DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Potential Disciplinary of the NASA ADS Expansion: An evaluation of disciplines, communities, and costs of expansion

Citation
Willmott, Mathew. 2016. Potential Disciplinary of the NASA ADS Expansion: An evaluation of disciplines, communities, and costs of expansion. Harvard Library Report.

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:32969785

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Potential Disciplinary Expansion of the NASA ADS

An evaluation of disciplines, communities, and costs of expansion

Mathew Willmott
6-30-2016

Contents

I. Project Overview

The Smithsonian Astrophysical Observatory/NASA Astrophysics Data System is a digital library supporting researchers in astronomy, astrophysics, and currently to a lesser extent, physics and geophysics. The ADS has been in operation since the early 1990's, and indexes virtually the entirety of published astronomy and astrophysics material; in addition, the system contains full-text scans of most astronomy and astrophysics research articles prior to 1997. The ADS is currently considered an essential component of research in these fields; usage analysis shows that the system is used almost daily by the majority of the astronomy research community, and the collaborations the organization has forged with publishers, data providers, scholarly societies, and other online resources vastly increase the capabilities and efficiency of information-searching in the community. The model has proven successful for this community, and the Principal Investigator and Project Scientist at the ADS are now interested in expanding the scope of the system to include additional disciplines; their stated goal is to build a "PubMed for non-medical sciences" before publishers or commercial entities move into the space.[1] The Harvard Library is considering partnering with the ADS on this expansion.

Therefore, the purpose of this project is to investigate the costs, risks, and potential benefits of expanding the scope of the ADS, on behalf of the Harvard Library. The examinations and discussions herein are intended to assist library administrators in making an informed decision about whether to move forward with this partnership, and if so, what level of effort to expend, who to partner with, and how to approach the process. This project included several components to illuminate various facets of the potential expansion, including discussions with ADS staff to understand the inner workings and support structure of the database and what an expansion would entail; discipline analyses, including a limited number of researcher interviews or interactions, to estimate the size of any potential expansion and identify other discipline-specific characteristics; a literature search for studies examining the information-seeking needs and behaviors of researchers in relevant disciplines; and an exploration of potential collaborators if the Harvard Library were to spearhead the project.

[1] Telling quotes from Michael Kurtz, Project Scientist, about the importance he sees in partnering with the Harvard Library to apply the ADS model to other disciplines: "All that's at stake here is the future of how you run a digital library." There are only going to be a few players in the digital library world in the future; if Harvard isn't one, then they're a "failed library."

II. The ADS Currently

A thorough understanding of the current state of the ADS, as well as the reasons for the success it has enjoyed within the astronomy and astrophysics community, is important when considering how to apply the model to additional disciplines.

Size and scope

The database currently maintains over 11 million bibliographic records. The main focus of the database is to serve users in astronomy and astrophysics by indexing the vast majority of published research in those fields, including journal articles and conference proceedings as well as less common reference types such as data sets, observing proposals, and even source code. In general, the content of the ADS is guided by what is being cited in the astronomy and astrophysics literature; as the practice in citations change, so does the type of content that ADS needs to index, hence the move to indexing alternate content types described above.

However, the astronomy and astrophysics contribution to the database only accounts for about 2M of the 11M records in ADS; the remainder of the records are in physics. Astronomy and physics are very closely related fields, and so much of the astronomical literature cites research in physics. Rather than discerning the appropriate physics literature to index, the ADS determined it would be more time- and effort-efficient to index all physics literature by obtaining metadata directly from publishers. Additionally, ADS maintains records for all items posted on the arXiv preprint server, and continually merges duplicate records to ensure that each preprint record is combined with its corresponding final published version, upon publication. The yearly volume of publications in each area of ADS is shown in Table 1; approximately 442K records are added each year.

Table 1: Volume of publications in ADS

	Astronomy	Physics	arXiv	Total
Published 2010	66K	307K	69K	427K
Published 2011	61K	299K	76K	424K
Published 2012	64K	329K	83K	462K
Published 2013	54K	315K	91K	449K
Published 2014	57K	305K	96K	446K
Five-year average	**60K**	**311K**	**83K**	**442K**

Note: due to some overlap, the total number of records is not simply the sum of each area.

In addition to simple bibliographic records, the ADS also provides the full text of a large number of documents (~0.4M), including many early issues of astronomy and astrophysics journals from the 1800's up to 1997 which the ADS scanned to put online. The ADS also stores a substantial amount of publisher- or repository-provided full text (~4.4M records) without exposing it to the public; this information is used for indexing and metadata creation to improve search functionality. Additionally, the ADS maintains a network of citation links between articles in the database, indexes bibliographies curated by outside data centers, and provides integrated search tools linking documents with astrophysical objects.

Current funding model

At present, the ADS is funded entirely by NASA; the budget until recently has been approximately $2.3M, but was recently increased to $2.7M with a projection of $3M by the year 2020. About 90% of this funding goes towards the staff who manage the system and provide development and curation

efforts; the ADS leadership expects to employ close to 10 FTE once hiring is completed after this recent budget increase. Additionally, ADS pays a small amount to a collaborative effort in France which indexes astronomical objects. The remaining 10% of the budget goes towards infrastructure, including hardware and the costs of cloud computing.[2]

The current budget and employee distribution is broken into functional areas in the 2015 program review; a brief summary and analysis of these numbers will inform future cost estimates. Using the budget scenarios presented in the review (on p.30), several data points are calculated and presented in Table 2. First, staffing and budget data for FY16 is presented, and the average cost per FTE for each functional area is calculated. This average cost estimates a baseline rate which will be applied to any projected staffing increases. Second, the anticipated staff level for FY20 is presented. This is the target staffing level for supporting the ADS's current disciplinary coverage moving forward, and therefore the staffing level that estimates for disciplinary expansion will be based on. Finally, the budget for a fully-staffed ADS is calculated, estimating what it would cost to fully support the ADS at the FY16 cost per FTE. This calculation is necessary because the actual budget presented in the program review includes estimates of salary increases over the five years from FY16 to FY20. In this project report, however, the focus is on an immediate increase in staff to support expansion to a new discipline.

Table 2: ADS Budget Data[3]

	FY16 staff	FY16 budget	FY16 $/FTE	Fully staffed ADS (FY20 anticipated staff)	Fully staffed ADS budget: FY20 anticipated staff at FY16 rate
Staffing Subtotal	**10.69**	**$2475.5**	**$231.6**	**11.25**	**$2612.8**
Bibliography Support	3.53	$748.5	$212.0	3.95	$837.6
Bibliographic ingest	*2.13*	*$475.7*	*$223.3*	*2.39*	*$533.8*
Indexing/archiving/databases	*1.4*	*$272.8*	*$194.9*	*1.56*	*$304.0*
Development	4.25	$805.9	$189.6	4.25	$805.9
search functionality	*0.8*	*$173.4*	*$216.8*	*0.8*	*$173.4*
user tools	*0.68*	*$122.5*	*$180.1*	*0.68*	*$122.5*
software maintenance	*2.78*	*$510.0*	*$183.5*	*2.78*	*$510.0*
User Support	0.93	$319.9	$344.0	1.07	$368.1
Management	1.98	$601.2	$303.6	1.98	$601.2
Infrastructure Subtotal	**0**	**$255.7**	**N/A**	**0**	**$255.7**
Hardware and Licenses	0	$255.7	N/A	0	$255.7
GRAND TOTAL	**10.69**	**$2731.2**	**N/A**	**11.25**	**$2868.4**

Note: budget values are in thousands of US dollars.

Currently, because funding comes from NASA, the primary focus of the database must be in astronomy and astrophysics. For this reason, while they do index much of the physics literature, ADS staff lack the resources to pursue gray literature, or to support connections to data and the like. This illustrates the

[2] Murray, Accomazzi, and Kurtz, "Astrophysics Archives Programmatic Review 2015: The NASA Astrophysics Data System."

[3] FY16 and FY20 staff and budget data is calculated using the existing staff levels (Table I in the program review) plus the added staff allocations based on the augmented budget request (Table IV in the program review); the augmented budget request was approved, and so this properly estimates full staffing for those fiscal years.

need for partnership to achieve an expansion of scope; without additional resources, the in-depth work that makes the ADS successful in its field cannot be replicated in other disciplines.

Usage statistics

Usage of the ADS is quite high given the relatively small size of the core astronomy and astrophysics community. Based on analyses performed by Michael Kurtz in 2015, "roughly 50,000 scientists use the ADS almost daily including essentially every working astronomer, and a few tens of thousands of physicists and geophysicists. About 250,000 scientists use it to download a couple of articles per week or month, and a few million individuals use it occasionally, often via Google or Wikipedia links." Over the course of the year ending March 18, 2016, ADS saw about 1.5M queries through the classic search page, resulting in 2.43M abstract views and 750K full-text views.

Recent developments

The ADS is constantly developing to keep pace with the astronomy community. An important example of this development is the work the ADS does to follow the evolution of citation trends. As researchers began citing non-traditional types of references, the ADS was quick to add records for those types of materials to the database, including discipline-specific materials such as observing proposals as well as natively digital materials such as data tables and astrophysics source code. These added types thoroughly enrich search results for users, particularly those who employ citation-following as a strategy for exploring relevant literature, allowing them to more efficiently identify and access the resources used in the creation of a particular piece of content.

