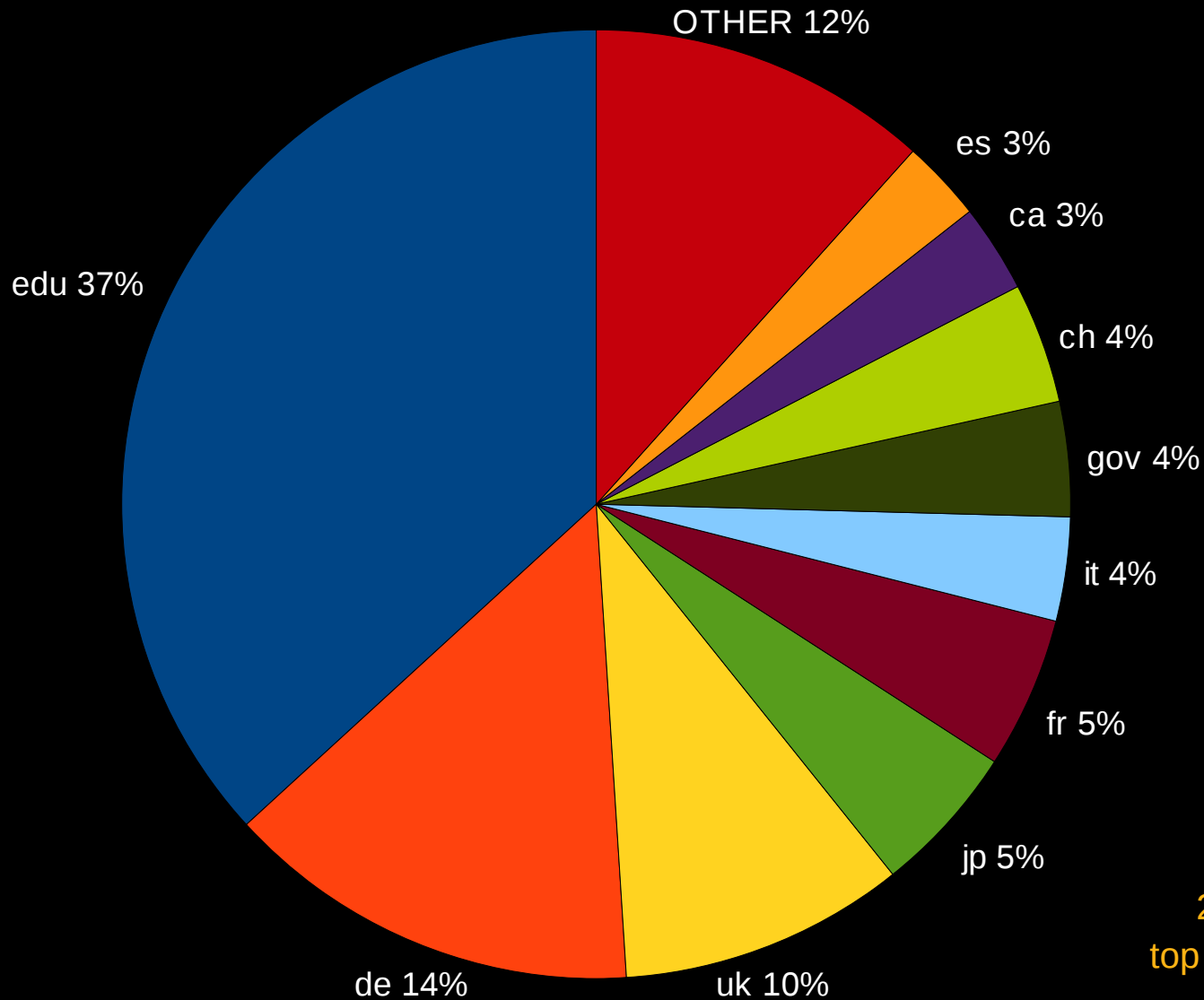
arXiv submissions by subject 1991-2009



arXiv submissions by subject 2008



Institutional downloads by domain



2008, main site,
top 200 institutions

Fraction of use by Cornell: 0.5-0.7%

Fraction of budget paid by Cornell: 80-100%

Obviously different from our IR in that arXiv is not designed to support Cornell researchers any more than those from other institutions