In addition to responding to trends from researchers, the ADS helps to drive innovation in various areas relating to astronomical information. For example, they are actively collaborating with organizations such as ORCiD, AAS, and Ringgold to develop standards for naming individuals and organizations, with the long term goal of implementing and promoting these naming standards to improve authority control in their database.

Insights on why the ADS is so successful

These explorations into the current status of the ADS have illuminated several key factors which have contributed to the success of the system within the community

Collaborations

The ADS has long employed extensive collaborations with many different types of organizations in order to enrich the services offered to users. Collaborations with other data providers allow for added features which provide large efficiency gains to users; for example, users can search for articles which reference a specific astronomical object from the SIMBAD database, they can restrict a search to only documents classified in the bibliography of a specific observatory, and they can follow links from bibliographic records directly to the article or the underlying data set. Collaborations with publishers allow for comprehensive coverage of the literature in a field, for higher-quality metadata, and for increased ability to digitize backfiles which would otherwise not be available online. Collaborations with academic societies help the ADS stay engaged in the development of trends in metadata and indexing, develop standards for the community, receive support (financial or otherwise) when necessary for various efforts, and publicize the system and its features to users. Collaborations with other online resources like arXiv and INSPIRE ensure interoperability and allow for combined efforts to save resources. Collaborations with libraries and librarians (especially at the Harvard-Smithsonian Center for

Astrophysics' Wolbach Library) offer opportunities for community engagement as well as volunteer work on content and bibliography curation.

These valuable collaborations have taken years of work to forge, and some were quite challenging to navigate (particularly with publishers). But they have borne very fruitful relationships that have helped the ADS to grow, improve, and maintain its position as a key research tool in its field.

Staff

Both the Principal Investigator and the Project Scientist at the ADS agree that one of the biggest keys to the ADS's continued success is its high-end staff, which they believe can compete at an intellectual level with staff at Google. ADS staff are highly efficient, highly skilled, and self-driven, with extensive subject knowledge that allows them to fully understand their own job and the role that ADS plays in the community at large. In comparisons with other databases, the ADS staff is generally smaller while maintaining a larger set of records.

Developing such a high-quality group of staff starts with recruiting. The ADS administration has placed a premium on recruiting staff with a deep subject understanding, often opting for employees with advanced degrees in astronomy or a related field who have taken on projects in information systems. This is in stark contrast to a model which would entail hiring staff with experience in information systems and teaching them the discipline and ADS's place in it. One staff member pointed out with pride that the ADS has never had a user group, because they already know what the users need. "Librarians could not have built ADS," quipped the Project Scientist.

In practice, this recruiting process often means hiring on the fly when the right person is available; of the two staffpeople interviewed for this project, one was hired after showing significant promise as an undergrad worker at the Center for Astrophysics, and the other was hired away from his previous job at a different database while he was in Cambridge to present his work to staff at the ADS. Additionally, the administration at the ADS provides the support and cultivation that these high-quality staff need to be successful. Whether this means job flexibility when needed or a voice in the future direction of the ADS, the staff feel respected, supported, and valued, and in turn they work hard to keep the system improving.

Internal culture

One of the great strengths of the ADS team is its agility; this gives the organization a great advantage over larger, bulkier development teams at other databases. As mentioned above, the staff that are hired are high-quality, and in turn they are given the independence to pursue projects that they believe are important. This may mean responding to user suggestions, triaging problems, and setting priorities. Staff don't wait around to be managed, and administration doesn't try to micro-manage; instead there seems to be an inherent trust that the staff know what they should be doing. Meetings are held on a regular basis to discuss what projects staff are working on, and to set and discuss priorities for the organization.

This culture allows the database to stay responsive and current at all times, creating a better relationship with the users that is reflected in the level of use the database sees. Maintaining such a culture if the organization were to grow significantly bigger would be a necessary challenge to overcome when expanding the scope of the database, as it is would be a major competitive advantage as compared to other systems.

Right place, right time

To an extent, the ADS was able to cement its position as a key research tool in the astronomy and astrophysics communities by virtue of being in the right place at the right time. The system's development in the early days of the internet before online resources were commonplace allowed it to seize upon several opportunities.

For example, in scanning available backfiles of important journals and distributing the content freely, ADS became the central resource for access to that material. Additionally, through negotiations, the ADS was able to shape publisher attitudes towards backfiles; they allowed the ADS to focus on historical digitization so that they could focus on converting current content to digital formats, and researchers benefited by getting free access to content from their offices. In the current environment, the ADS would not enjoy these benefits when expanding to a new discipline: almost all major journal backfiles are already digitized, and many require a subscription or one-time purchase for access.

Another example is the relative lack of competition for a bibliographic database encountered in the early days of the digital era. For astronomers, the ADS was likely the only option for an online subject-specific database when it was made available; they adopted it and have since built it into their research process. In the present day, researchers in the disciplines under consideration for expansion have a glut of database options to choose from, many of which are simpler (such as Google Scholar with a single search box) or which researchers may be more familiar with. Convincing researchers to switch tools is more difficult than convincing them to start using a new tool with huge efficiency gains.

These examples are not intended to imply that there is no place in these markets for a system like the ADS. But these footholds that the ADS was able to use early in its development may not be sufficient for encouraging adoption by researchers in other fields today. Instead, the ADS (along with the Harvard Library) will need to identify unmet needs or significant efficiency gains to offer potential users.

Researcher-driven

Perhaps as a result of many of the factors discussed above, the ADS has a very engaged user community; they offer suggestions, they help identify duplicate records, and perhaps most importantly, they advocate for use of the database to their colleagues and their students. Chris Stubbs, a physics faculty member who is a faithful user of the ADS, is effusive in his praise of the system, describing a "tremendous efficiency gain" and even going so far as to say that taxpayers and granting agencies win when paying for ADS because it makes his work more efficient. Prof. Stubbs teaches an undergraduate class on observation, and always devotes an entire session to teaching his students how to use the ADS.

Developing such an evangelistic following in other disciplines will be key to the database's adoption; potential users are much more likely to use a system recommended by a colleague or teacher. The ADS will need to work on identifying influential advocates in the fields being targeted for expansion to help with encouraging new users to try it.

III. Evaluation of Related Disciplines

The disciplines under consideration for the initial expansion in scope of the ADS are physics, mathematics, and earth sciences. These are all hard sciences which have a reasonable disciplinary connection to astronomy and astrophysics, where the ADS focuses its coverage currently. In this section, we observe various characteristics about each of these disciplines through analyses, literature reviews, targeted investigations, and in some cases, interactions with community members. The observations herein will provide important points of comparison and contrast, allowing for predictions about how the existing operations of the ADS could scale to other disciplines. Additionally, much of the information gathered can assist in predicting how an expansion in scope would be received by researchers in these disciplines, as well as identifying areas of current struggle where development efforts at the ADS could make a significant impact on the research community.

In general, the literature searches discussed below are intended to find surveys of user needs and behavior in these disciplines, thereby illuminating the similarities or differences in behavior which could affect uptake of the ADS in a new discipline or which could identify community-specific needs for the ADS to meet. Many of the user behavior studies published in the last twenty years were designed to track the transition from traditional print resources to online resources; depending on the research behaviors investigated these may not have valuable data for the purposes of this project. Moreover, many of those studies do not present data specific to individual science fields, even when available, because sample sizes are small; a 2007 study at the University of North Carolina at Chapel Hill, for example, recognizes the differences observed between subfields in the sciences in other studies, but still combines subfields into two categories (basic and medical science) for analysis purposes[4]. Nevertheless, when these studies do present data at the individual field level, the information can be quite valuable for our purposes.

Astronomy and Astrophysics

Size and Scope

The ADS is the primary abstracting and indexing database for astronomy and astrophysics, and is therefore the best measure of the volume of publication in the discipline. However, Web of Science also indexes astronomy and astrophysics literature, and can provide a valuable comparison point for other disciplines. This analysis includes all documents in Web of Science with the subject category Astronomy & Astrophysics. Additionally, a large portion of the published literature in the field appears first in preprint form in the arXiv's astro-ph subject area, offering a third point of comparison for the volume of literature in the field[5].

Table 3: Astronomy and Astrophysics Publishing Volume

	ADS	Web of Science	arXiv
Total documents	2,179,278	719,433	178,946 (as of 2015)
Total documents, 2010-2014	300,413	119,554	60,752
Published 2010	65,689	24,707	11,616
Published 2011	61,182	24,067	11,954

[4] Hemminger et al., "Information Seeking Behavior of Academic Scientists."
[5] "arXiv.org E-Print Archive."

Published 2012	63,749	24,731	12,122
Published 2013	54,010	22,388	12,475
Published 2014	55,783	23,661	12,585

Note: data current to April-June 2016, except where otherwise noted.

A more direct comparison between ADS and Web of Science can be accomplished by breaking the yearly ADS totals into refereed and non-refereed publications, and the Web of Science totals into Articles, Proceedings, and other document types. In this case, refereed articles in ADS are roughly comparable to documents of type Article in Web of Science, non-refereed articles are roughly comparable to documents of type Proceedings, and the remainders are roughly equivalent.

Table 4: Astronomy and Astrophysics ADS and WoS volume

	ADS			Web of Science		
	Refereed articles	Non-refereed articles	Other (refereed and non-refereed)	Article	Proceedings	Other
2010	24,022	21,979	19,688	17,120	6,612	1,328
2011	26,571	14,805	19,806	18,674	4,731	896
2012	25,173	18,355	20,221	18,872	4,836	1,068
2013	23,996	12,267	17,747	18,751	2,802	877
2014	23,425	16,925	15,433	19,421	3,562	784
Total	**123,187**	**84,331**	**92,895**	**92,838**	**22,543**	**4,953**

Note: data current to April-June 2016, except where otherwise noted.
Also note: Web of Science has some overlap between Article and Proceedings (i.e. some documents are classified as both)

These tables demonstrate where ADS's coverage of the literature is greater than that of Web of Science. The difference in coverage is moderate for standard, refereed journal literature, although Web of Science appears to be closing the gap in more recent years (in 2014, WoS's coverage was up to 83% that of ADS). However, for coverage of alternate sources such as conferences and other materials, ADS has an enormous advantage over Web of Science. It is reasonable to infer, particularly given the information discussed in section II, that this expansive coverage is the result of deep collaborations with information providers in astronomy and astrophysics, of the ADS's position as the preeminent database in the field leading to holders of gray literature seeking out the ADS for digitization or indexing, and of the ADS's proactive work in expanding the scope of what can be indexed. We can therefore anticipate that an expansion to other disciplines could potentially result in similar coverage patterns, especially over the long term as the ADS fosters relationships and collaborations within each discipline and develops its standing among researchers, libraries, and information providers.

Information-Seeking Behavior

Astronomers were some of the earliest adopters of online journals, due in part to the availability of the ADS and its integration with the American Astronomical Society's electronic journal system; the standards used in this collaboration were quickly adopted by other astronomy publishers. Additionally, the early integration of astronomical data sources such as CDS, NED, and SIMBAD helped to advance the adoption of online resources in astronomy[6]. As Tenopir et al also point out, astronomy has several specific characteristics which facilitated the early adoption of online resources: virtually all research is

[6] Tenopir et al., "Relying on Electronic Journals."

published in English; citations are relatively compact within the field, with a large percentage of citations pointing to a small number of journals; and scientists have a strong desire for rapid publication through preprints. The latter of these points is of note both because it is a desire shared by physicists, and because ADS maintains a collaboration with the two main preprint servers in these disciplines, arXiv.org and INSPIRE.

Finally, Tenopir et al surveyed members of the American Astronomical Society and compared the results with previous studies of information-seeking behavior in scientists. In some cases, behavior patterns were found to be comparable between astronomers and other scientists, including currency of articles being read, amount of time spent reading per article, and percentage of articles read which contain information unknown by the readers. However, other cases offer more distinctions: astronomers read more articles for their primary research and fewer for current awareness than scientists in other fields. Astronomers are also more likely than other scientists to identify articles they want to read through online searching and less likely than other scientists to identify articles they want to read by browsing journals (either electronic or print), although this distinction may be related to the availability of quality online searching from the ADS at the time of the survey (2002).

Interdisciplinary Profile

As Tenopir references, one remarkable characteristic of the field of astronomy is its compactness. In fact, multiple studies have observed citation patterns in astronomy and found it to be perhaps the most compact field in all of the sciences. A 1998 citation analysis performed by Claspy observed citation patterns in reference lists of several major astronomical journals, finding that "92% of the references, regardless of type of publication, were to some type of astronomical literature."[7] More general studies can put this compactness into perspective; a 2000 survey of citation data from ISI's Journal Citation Reports database found rates of cross-disciplinary citations for each of 119 disciplines.[8] In this study, Astronomy and Astrophysics had, by far, the lowest rate of cross-disciplinary citations, at 14.3%; overall rate across the study was 69%. Finally, the National Science Foundation's 2000 Science and Engineering Indicators report identifies the disciplinary distribution of cited articles in a variety of broad and fine fields, finding that 89.8% of references in Astronomy and Astrophysics articles were within the field (the highest percentage within a "fine" field by almost 10 percentage points).[9] By Claspy's survey, Physics is the second-most cited field by astronomers, garnering 4.7% of citations. The NSF study, on the other hand, finds Biomedical Research to be the second-most cited field, with 3.8% of citations (Physics is third with 2.7%).

Taken on the whole, these studies create a picture of a community of scientists who rarely need to explore published research outside of their discipline. This characteristic of compactness has likely contributed to the success of the ADS as a contained, single-subject digital library; it could also create a challenge in expansion to fields which require more interdisciplinary research.

Disciplinary Observations

Compared to other scientific communities (as demonstrated in the next several sections), the community of researchers in astronomy and astrophysics is quite small. The primary global professional

[7] Claspy, "Information Use in Astronomy."
[8] van Leeuwen and Tijssen, "Interdisciplinary Dynamics of Modern Science."
[9] National Science Foundation, "Science and Engineering Indicators: 2000."

organization is the International Astronomical Union, currently comprised of approximately 12,500 members worldwide (of which approximately 22% are in the U.S.).[10] The American Astronomical Society, the major society in North America, is comprised of 7,000 members, which includes scientists in tangential disciplines such as mathematics, physics, and earth science.[11] It is also notable that societies like the IAU and AAS and research organizations play a large role in publishing in the field. According to Thomson Reuters' Journal Citation Reports, the eight most-cited journals in the field account for over 80% of all citations to Astronomy and Astrophysics journals; only one of these journals (Physics Letters B) is published by a commercial publisher. The others, seen in Table 5, are published by or on behalf of scholarly societies or research organizations.

Table 5: Most-cited journals in Astronomy and Astrophysics[12]

Title	Publisher	Percent share of citations to astro journals
Astrophysical Journal	Institute of Physics, on behalf of AAS	22.1%
Physical Review D	American Physical Society	16.1%
Monthly Notices of the Royal Astronomical Society	Oxford University Press, on behalf of the Royal Astronomical Society	13.1%
Astronomy & Astrophysics	EDP Sciences, on behalf of the European Southern Observatory	11.7%
Physics Letters B	Elsevier	6.5%
Astrophysical Journal Letters	Institute of Physics, on behalf of AAS	5.0%
Astronomical Journal	Institute of Physics, on behalf of AAS	3.7%
Astrophysical Journal Supplement Series	Institute of Physics, on behalf of AAS	2.7%

Finally, it is worth noting the prevalence of linked data sources within the field. Of the approximately 123,000 refereed articles indexed in astronomy in ADS from 2010 to 2014:

- 17,000 (14%) have links to data sets used in the article
- 35,000 (28%) have links to astrophysical objects through either SIMBAD or NED
- 23,000 (19%) are included in at least one of the 29 curated bibliography lists from observatories or research groups
- 60,000 (49%) have links back to their arXiv e-print
- 109,000 (89%) include reference lists
- 102,000 (83%) include lists of citing documents

[10] International Astronomical Union, "Geographical Distribution of Individual Members."
[11] American Astronomical Society, "What Is the AAS?"
[12] Thomson Reuters, "2015 Journal Citation Reports®."

While reference and citing document lists are common across all fields, the prevalence of the other four objects in the astronomy research paints a picture of a field that, as a whole, relies on interconnected data and information, much of which is unique to this particular field.

Physics

Size and Scope

By all measures, physics is a larger discipline than astronomy and astrophysics. Web of Science offers thorough coverage of the field, and can therefore provide a good assessment of the volume of publication. Other interdisciplinary databases, such as Scopus and Google Scholar, are also commonly used to search the literature; Web of Science is used for this analysis because Scopus is not licensed at Harvard, and Google Scholar does not allow the type of fielded or subject-based searching that would make this analysis reliable. Additionally, Inspec, produced by the Institution of Engineering and Technology and distributed through multiple platforms (for Harvard, available through EBSCOhost), is the most commonly-used subject-specific database, thoroughly indexing the literature in physics, electrical engineering, and computer science. As described above, ADS also indexes physics literature in order to provide thorough coverage of the references that astronomers cite in their work. Finally, we can observe the volume of literature distributed through the arXiv, as a large portion of physics publications are posted there before being published in a journal.

The volume of physics publication in each of these sources is displayed in Table 6. For Web of Science, this includes articles in the subject categories of Biophysics, Optics, and all subcategories of Physics (Applied; Atomic, Molecular & Chemical; Condensed Matter; Fluids & Plasmas; Mathematical; Multidisciplinary; Nuclear; and Particles & Fields)[13]. For Inspec, this includes all publications indexed in the Physics Section (Section A). For arXiv, this includes cond-mat, gr-qc, all hep subjects, math-ph, nucl-ex, nucl-th, physics, and quant-ph.

Table 6: Physics Publishing Volume

	ADS	Web of Science	Inspec	arXiv
Total documents[14]	8,042,984	6,026,014	9,145,076	579,786 (as of 2015)
Total documents, 2010-2014	1,541,973	1,066,933	1,662,864	188,591
Published 2010	303,697	201,928	308,268	33,549
Published 2011	296,297	206,867	325,605	36,251
Published 2012	326,963	214,015	343,222	38,014
Published 2013	312,968	222,646	359,119	39,823
Published 2014	302,048	221,477	326,650	40,954

Note: data current to April-June 2016, except where otherwise noted.

[13] Note that because Web of Science can assign multiple categories to a single journal, there can be overlap between subjects. Physical Review D, for example, is included both in the Astronomy & Astrophysics counts and the Physics counts.

[14] Total for arXiv represents their 2015 cumulative submission count. The submission count for math and math-ph is provided together; their separate shares are estimated based on the ratio of documents in arXiv in each subject for 2010-14; only math-ph is included here.