(Budget ~\$400k 2010, arXiv was funded by grants 1991-2001, mix of Cornell Library and grants since 2001. Grant funding for normal operation and mainstream development difficult)

Sustainability

“sustainability is the ability to generate or gain access to the resources needed to protect and increase the value of the content or service for those who use it

- covers its operating costs through a combination of revenue sources and cost-management strategies

- continues to enhance its value based on the needs of the user community”

Mission

perhaps... ***“To provide an effective means of rapid open scholarly communication and archiving to the physics, math, cs and related academic communities at minimal cost.”***

rapid – daily turnaround

serves *academic* communities – not all people

minimal cost –effective cost per submission <\$7

effective cost per download ~1.3c

Principles

free access to articles for individuals. Anyone with an internet connection should be able to download articles for free. (have charged/negotiated for bulk access.)

free submission of articles for individuals. Anyone with an internet connection should be able to upload articles for free. (advisory committees and moderators provide guidance on appropriateness.)

community focus. We serve the scholarly communities first, their needs are put before those of individual authors (hence policy on recording versions, not allowing deletions etc.)

fair/honest/unbiased operation in accord with community standards. (advisory committees provide guidance.)

Process

- Late 2008 and 2009 – spoke with people running other disciplinary scholarly resources (e.g. SSRN, Stanford Encyclopedia, Legal Information Institute)
- Summer 2009 – input from 15+ senior library staff at other universities and laboratories, and input from arXiv scientific advisory board
- Fall/Autumn 2009 – Decided on 3 year collaborative support model requesting voluntary contributions from heavy user institutions
- Late 2009 – sought pledges of support from top 20 institutions, built FAQ addressing concerns
- Feb 2010 – announced plan, started to send requests for support



FOR RELEASE:

Contact: Gwen Glazer
Phone: (607) 254-8390
E-mail: gglazer@cornell.edu

Cornell University Library Engages More Institutions in Supporting arXiv
Collaborative Business Model Changes Funding Structure

ITHACA, N.Y. (Jan. 21, 2010) - In a move to expand support for sustaining arXiv, Cornell University Library is broadening the funding base for the online scientific repository. Nearly 600,000 e-prints - research articles published online in physics, mathematics, statistics, computer science and related disciplines - now reside in arXiv, which is an open information source for hundreds of thousands of scientific researchers.

arXiv will remain free for readers and submitters, but the Library has established a voluntary, collaborative business model to engage institutions that benefit most from arXiv.

"Keeping an open-access resource like arXiv sustainable means not only covering its costs, but also continuing to enhance its value, and that kind of financial commitment is beyond a single institution's resources," said Oya Rieger, Associate University Librarian for Information Technologies. "If a case can be made for any repository being community-supported, arXiv has to be at the top of the list."

The 200 institutions that use arXiv most heavily account for more than 75 percent of institutional downloads. Cornell is asking these institutions for financial support in the form of annual contributions, and most of the top 25 have already committed to helping arXiv.

Institutions that have already pledged support include:

- California Institute of Technology
- University of California, Berkeley
- University of Cambridge (UK)
- CERN - European Organization for Nuclear Research (Switzerland)
- CNRS - Centre National de la Recherche Scientifique (France)
- Columbia University
- DESY - Deutsches Elektronen-Synchrotron (Germany)
- Durham University (UK)
- ETH Zurich - Eidgenössische Technische Hochschule Zürich (Switzerland)
- Fermilab
- Harvard University
- University of Illinois at Urbana-Champaign
- Imperial College London (UK)
- Los Alamos National Laboratory
- Massachusetts Institute of Technology
- Max Planck Society (Germany)
- University of Michigan
- University of Oxford (UK)
- University of Pennsylvania
- Princeton University
- SLAC National Accelerator Laboratory

Collaborative support model

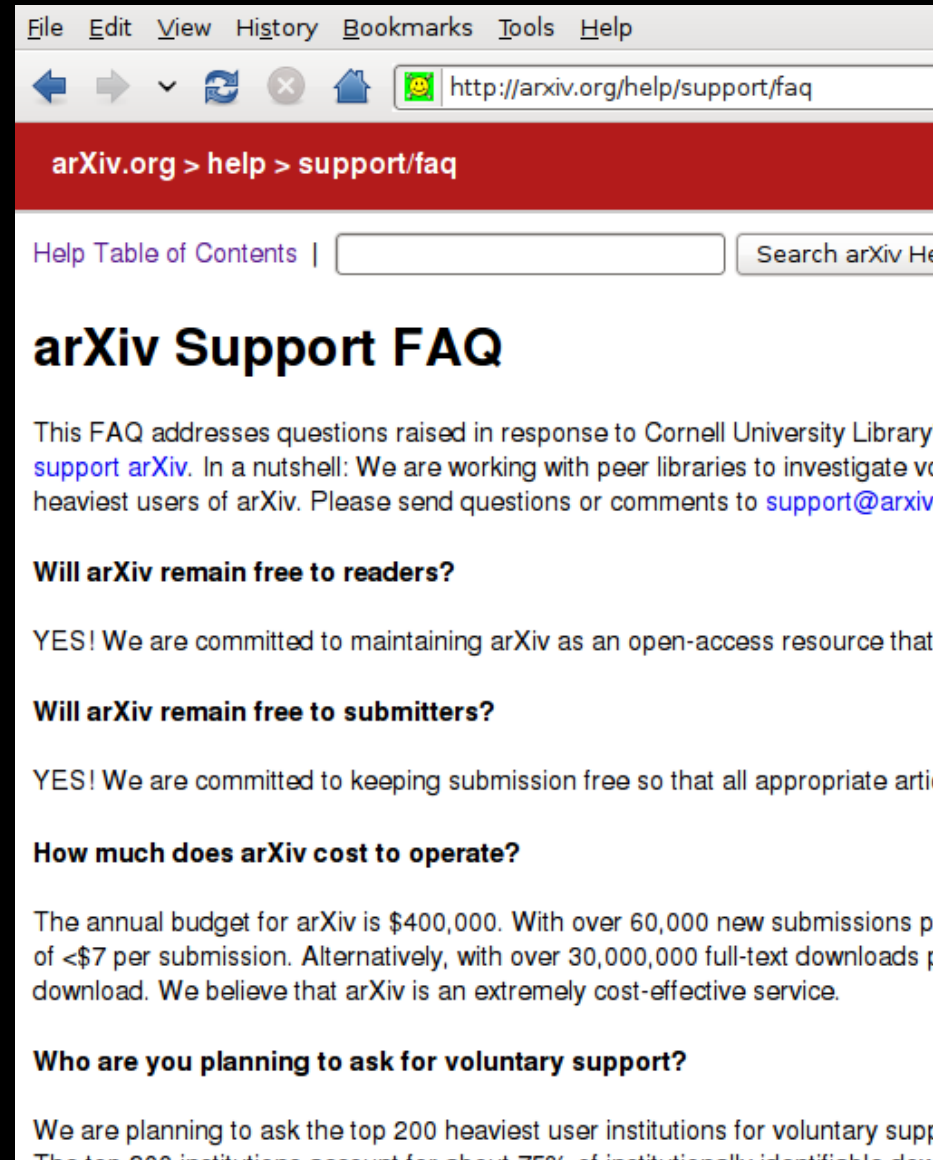
- **many possible models, no obvious answer:** users or user institutions (submissions, downloads), funding bodies, scholarly societies, donations, endowment, advertising, sponsorship, freemium services
- **free riders:** unavoidable issue when principles include free submission and free reading!
- **ask top 200 user institutions for support** at between \$4k (top tier) and \$2.3k (bottom tier) per institution/year
 - initial targets were to raise 25% 2010, 50% 2011
 - short term model (3 years)
 - benefits to supporters (“supported by...”, OpenURL and stats)
 - create additional sustainability advisory group
- **support model brings an additional cost:** ~0.4 FTE admin/mgmt at present, should decrease + dev time

Issues raised by library community

- when is arXiv going to replace the formal journals?
- how will you address the free rider problem?
- why not charge scholars per submission?
- what are the benefits for my institution?
- how will you structure a governance model?
- are you opening a floodgate?
- what are the other potential sources of revenue?
- what is your long-term plan?

Issues raised by user community

- will there be charge for using arXiv?
- will contribution model be mandatory?
- might this model harm the open access cause?
- did you try other ways of raising money? (surely agency X will support arXiv)
- what is your long-term plan?
- isn't this the most important thing Cornell University Library does?