A comparison of the conference literature indexed in these databases can give a generalized view of the profile of publications in physics, helping to identify whether there are additional areas which ADS would need to index to provide full coverage of the literature. It is important to remember that non-refereed articles in the ADS represent both conference publications and other formats (including monographs, etc.).

Table 7: Conference literature in physics databases

	ADS		Web of Science		Inspec	
	Total documents	Non-refereed articles	Total documents	Proceedings	Total documents	Conference articles
2010	303,697	68,229 (22.5%)	201,928	55,832 (27.6%)	308,268	62,969 (20.4%)
2011	296,297	38,235 (12.9%)	206,867	46,215 (22.3%)	325,605	61,388 (18.9%)
2012	326,963	51,287 (15.7%)	214,015	50,402 (23.6%)	343,222	62,826 (18.3%)
2013	312,968	45,571 (14.6%)	222,646	47,938 (21.5%)	359,119	61,527 (17.1%)
2014	302,048	44,492 (14.7%)	221,477	42,401 (19.1%)	326,650	50,602 (15.5%)
Total	1,541,973	247,814 (16.1%)	1,066,933	242,788 (22.8%)	1,662,864	299,312 (18.0%)

In both tables, Inspec has a slightly higher publication volume in physics; Inspec contains about 121K more total physics documents than ADS for the years 2010 to 2014, representing an 8% overall increase in physics coverage over the ADS. At least 51K of these documents are conference publications, representing a 21% increase in coverage over the ADS. This is not unexpected: conference publications outside of major publishers are more elusive to locate for indexing, and to date, ADS has not been focusing significant effort on indexing in physics.

An overall conclusion from these tables is that while ADS currently provides close to comprehensive coverage in physics, a concerted expansion of scope into physics would still require a moderate push to expand this coverage, especially pertaining to conference literature. Because ADS already has agreements with major physics publishers, including SPIE and AIP for physics conference proceedings, it is likely that much of this additional literature will be more difficult to locate, requiring collaborations or agreements with smaller publishers or organizations.

Information-Seeking Behavior

A survey of physicists and astronomers in 2008 helps to distinguish between some of the needs of each community separately[15]. While the sample size for the survey is rather small and mostly PhD students rather than faculty or more mature researchers, we can still observe some overall trends. Notably, the aforementioned preference of astronomers towards online searching for articles rather than browsing is not necessarily shared by physicists; physicists are more likely to browse journals or to track references to find their research.

Interdisciplinary Profile

Physics is significantly more interdisciplinary than astronomy. Van Leeuwen's study of cross-disciplinary citations[16] relied on Web of Science subject categories for its subject-specific analysis; the paper only reported the top and bottom 15 cross-disciplinary citation rates by subject, and physics subjects were

[15] Jamali and Nicholas, "Information-Seeking Behaviour of Physicists and Astronomers."
[16] van Leeuwen and Tijssen, "Interdisciplinary Dynamics of Modern Science."

observed in each category: Biophysics had the ninth-highest CDC rate, at 92.4%, while Physics-General; Physics-Atomic, Molecular, Chemical; and Optics rated in the bottom 15, with CDC rates of 59.9%, 59.4%, and 58.4%, respectively. This spread implies that physics in general is much closer to the average CDC rate of 69% across the study, which is significantly greater than that of astronomy and astrophysics. However, it should be noted that one challenge of interpreting the van Leeuwen study in this context of this project is that if an article in one physics subfield cites an article in another physics subfield, it counts as a CDC. However, if the ADS were to cover all areas of physics, both the citing and cited articles would be included in the database.

The NSF's Science and Engineering Indicators, on the other hand, offer a more generalized view of interdisciplinarity in citation behavior at a broader level.[17] According to the NSF study, 81.7% of references in the broad field of Physics articles stayed within the field. Outside of Physics, significant percentages of references pointed towards articles in the broad fields of Chemistry (6.9%), Biomedical Research (4.3%), and Engineering & Technology (3.8%). For Chemistry and Engineering & Technology, these patterns were reciprocated: 11.1% of Chemistry references and 21.3% of Engineering & Technology references pointed towards Physics articles. Additionally, 8.2% of Mathematics references pointed towards Physics articles, even though less than one percent of Physics references pointed towards Mathematics articles. Furthermore, the fine field of Biophysics is categorized by the NSF under Biomedical Research; 75.5% of references in this fine field point to other articles in Biomedical Research; only 16.7% of references stay within Biophysics, and so the remaining 58.8% of references point to other fine fields within the broad field of Biomedical Research, rather than Physics.

On the whole, Physics is significantly more interdisciplinary than Astronomy and Astrophysics, including a higher percentage of references pointing outside of the field as well as a greater number of citations from other fields pointing into the field. This information must be taken into account when considering the effectiveness of some of the ADS's more prominent features, reference lists and citation counts: if the ADS were to expand into physics, either these lists would be incomplete because they would point to material outside of the scope of the database, or the ADS would need to perform a cursory indexing of related fields like Chemistry and Engineering (similar to the way Physics is currently indexed to enrich the reference and citation lists for Astronomy).

Disciplinary Observations

Several professional societies exist in physics, generally representing various geographical areas. Major societies or society amalgamations include the American Physical Society in the U.S. (53K members[18]), the Institute of Physics in the U.K. and Ireland (50K members[19]), the European Physical Society (120K individual members across their member societies[20]), and the Association of Asia Pacific Physical Societies (includes the Physical Society of Japan, 17K members[21]; the Japan Society of Applied Physics, 23K members[22]; and the Chinese Physical Society, 40K members[23]). While there may be some overlap in membership, these listed societies alone represent a community of over 300,000 physicists. In addition,

[17] National Science Foundation, "Science and Engineering Indicators: 2000."
[18] https://www.aps.org/about/governance/annual-reports/upload/annrep2015.pdf
[19] http://www.iop.org/about/index.html
[20] http://www.eps.org/?page=membership_ms
[21] http://www.jps.or.jp/english/concept.html
[22] https://www.jsap.or.jp/english/aboutus/members-organizations.html
[23] http://www.cps-net.org.cn/english.htm

the American Institute of Physics consists of many smaller, more specialized member societies, including the Acoustical Society of America, the American Association of Physicists in Medicine, the American Association of Physics Teachers, the Optical Society of America, and the Society of Rheology.

In addition to representing the scientific community towards which an expansion in the scope of the ADS would be targeted, most, if not all, of the societies above publish discipline-specific journals which contribute to the scientific literature in physics. Some are published and distributed by the society itself, such as the APS journals, while others are published by a parent organization such as the AIP, or a commercial publisher.

Because of the granularity and disparate nature of the professional physical societies, as well as the wider range of research taking place in physics in general, the distribution of citations among journals is much greater than for astronomy and astrophysics; the top 25 journals account for only 50% of all citations in the field, according to Journal Citation Reports. Additionally, while 10 of the top 11 journals by citation count are published by scholarly societies, half of the remaining titles in the top 25 are published by commercial publishers, indicating that the commercial presence in the field is greater than that of astronomy.

Many areas of physics—experimental and applied physics fields in particular—rely on various data sources, including both curated datasets and experimentally verified reference data. Notable examples include the National Nuclear Data Center[24], maintained by Brookhaven National Laboratory, and the National Institute of Standards and Technology's Physical Reference Data[25] resource, which includes links to many of the databases and other data products maintained by NIST. Linkage of the literature to these data sources, either as a citation or as a linked data feature (similar to the existing connection between SIMBAD and the ADS) could greatly improve search functionality. Additionally, there are numerous academic and governmental research laboratories and organizations that would benefit from embedding their work into a literature search tool like the ADS, through bibliography curation as well as through indexing of various types of gray literature, including laboratory reports, publicly available datasets, and experiment proposals and results.

Interactions with physicists in various subfields showed differing levels of interest in expanding the scope of ADS into physics. Christopher Stubbs, a professor in physics and astronomy at Harvard, is an avid user of the system, accessing it weekly—if not daily—to search for a known author or paper, to prepare bibliographies for his own work, and to search the literature by following citations. He sometimes uses Google Scholar when he is trying to search for items out of scope or on a broader scale; he notes that Google Scholar is generally less effective, with a higher noise-to-signal ratio. He stresses that the ADS is key to his research, stating that taxpayers and granting agencies win because his work is more efficient. Stubbs, however, confesses that his research patterns are likely more in line with astronomers than with physicists.

Physicists who are less familiar with the system had other opinions. David Nelson, a faculty member working in theoretical physics and physical biology, expressed disinterest in adding to the suite of research resources available to physicists, stating that his research needs are well-met by a combination of Harvard's journal subscriptions, Google Scholar, and (most importantly) the arXiv. Nelson posits that

[24] http://www.nndc.bnl.gov/
[25] http://www.nist.gov/pml/data/index.cfm

"at a time of budgetary constraints and burgeoning bureaucracy here, I suspect there are better ways to spend our resources." While some of these statements may be debatable (especially considering the shortcomings of Google Scholar, particularly those mentioned previously by Prof. Stubbs; and the incomplete coverage of the literature provided by arXiv and evidenced in Table 6), the sentiment may be a common one, that the existing resources are sufficient. Possibly echoing this sentiment, no response was ever received from the other faculty contact recommended by Harvard's Physics Librarian.