The screenshot shows a web browser window with the address bar displaying <http://arxiv.org/help/support/faq>. The browser's menu bar includes File, Edit, View, History, Bookmarks, Tools, and Help. Below the address bar, the breadcrumb navigation reads "arXiv.org > help > support/faq". The page content includes a "Help Table of Contents" link, a search box, and the main heading "arXiv Support FAQ". The text of the FAQ begins with: "This FAQ addresses questions raised in response to Cornell University Library support arXiv. In a nutshell: We are working with peer libraries to investigate voluntary support from our heaviest users of arXiv. Please send questions or comments to support@arxiv.org".

Will arXiv remain free to readers?

YES! We are committed to maintaining arXiv as an open-access resource that

Will arXiv remain free to submitters?

YES! We are committed to keeping submission free so that all appropriate arti

How much does arXiv cost to operate?

The annual budget for arXiv is \$400,000. With over 60,000 new submissions p of <\$7 per submission. Alternatively, with over 30,000,000 full-text downloads p download. We believe that arXiv is an extremely cost-effective service.

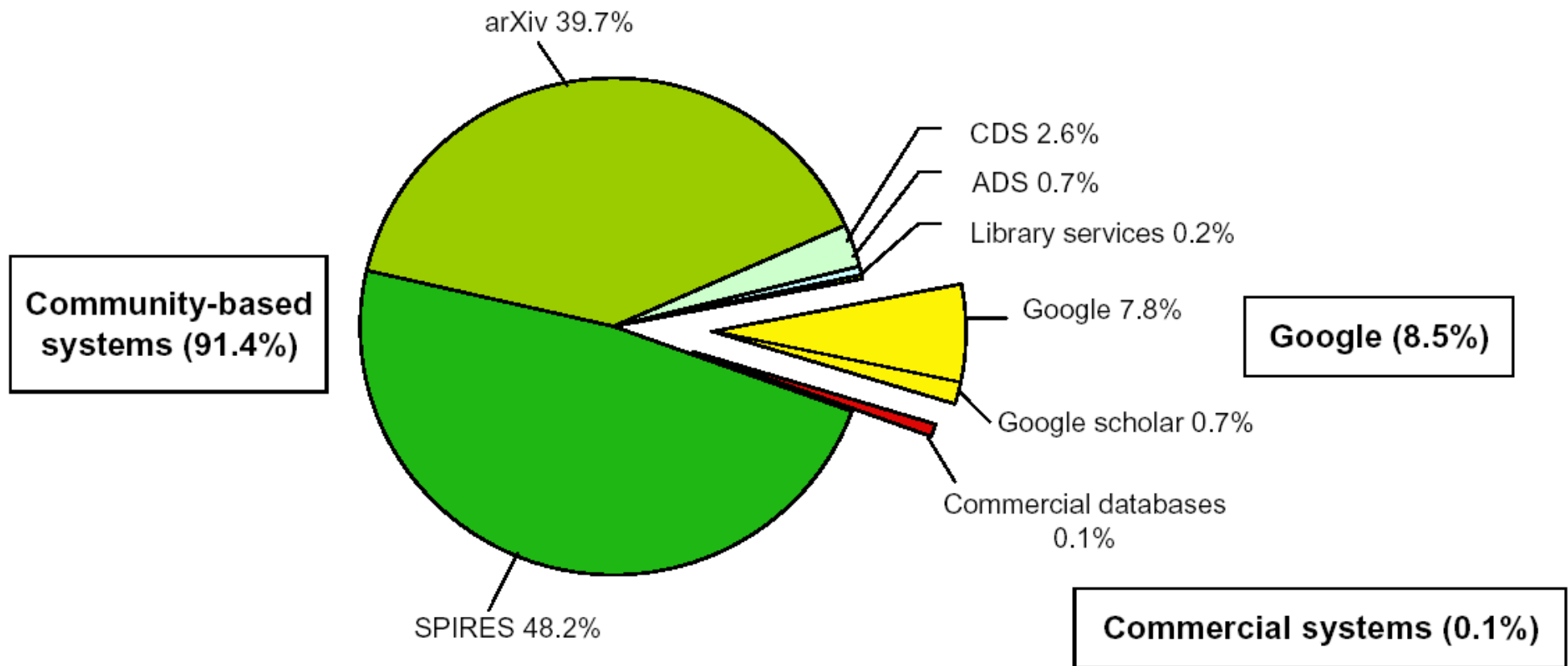
Who are you planning to ask for voluntary support?

We are planning to ask the top 200 heaviest user institutions for voluntary supp The top 200 institutions account for about 75% of institutional identification

Principles for supporting open access scholarly resources

- deep integration into a scholarly and/or scientific community and engagement by the scholarly community that demonstrate the value and importance of the resource
- commitment of the home institution to innovation
- transparency of and confidence in the operation of the resource
 - mandate
 - collection policies and submission guidelines
 - technology platform
 - end user support
- governance structure with clear roles
- technical stability of the delivery system and management mechanisms
- digital preservation strategies and proven track record in maintaining the system over time

Which HEP information system do you use the most?



Platform

- arXiv is 19 years old, predates most other repository systems, has “evolved”
- Many good decisions made early (e.g. URL structure, ids)
- Some specialized elements
 - Submission system and rapid moderation/announcement workflow
 - TeX engine
 - Interoperability with other systems (though most now using standard interfaces). Close data sharing with SPIRES (soon to be Inspire) and NASA ADS
- Bulk of access interface standard
 - Good candidate for replacement with other system
 - Hope to free effort in long run

Preservation

- currently arXiv content “*well managed*” but not preserved
- have been careful to accept reasonable formats
 - primarily: tex, pdf, docx
 - explicit click-through license
 - rare case of system that has already seen format transition (ps -> pdf)
- not that large (<1TB)
- content will go into Cornell University Library archival repository (Fedora based, being developed now)
- investigating other (external) options in parallel

Focus on the institution

- New task of managing institutional accounts
- Promised “*supported by*” banner
- Will add support for local OpenURL endpoints

Requests to help institutions show/demonstrate value:

- Download data relatively easy by IP
- Submission data difficult because we don't collect it. Have made estimates on request based on registration data and email domains – should we change submission policy/interface?
- Possibilities of new feature facilities related to author ids, claiming, connection to research information systems (e.g. VIVO, BibApp), SWORD deposit

e.g. Cornell use

=== Statistics for arXiv use from Cornell ===

730 registered users with Cornell email or claimed affiliation (0.5% of 154512 user accounts)

2360 articles owned by users with Cornell email or claimed affiliation (0.4% of 614489 documents)

2722 articles submitted by users with Cornell email (0.4% of 614489 documents)

=== Submissions in 2009 ===

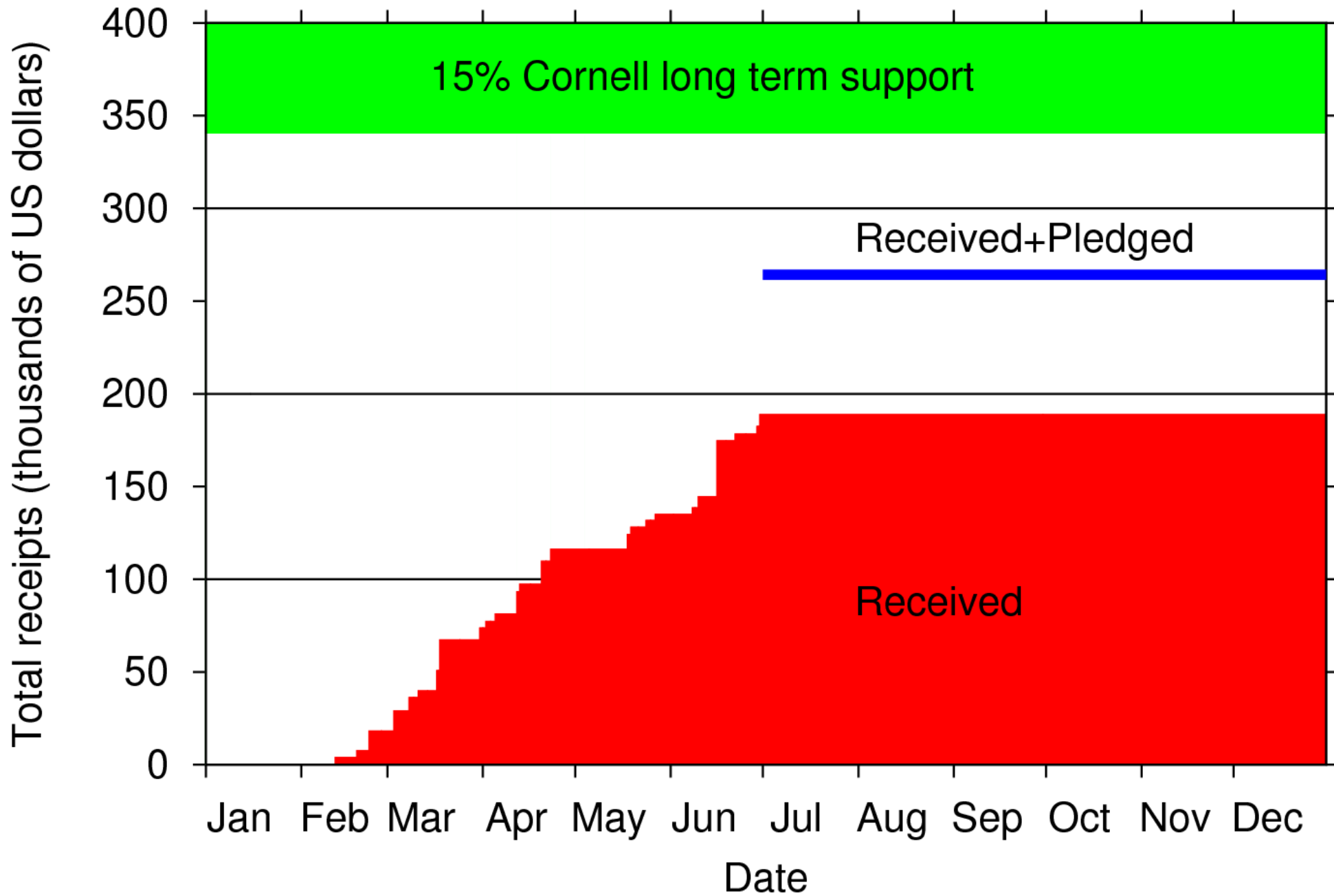
Found 223 articles in 2009 owned by users with Cornell email or claimed affiliation (0.3% of 64046 total articles in 2009)