Mathematics

Size and Scope

For Mathematics, ADS does not currently offer any coverage of the literature. We estimate the volume of publication output in the field, again using data from Web of Science. Additionally, for comparison, we observe the publication volume in MathSciNet, an abstracting and indexing database run by the American Mathematical Society (discussed in further depth below). MathSciNet is available for subscription directly from the AMS, and is the primary subject-specific database in the field. Finally, counts from arXiv's "math" subject are presented for comparison; as the ADS does index all arXiv preprints (although in the case of mathematics, where the journals themselves are not indexed, ADS generally does not reconcile the preprint with the final published article).

Table 8 summarizes the current publishing volume in the field of mathematics. For Web of Science, this includes articles in the subject categories of Logic; Mathematical & Computational Biology; Mathematics; Mathematics, Applied; Mathematics, Interdisciplinary Applications; or Statistics & Probability.

Table 8: Mathematics Publishing Volume

	Web of Science			MathSciNet			arXiv
	Article	Proceedings	Total	Article	Proceedings	Total	Total
Total documents[26]	2,485,255	323,526	3,401,401	2,754,855	394,047	3,250,898	221,327
Total documents, 2010-2014	515,834	57,798	627,943	467,069	48,070	523,451	119,656
Published 2010	79,650	15,438	103,498	87,732	9,628	98,972	18,765
Published 2011	90,124	10,234	109,653	90,788	9,816	102,344	21,287
Published 2012	103,363	10,307	124,250	96,311	9,447	107,298	24,176
Published 2013	119,121	11,897	143,117	98,559	9,923	110,294	26,785
Published 2014	123,576	9,922	147,425	93,679	9,256	104,543	28,643

[26] Total for arXiv represents their 2015 cumulative submission count. In this case the submission count is given for math and math-ph together; math-ph count is estimated based on the ratio of documents for 2010-14, and removed from this total.

Note: data current to April-June 2016, except where otherwise noted.
Also note: Web of Science has some overlap between Article and Proceedings (i.e. some documents are classified as both), as well as many documents which are not classified as either.

Both Web of Science and MathSciNet display a similar yearly publication volume and total historical corpus size. Additionally, the volume of conference proceedings indexed by each database is similar. Overall, the size of the mathematics literature produced yearly is approximately twice that of astronomy and astrophysics, although the proportion of conference proceedings represented is notably less, approximately 10% in mathematics as compared to upwards of 20% in astronomy. It is also worth noting that of the documents which are not classified as either articles or proceedings in Table 6, many are monographs, a format that the literature review below shows is more important in mathematics than in physics or astronomy. If the ADS were to expand into mathematics, it would need to be thorough in indexing monographs.

Information-Seeking Behavior

A 1999 survey at the University of Oklahoma compared information-seeking behavior between scientists in different disciplines, including physics and astronomy (combined into one category) and mathematics, finding several notable contrasts between the two fields[27]. Mathematicians are much more likely than physicists or astronomers to support their research activities using monographs (85% of mathematicians vs. 53% of physicists/astronomers), conference attendance (92% vs. 60%), personal communications (97% vs. 33%), and preprints (92% vs. 67%). Similarly, Brown finds contrasts in staying abreast of current developments in a field: mathematicians are more likely than physicists/astronomers to scan current issues of journals (91% vs. 69%) or rely on personal communications (85% vs. 62%). Some of these differences lend themselves to existing features within the ADS, and others present opportunities for development. ADS has extensive experience in handling preprints, including an ongoing relationship with arXiv and merging duplicate records; this existing system would benefit mathematicians. However, an innovative opportunity may exist in the concept of indexing or even storing personal communications used in mathematics research, particularly those which are cited in the literature, similar to the way the ADS has recently begun indexing additional sources in astronomy (e.g. observing proposals, software packages, etc.).

Interdisciplinary Profile

The field of mathematics shows some similarities to astronomy and astrophysics in terms of its compactness of citations. In van Leeuwen's study of cross-disciplinary citations, the basic field of Mathematics is found to have the second-lowest percentage of cross-disciplinary citations, 26.1%, higher than only Astronomy & Astrophysics. While there is also a "Mathematics-miscellaneous" field in the top range, with a CDC rate of 94.1%, this alternate discipline encompasses only five journals, while the main field contains 104 journals.[28]

The NSF Science & Engineering Indicators are able to give a more nuanced view of citation behavior. In the broad field of Mathematics, 77.4% of references point to other Mathematics articles, while 8.2% point to Physics and 7.5% point to Engineering & Technology. This interdisciplinarity is mostly observed in the fine field of Applied Mathematics, where only 62.1% of references point to other Mathematics articles; in General Mathematics, 89.8% of references point to other Mathematics articles. Additionally,

[27] Brown, "Information Seeking Behavior of Scientists in the Electronic Information Age."
[28] van Leeuwen and Tijssen, "Interdisciplinary Dynamics of Modern Science."

it is worth noting that the volume of citations in Mathematics is significantly less than that of other fields: NSF records about 5,000 citations from all articles in the broad field of Mathematics in 1997, as compared to about 24,000 in the fine field of Astronomy & Astrophysics, and over 120,000 in the broad field of Physics. In addition to a much lower volume of references to track, this phenomenon also means that while less than 0.4% of all Physics references and 1.7% of all Earth Science references point to an article in Mathematics, each subject accounts for almost 10% of all citations to Mathematics articles.

Disciplinary observations

In the U.S., the American Mathematical Society is the main professional society for mathematicians, with approximately 30,000 members[29]. Other societies include the American Statistical Association (18K members)[30] and the Society for Industrial and Applied Mathematics (14K members).[31]

The AMS also produces MathSciNet, the primary subject-specific database in the field. MathSciNet, in addition to indexing the mathematical literature, also curates article reviews written by mathematicians; these reviews are linked to the documents they discuss and are indexed and searchable. These reviews could be interpreted as a serialized form of author-to-author communication along the lines of the personal communication described by Brown above. Until 1996, these reviews were distributed in journal form under the name Mathematical Reviews. The database also includes some discipline-specific features, including the Mathematics Subject Classification and easily viewable typeset mathematics. However, MathSciNet has some drawbacks as well, not the least of which is the lack of e-mail alerts or RSS feeds for current awareness[32]. Brown found in her survey that 31% of mathematics faculty used a current awareness service even in 1999; it is likely that demand for these features is even higher now.

The central international body representing mathematicians around the world is the International Mathematical Union.[33] The IMU does not have individual members, but rather represents the joint interests of member countries in various endeavors, including education and fostering the growth of mathematics communities in the developing world. Interestingly, the IMU has, in recent years, begun the process of studying what a Global Digital Mathematical Library would look like, supporting research by the National Academy of Sciences and the National Research Council into this topic.[34] This study has led to recommendations for features and development which are closely aligned with the existing infrastructure of the ADS, including the following findings and recommendations:

- "A primary role of the Digital Mathematics Library should be to provide a platform that engages the mathematical community in enriching the library's knowledge base and identifies connections in the data."
- "While fully automated recognition of mathematical concepts and ideas (e.g., theorems, proofs, sequences, groups) is not yet possible, significant benefit can be realized by utilizing existing

[29] http://www.ams.org/membership/membership
[30] http://www.amstat.org/about/asamembers.cfm
[31] https://www.siam.org/membership/
[32] Mounts, "MathSciNet."
[33] http://www.mathunion.org/
[34] Committee on Planning a Global Library of the Mathematical Sciences, "Developing a 21st Century Global Library for Mathematics Research."

scalable methods and algorithms to assist human agents in identifying important mathematical concepts contained in the research literature—even while fully automated recognition remains something to aspire to."

- "The Digital Mathematics Library should rely on citation indexing, community sourcing, and a combination of other computationally based methods for linking among articles, concepts, authors, and so on."

These findings imply that the mathematics community as a whole would realize great benefit from a digital library working to encapsulate literature and knowledge within their field into one system in the way that the ADS has for astronomy. That said, when contacted to try to set up a discussion on some of these topics, the editor of Mathematical Reviews, who oversees MathSciNet, did not respond.

The distribution of citations among mathematical journals is even greater than that of physics; according to Journal Citation Reports, the top 50 journals (by total citations) still only account for less than half of all citations in the field. Additionally, many of the top journals are interdisciplinary journals representing the intersection of mathematics and other fields: Bioinformatics, Econometrica, and Computer Methods in Applied Mechanics and Engineering all rank in the top five mathematics journals.

Earth Science

Size and Scope

Earth Science comprises a much broader and more loosely-defined set of subfields than the subjects discussed above. For the purposes of this report, the scope of the field was defined by the subfields listed for Harvard's Department of Earth and Planetary Sciences: Climate, Atmosphere, and Oceans; Earthquake Science and Active Tectonics; Earth's Interior and Surface; Geobiology and Earth History; and Planetary Science and Cosmochemistry. The volume of publication output in these fields is again measured using Web of Science and a primary subject-specific database in the field, in this case, GeoRef, which covers the majority of the aforementioned subfields.

Table 9 summarizes the current publishing volume in the field of earth science. For Web of Science, this includes documents in the subject categories of Geochemistry & Geophysics; Geology; Geography, Physical; Geosciences, Multidisciplinary; Meteorology & Atmospheric Sciences; and Oceanography.

Table 9: Earth Science Publishing Volume

	Web of Science			GeoRef		
	Article	Proceedings	Total	Peer-Reviewed Document	Proceedings	Total Documents
Total documents	1,052,405	218,830	1,468,529	1,117,417	1,292,875	3,924,279
Total documents, 2010-2014	231,608	30,511	282,923	166,634	174,573	383,487
Published 2010	41,589	6,402	53,142	37,291	31,804	81,049
Published 2011	43,551	4,621	51,362	38,457	36,478	83,780
Published 2012	45,045	5,972	54,921	33,403	39,789	82,790
Published 2013	49,785	6,928	60,615	35,716	39,741	81,338
Published 2014	51,638	6,588	62,883	21,767	26,761	54,530

Notes:
Data current to April-June 2016, except where otherwise noted.
Web of Science has some overlap between Article and Proceedings (i.e. some documents are classified as both)
Both Web of Science and GeoRef index additional types of materials

Based on these data, Web of Science indexes slightly more journal literature than GeoRef. However, GeoRef indexes vastly more conference literature, as well as noticeably more alternate types of documents (generally, reports and other gray literature discussed below). As such, a reasonable estimate of the overall volume of literature would be an amalgamation of the two sources, using Web of Science for journal literature and GeoRef for conference and other literature. This rate is estimated in Table 10 (calculations are for 2010-13 only, as the sharp decrease in volume in GeoRef for 2014 implies that perhaps the ingest of material into the database is incomplete for that year).

Table 10: Estimated yearly publishing volume in Earth Science

	Journal Articles (WoS)	Conference Proceedings (GeoRef)	Other Resource Types (GeoRef)	Total Documents
Published 2010	41,589	31,804	11,954	85,347
Published 2011	43,551	36,478	8,845	88,874
Published 2012	45,045	39,789	9,598	94,432
Published 2013	49,785	39,741	5,881	95,407
Yearly Average	**44,993**	**36,953**	**9,070**	**91,015**

Information-Seeking Behavior

While it predates much of the technological infrastructure now used to transmit scholarly information, a 1989 study of the information-seeking behavior of geoscientists provides illuminating conclusions relating to the types of information used by these researchers, as well as points of frustration these researchers experience which a properly-designed system could alleviate[35]. The study found that many geoscientists relied heavily on their personal network for awareness of published information; these networks include professional contacts who send unsolicited information as well as graduate students who understand the needs and interests of their faculty and help them stay current. In general, researchers are pressed for time and did not use search tools as effectively as they could have. Additionally, researchers faced further challenges in obtaining information, including delays in reports being released, particularly from for-profit entities, as well as encountering literature in foreign languages. These could all represent opportunities for entry into the discipline: a database with effective alerting abilities to replicate those of the professional network, a database which facilitates the rapid release of results, and a database which helps users to explore foreign literature (perhaps through translations or short synopses) could help to meet the needs of geoscience researchers.

Finally, researchers expressed frustration with GeoRef, the major subject-specific database in this field, particularly for being difficult to search on a given concept. While this assessment is now 25 years old, it distinctly echoes the language of the National Research Council's report on a digital mathematics library, discussing the need for effective ways to search concepts across the literature.[36] Clearly, this database

[35] Bichteler and Ward, "Information-Seeking Behavior."
[36] Committee on Planning a Global Library of the Mathematical Sciences, "Developing a 21st Century Global Library for Mathematics Research."

requirement is not specific to one subject, but rather it has the potential for positive impacts across disciplines.

A later study by the same author further investigates various types of gray literature used by geoscience researchers[37]; the author posits that due to the nature of the discipline, geologists require a wider variety of gray literature than in many other scientific disciplines. Examples of gray literature discussed by Bichteler include geologic field trip guidebooks, USGS and state survey open-file reports, research newsletters, maps, and dissertations and theses. These literature types are, in many cases, crucial for earth science research: because much research is geographically oriented, maps play a vital role; similarly, locally-produced publications like field trip guidebooks written by area experts or theses from a nearby institution are often the best source of information on the geology of a specific area. Open-file reports are a common way to disseminate the results of surveys quickly without the extensive effort and time that goes into official reports.

Interdisciplinary Profile

Van Leeuwen's analysis of cross-disciplinary citations does not specifically pinpoint the interdisciplinary behavior of any of the Web of Science subject categories identified as a part of Earth Science. None of the categories are listed in either the top or bottom 15 disciplines, implying that they exhibit interdisciplinary behavior close to the overall average in the sciences. However, as previously discussed, Astronomy & Astrophysics, Mathematics, and several subfields of Physics do appear in the list of lowest CDC rates, and so it is reasonable to expect that Earth Science fields have more interdisciplinary citation patterns than these other disciplines.[38]

The NSF's Science & Engineering Indicators report requires some amount of manipulation to interpret properly; the NSF has assigned the broad field of Earth & Space Sciences to include the fine fields of Astronomy & Astrophysics, Earth & Planetary Science, Environmental Science, Geology, Meteorology & Atmospheric Science, and Oceanography & Limnology. We can remove Astronomy & Astrophysics (its own discipline in our analysis) and Environmental Science (part of a separate department at Harvard) from the broad field to achieve a calculation approximating the discipline we are interested in. In this constructed discipline, 80.9% of citations point back to the Earth & Space Science broad field.[39] Interestingly, 10.7% of references point to articles in Biomedical Research, and 3.0% point to articles in Biology, indicating a much stronger interdisciplinary connection to these areas than any of the other disciplines considered in this report.

Disciplinary Observations

Earth Science has a significant research community; the major professional society in the U.S. is the American Geophysical Union, with 60K members,[40] but other subfield-specific societies also boast significant membership, including the American Association of Petroleum Geologists (40K members[41]) and the Geological Society of America (26K members[42]). The American Geological Institute (AGI) acts as

---

[37] Bichteler, "Geologists and Gray Literature."
[38] van Leeuwen and Tijssen, "Interdisciplinary Dynamics of Modern Science."
[39] These references may still be pointing to the removed fine fields of Astronomy & Astrophysics or Environmental Science; there is no way to remove those from the receiving end of citations in this report.
[40] http://about.agu.org/about/files/2016/02/2014_Annual_Report.pdf
[41] http://www.aapg.org/about/aapg/overview
[42] http://www.geosociety.org/aboutus/index.htm

a network of geological societies in the U.S. and abroad, citing a cumulative membership of 250K scientists across their 51 member societies.[43] Additionally, other international federations exist to represent earth scientists outside of the purview of the AGI, including the European Geosciences Union (12,500 members[44]). Overall, the community size outpaces that of Astronomy and Astrophysics or Mathematics, and is comparable to that of Physics.

Aside from the Journal of Geophysical Research, published by the AGU and accounting for almost 10% of all citations to earth science literature, citations are spread among many journals in the field. Including JGR, the top 35 journals contain account for only 50% of the citations to the earth science literature, likely due to the preponderance of journals covering narrow subfields; top journals by citation counts include Journal of Climate, Journal of Hydrology, Chemical Geology, and Tectonophysics. Additionally, while many journals are published by professional societies, there is also a substantial commercial presence in the journal market, from Elsevier in particular: Elsevier owns six of the top 20 titles by citation count, and publishes two more on behalf of societies.

GeoRef, created by the AGI and offered on several different platforms (available to Harvard through ProQuest), is the most comprehensive index of scholarly literature in the earth sciences. According to the AGI website, more than 100,000 references are added each year. Additionally, GeoRef has grown to address some of researchers' needs regarding gray literature, including indexing U.S. and Canadian theses and dissertations; many open-file reports; and geologic field trip guidebooks, compiled by the Geoscience Information Society with the support of AGI[45]. However, the database does not include any full-text: these gray literature items are cited so that researchers know they exist, but full text is stored elsewhere, if available at all. Additionally, GeoRef does not comprehensively index certain major disciplinary sources, such as the USGS Publication Warehouse. Finally, GeoRef lacks citation indexing, a feature that would be particularly useful for traversing gray literature such as surveys, reports, and studies which are not published in journals.

As described in Bichteler's study, maps are a vital resource to earth scientists. Maps from many different publishers—USGS and state surveys, scholarly societies, universities, and the private sector—are cataloged, with latitude and longitude coordinates in the USGS's National Geologic Map Database[46]. However, many independent sources for maps also exist, as can be seen from numerous university library guides; these sources are often specialized for a particular purpose or localized to a geographic region. Maps and other location-specific resources and studies could offer the potential for a linked data system similar to that implemented in the ADS with astronomical objects. In fact, many GeoRef documents with location-specific information are already indexed with latitude and longitude coordinates.

[43] http://www.americangeosciences.org/about
[44] http://www.egu.eu/about/
[45] "Geologic Guidebooks of North America Database | American Geosciences Institute."
[46] "USGS National Geologic Map Database."

IV. Potential ADS Expansion

Logistical Needs

The practical needs of expanding the ADS to one or more additional disciplines are fairly straightforward, and comprise two main areas: hardware/physical infrastructure, and staffing resources.

Hardware/Physical Infrastructure

Currently, the ADS only commits about 4% of its budget to hardware and web hosting charges ($86K out of a budget of $2.3M in FY16). Expansion to additional disciplines has the potential to increase that cost by requiring additional data storage and additional bandwidth for the expanded user community. Data storage would not be a significant increase: the ADS already indexes a majority of the physics literature, which is larger than both mathematics and earth science put together. And so while the increase in data storage needed (and the costs associated with it) would not be negligible, for the disciplines discussed in the project the increase would be minimal. The increased user traffic resulting in such an expansion, however, would be notable: as shown previously, the research community in Astronomy and Astrophysics is relatively small. Assuming that the end goal of the expansion of scope would be user engagement in other disciplines comparable to the current level of engagement in Astronomy and Astrophysics (that is, basically, daily use by every active researcher), the ADS should expect an increase in user traffic by at least an order of magnitude in Physics or Earth Science, and likely in Mathematics as well. While ADS staff agree that the relationship between a bigger database or increased user traffic and the hardware resources required is not quite linear, this could still potentially be a significant increase in cost.