Breakdown of articles submitted in 2009 by subject areas:

| | |
|----------------------|------------|
| Mathematics | 49 (22.0%) |
| Astrophysics | 43 (19.3%) |
| Cond. Matter Physics | 42 (18.8%) |
| High-Energy Physics | 35 (15.7%) |
| Computer Science | 29 (13.0%) |
| Other Physics | 22 (9.9%) |
| Quantitative Biology | 2 (0.9%) |
| Statistics | 1 (0.4%) |
| TOTAL | 223 |

After all that, how are we doing?

arXiv Support 2010 (to 2010-06-28)



Current work

- Form Sustainability Advisory Group in addition to Scientific Advisory Board
- Discuss broadening of support with publishers and societies
- Review pricing structure for 2011, address non-linearity
- Add facilities for account management and “*supported by*”
- Develop technology plan/roadmap
- Continue to work on long-term business model

Thanks to our supporters

Tier 1

Australian National University
Brookhaven National Laboratory
California Institute of Technology
CERN (Switzerland)
Columbia University
ETH Zurich (Switzerland)
Fermilab
Harvard University
Institute for Advanced Study
Johns Hopkins University
Los Alamos National Laboratory
Research Library
MIT Department of Physics and
MIT Libraries
Nagoya University
New York University
Ohio State University
Penn State University
Princeton University
Purdue University
Rutgers University
Stanford University
Texas A&M University
The University of Tokyo

Tohoku University

UC - Berkeley
UC - Davis
UC - Irvine
UC - Los Angeles
UC - San Diego
UC - Santa Barbara
University of Michigan
University of Pennsylvania
University of Toronto
University of Waterloo
University of Wisconsin
Tier 2
DESY (Deutsches Elektronen –
Synchrotron)
Duke University
Indiana University
Michigan State University
Rice University
Technion - Israel Institute of
Technology
The Institute of Mathematical
Sciences

Tokyo Institute of Technology

University of Minnesota
University of Tsukuba
Waseda University

Tier 3

Boston College
Macquarie University
UC – Merced
UC – Riverside
UC - Santa Cruz
University of Iowa
University of Oregon
University of Rochester
Washington State University

...and thank you
for listening!

Questions?

`simeon.warner@cornell.edu`

`rieger@cornell.edu`

`support@arxiv.org`

PS. We will soon be looking to hire a developer with repository and scholarly communication experience. Email me!

Interoperability: repository neighbors