Fortunately, the ADS's recent migration to using Solr for search and using Amazon Web Services for cloud computing means that the software on the back end of ADS is scalable; while the required resources may be greater as a result of increased corpus size and user traffic, the database functionality can handle such an increase. Additionally, while the staff may spend development time building discipline-specific services and functionality, the API-based infrastructure allows for the easy addition of new services as needed.

Staffing resources

Expansion to additional disciplines will certainly require an increase in the staffing levels at ADS, as the tasks currently performed across the organization will need to broaden to incorporate more data, develop and support more services, manage user interactions—both support and outreach—for a much larger community, and negotiate collaborations and cooperation with a wider range of data providers, publishers, and scholarly organizations.

Staff will need to be recruited for data ingest and curation who can offer a thorough understanding of the subject matter being indexed, including the bibliographic metadata and citations as well as the full text of the documents and any related discipline-specific data linking. Subject knowledge is especially important for curation related to linked data capabilities, as these may include important disciplinary concepts which will need to be properly reconciled (a good example is if the ADS were to expand to mathematics and offer equation searching using MathML; the curator would need an understanding of equation syntax to properly implement this system).

Expansion into a new field would benefit from a large, short-term increase in the volume of curation staff to take on the immediate task of ingest and curation of historical literature. Given sufficient

system automation, this could potentially be accomplished by hiring temporary student workers in the target disciplines. After the initial ingest process, an elevated level of curation staff would need to be retained in order to process newly-published literature on a regular basis; these are the staff for whom facility with the discipline is most important. While some of the ingest processes are likely transferrable from subject to subject (particularly for cross-disciplinary publishers such as Elsevier), much of this work will involve interaction with a completely new set of publishers. Additionally, the current staffing level lacks sufficient redundancy in procedural knowledge. As such, it is expected that the need for ingest and curation staff will grow approximately linearly with yearly publication volume.

The need for development staff will also increase if the ADS expands to other disciplines, as expansion would necessitate the need for additional, discipline-specific services, and these services will need to be developed, tested, maintained, and kept current as the database (and the disciplinary community) evolves. However, because the database will continue to run on the same general infrastructure, the increase in development effort does not necessarily need to be linear with either publication volume or community size. In fact, the increase in size of the development group should be slightly higher than what is needed to maintain the system under the new constraints of size and user traffic; this strategy will free time for developers to invest in creating new, subject-specific, community driven services to enrich the database. Additionally, if discipline-specific data and search needs can be identified at the beginning of the expansion process, project developers could be brought on for the initial development efforts, creating services which would then be maintained later by ongoing staff.

If we expect that the new user communities will engage with the ADS at the same level as astronomers do currently, and we intend for the ADS to offer the same level of service to these user communities, then the staff effort spent on user support will likely increase approximately linearly with community size. For any of the communities under consideration, this is likely to result in a large organizational commitment of at least three full staff members (based on the FY16 effort level of 0.93FTE). While this responsibility is currently distributed around the organization, such a large organizational commitment would likely require hiring staff dedicated primarily to user support and implementing an infrastructure around day-to-day user interactions, to ensure prompt consistent messaging with users and to improve the internal efficiency of the user interaction function. This increased focus and efficiency could potentially reduce the staffing level needed for user support. It would likely also reduce the per-FTE budget for this functional area, by improving specialization and allowing for the hiring of lower-level staff to handle the easier interactions with users.

The administrative staff role will need to show significant growth as well: the current PI and PS expend significant effort in leadership and evangelism: leadership within the ADS to guide development and growth that mirrors the evolution of the field of astronomy; and evangelism to the astronomy community on behalf of the ADS, to promote the database in the community and to forge collaborations that will allow the ADS to both respond to and drive changes in the way the community accesses and uses information. Each additional discipline added would likely need at least an equal level of effort, and likely more depending on the community size, from practicing researchers at the same academic level (full faculty members). These staff will be needed to plot the ADS's development relevant to the discipline, and to fully collaborate with each scientific community, building new relationships with a wide array of organizations and advocating for the use of the ADS even despite other, more well-established disciplinary resources which may be available.

Political Needs

In addition to these logistical considerations, the ADS would also need to build its standing in each discipline it intends to expand to. While hiring researchers to act as ADS administrators for a particular discipline is necessary, equally important is recruiting prominent researchers in each field to buy in to the idea of ADS, even as it is in its developmental stages. Chris Erdmann paraphrased one astronomy researcher as saying "giving credit to the ADS is like giving credit to the air I breathe." Fostering a culture like this in other disciplines is key to a successful expansion, and user engagement has played a significant role in the ADS since its inception.

Just as important is gaining the support and even active participation of scholarly societies in the target disciplines. These societies hold significant sway over their members, and publicity through the society directly, through invited talks at society events, as well as other official channels can go a long way towards cultivating an audience within a discipline. This task will be particularly challenging in fields like physics and earth science, where specialized societies within the discipline are plentiful. Additionally, many societies are currently operating existing bibliographic databases within their disciplines (e.g. IET in physics, AMS in mathematics, and AGI in earth science); gaining the support, cooperation, and collaboration of these societies will be paramount to ensure that an expansion of scope by the ADS is seen as a benefit to the community and not an attempt to compete for market share.

Finally, it will be important to earn the support of research organizations such as national laboratories or geological surveys, as well as that of libraries and librarians within the relevant fields. Not only do these types of organizations also hold sway over the tools that researchers in these disciplines use, they also act as important collaborators that help in the development of the database, adding content and curating data which vastly improves the impact of the system as a whole. Astronomical observatories, for example, curate their own bibliographies within ADS, allowing users to see research relevant to or produced by that particular organization. Labs and/or geologic surveys will need to contribute in the same way. Libraries also have important roles to play, not just as advocates but also as data providers. Much of the historical scanned content in astronomy came from the holdings of the Center for Astrophysics' Wolbach Library, and Wolbach Library staff also made significant contributions towards the curation of content in various ADS collections. It will be critical to find libraries to fill both of these functions if the expansion to a new discipline is to be successful.

Obstacles to Expansion

Several significant factors could present obstacles to a successful ADS expansion into additional disciplines:

Current status of information searching

If the response quoted above from David Nelson is any indication, one potential obstacle to success is researchers' impressions that they already have everything they need, and that further investment (by them, or by the institution) is not necessary. The relatively recent change in information delivery from the print world to online was transformative, and compared to that transition, a "simple" change in search tool may seem paltry. Indeed, the ADS was representative of that change to astronomers, as the early developer of online information delivery in the field. Moving into these other fields, it will need to establish itself without that immediate draw. Additionally, Google Scholar has won many devoted users based on the simplicity of a single search box, while preprint servers like arXiv have developed strong user groups through their immediacy and their alerting systems. Faculty and other researchers have

heavy competition for their attention, and so a new resource must offer clear benefits over these existing tools to justify the time and effort required for researchers to learn a new system and shift their habits and behaviors. If the ADS is not able to gain traction with researchers over the omnipresent noise in their professional lives, then its ability to build a user community will be severely hindered.

Distributed nature of target disciplines

All three of the disciplines targeted for expansion are notably more distributed, both within their field and on an interdisciplinary level, than astronomy and astrophysics. As shown above, astronomy and astrophysics has in fact been demonstrated to be the most compact field in the sciences. This distribution manifests itself as a challenge in multiple ways.

First, reference lists and citation links will be less comprehensive in these fields, as researchers are more likely to cite resources or be cited by articles outside of their field. This could reduce the reliability of the citation counts, especially for evaluation. While astronomers evaluate themselves almost exclusively on citation counts in ADS, earth scientists may need to use a combination of ADS and Web of Science or Google Scholar to get a more complete picture. In fact, the ADS originally had this problem with astronomy literature, which is why they began indexing the physics literature in the first place: to ensure more complete reference lists for astronomy papers. This is a potential solution here, also, although the target disciplines cite research in chemistry, biology, biomedical research, and engineering and technology, and the sheer volume of publications in these additional fields may be prohibitive.

Additionally, the distributed nature of research communities within these disciplines means that uptake within the community will probably be slower or less likely to spread. An atmospheric chemist, for example, who discovers and loves the database, may help spread the word to her own community, but is less likely to help spread the word to an earthquake scientist. This increased disconnect within each community means that the work of the on-staff ADS evangelists is much greater, as they need to connect directly with more people.

Finally, distributed research profiles mean distributed journal profiles, and while the vast majority of astronomy research can be obtained through agreements with a couple of societies, research in these additional disciplines requires much more work in the way of collaborations with societies and publishers and preparation of infrastructure for ingest.

Need to cooperate with publishers

In order to overcome many of the challenges enumerated above, the ADS will need to collaborate with all major publishers in each discipline. A large component of the success of the ADS comes from the comprehensive coverage of the subject, and without the publishers' cooperation, such coverage is difficult to achieve. Additionally, non-cooperative publishers make for more work during ingest, as the ADS would not be able to negotiate with them for properly formatted metadata. Many astronomy publishers also currently provide the full text of their publications to the ADS for full-text indexing and data mining, creating a more powerful search tool (note that this full text is not shared publicly, just the results of the indexing and data mining are); publishers who do not cooperate with this process weaken the search capabilities of the database. However, the ADS does not have the same history and name recognition in these fields that it did in astronomy, and so some of these partnerships may be difficult to forge.

Need to regularly and sustainably engage with researcher needs

It is not sufficient for the ADS to expand its scope to a new field, build some useful linked data services, create a user community, and then relax development and maintain the status quo. The hallmark of ADS throughout its existence is its continued engagement with the astronomy community, and its evolution as a digital library serving that community. This status requires responsiveness to new and emerging information and data needs, types, and uses. Moreover, the database needs to identify opportunities and drive transformational change in information search and usage in these fields. Staff at the ADS pride themselves on not needing an official user group, because they know the community they work with so well. Any disciplinary expansion should be equally engaged with the community, to ensure that they are providing the services that users need most. This engagement has added benefits back to the ADS as well, in the form of user-contributed content and corrections.

Competition in this space

ADS defined the information searching space in the field of astronomy, establishing itself as a free resource catering to the needs of the community. However, this was not necessarily the case in the target disciplines under discussion for expansion. As databases such as IET's Inspec in physics, AMS's Mathematical Reviews (becoming MathSciNet) in mathematics, and AGI's GeoRef in Earth Science transitioned to online presences, they also developed for-pay business models by which they recouped at least some of their operating costs through subscription fees. It is not clear to what extent these societies rely on subscription income to subsidize other society operations; while Mathematical Reviews was originally published at a reduced cost to foster distribution,[47] many scholarly societies today publish journals and other resources at a profit to benefit the rest of the organization.

For these organizations, then, the ADS would potentially be encroaching on their market space, which is already small due to large interdisciplinary resources such as Web of Science, Scopus, and Google Scholar. Particularly if ADS continues its business model as a funded free resource, these societies may anticipate having trouble retaining subscription income with libraries facing budget pressures and therefore feel the need to compete for market share to retain subscriptions, rather than join a collaboration to improve the status of their community's information resources. Additionally, in an extreme worst-case scenario, commercial distributors could get involved if they feel their interests are threatened, particularly for Inspec and GeoRef, both of which are distributed through a variety of platforms (including Elsevier's Engineering Village, EBSCO, Ovid, and ProQuest). It will be important to approach all of these entities as potential collaborators, to help the transition into these fields.

ADS culture

A possible concern in the expansion of the database is the effect that such a transformative change could have on the small, agile, highly efficient group of staff currently employed at ADS. Currently the culture is such that staff are trusted to be independent and flexible. While they are aligned in overall strategy, in many cases they have the freedom to develop and advocate for their own way of furthering the organization towards that strategy. A rapid expansion of staff could mean that more structure needs to be instated in the organization, potentially restricting the independence that existing staff are accustomed to and thereby reducing the ingenuity and per-capita efficiency that the organization as a whole currently enjoys. Additionally, much of the hiring to date was opportunistic: one interviewee was offered a job as the result of a presentation he gave while working for a different database, while

[47] American Mathematical Society, "Announcing Mathematical Reviews."

another was offered a job straight out of college after working on the system as an undergrad. However, for a deliberate expansion such as this, the ADS will not have the luxury of patiently waiting to find and recruit a potential staffperson with the optimal personality fit or skill set; they will have to opt for the best available people at the time. Furthermore, many of the most appropriate potential employees would likely have the opportunity to make significantly more money in industry, and so the ADS will need to sell the positions based on, among other things, the organizational mission and the prestige of working at Harvard.

Reduction of support from NASA

While the ADS is currently fully supported by NASA for the benefit of NASA-funded astrophysics research, an expansion of scope into other areas could be seen as fundamentally changing that mission, even if the expansion does not necessarily mean a reduction in the resources and attention devoted to astronomy and astrophysics. As such, NASA may be inclined to reduce funding from current levels, especially in areas (such as hardware costs) where resources are jointly supporting all disciplines covered by the database. It would be prudent to discuss this possibility with NASA before embarking on a potential disciplinary expansion, to avoid unexpected funding changes in the future.

Potential Winning Impacts

Based on the observed successes of the ADS within the astronomy community, as well as the characteristics of the target disciplines discussed previously, we can identify several areas where the ADS could make a significant positive impact that would help it to gain traction as it builds a user base in each field.

Linked data, objects, and concepts

As the concept of the "Semantic Web"[48] becomes more commonplace, many researchers are beginning to consider how the idea of linked data, objects, concepts, and the like can benefit research in the sciences. As discussed previously, the International Mathematical Union has been discussing and studying these concepts to plan and build a global digital mathematics library for some time.[49] Many of the benefits envisioned through the IMU's work are the result of linked data, such as the ability to explore the published literature by mathematical concept, by equations, and by citations.

This is almost precisely what the ADS has developed in the field of astronomy: a full set of bibliographic data, traversable by citations, and linked with astronomical objects, observatory bibliographies, data sets, and software. With the updated Bumblebee interface, additional linked data objects are available, including concept terms pulled from the full text of articles and author networks. Second-order operators are also possible, allowing a search for, say, all papers that cite articles about observations made at the Gemini Observatory. Successfully demonstrating this functionality, translated to data objects in the target disciplines, would likely help the ADS to build devoted user groups to act as advocates among their peers. Potential applications could include linking mathematical literature to the data contained in the NIST Digital Library of Mathematical Functions,[50] or implementing a search of the earth science literature by geographic location or proximity to a set of physical coordinates.

---

[48] https://www.w3.org/standards/semanticweb/

[49] Committee on Planning a Global Library of the Mathematical Sciences, "Developing a 21st Century Global Library for Mathematics Research."

[50] http://dlmf.nist.gov/

Scanning and indexing historical and gray literature

In the early stages of the development of the ADS, most journal literature had not yet transitioned from print to electronic. Most of the journals that had transitioned were only creating electronic copies for current content moving forward. As such, the ADS forged agreements with many astronomy publishers, allowing them to digitize and freely distribute all historical content back to volume 1; this allowed publishers the ability to focus on their content moving forward. ADS staff developed a scanning strategy, obtained print copies (many from the CfA's Wolbach Library), and fully digitized the backfiles of most major astronomy journals from the early 1800's up to around 1997. Because this was the only place to find this material electronically, this drove early traffic to the ADS.

Most publishers in other fields have now digitized their backfiles and offer them for purchase. However, a concerted effort to locate, obtain, and digitize additional types of material—particularly historical gray literature which would be otherwise difficult to search for, let alone obtain—could provide potential users added incentive to use the ADS. This strategy could be particularly useful in a field like earth science, which has a strong history of gray literature publication, including guidebooks, open-file reports from USGS and state surveys, maps, and the like. Additionally, much of the gray literature in earth science resides in uncataloged or minimally-cataloged collections within libraries. Assembling these sets of objects into digital, findable collections would do a great service for the earth science community, and would help drive traffic to the ADS out of necessity. However, such a strategy could also be generalized to fields like physics and mathematics, where these non-standard document types are less prevalent but where much historical conference and technical report literature remains difficult to locate and unavailable electronically. It is worth noting, however, that this these types of resources could take significant effort to locate and obtain; partnerships with prominent libraries in these fields could remove some of this onus and help the ADS to identify appropriate targets for this work.

Improving citation mapping and other metrics

Astronomers currently use the citation data indexed in the ADS to evaluate the impact of the research published by a specific author, research group, organization, etc. This is not only because of the well-curated nature of the database and its citation links, but also because citations to and from a wide range of resources are tracked (data sets, observing proposals, software, curated bibliographies) and because the structure of the ADS allows for more in-depth searching using citation links as second-order operators. Additionally, in the new Bumblebee interface, the ADS has incorporated several visualizations around citation links as well as authorship patterns in a group of papers; usage data tracking full-text requests from within the ADS; and related papers, observing what documents users regularly read in sequence.

Citations are indexed in many different systems (Web of Science, Scopus, Google Scholar, and many subject-specific databases), but the in-depth, well-curated links in ADS, combined with the search functionality that allows these links to be used as paper-, author-, or organization-level metrics, could prove to be a large benefit to disciplines which are looking to better define a method of measuring and tracking output and impact. In fact, the report on developing a 21st century digital mathematics library

spends three pages (47-49) discussing the desirability and impact that much of this existing functionality would have on the field of mathematics.[51]

Potential Collaborators

Collaboration with various partners across the field of astronomy has been crucial to the successful development of the ADS since its creation. In the process of gathering data about the ADS and about the fields targeted for a possible expansion of scope, several potential partners were identified which could aid in the growth of the system.

Publishers

Cooperation with journal and conference publishers will be key in expansion of the database. The ADS already has agreements with several interdisciplinary publishers that cover some of these target disciplines, but further development of relationships with discipline-specific publishers will be necessary. In astronomy and astrophysics, the American Astronomical Society journals were offered online and indexed through ADS in 1995, and 1996 saw "nearly every astronomy journal which had not already joined into collaboration with ADS join."[52] The high concentration of astronomy research in AAS journals likely contributed to such a quick acceptance of the ADS from the publishing world—other astronomy publishers needed to collaborate with the ADS or risk irrelevance.

This tipping point may not be reached quite as quickly in other fields, because as shown in the discipline evaluations, these target disciplines are more distributed. The best strategy will be to expand existing collaborations with the commercial publishers that cover content in each discipline (e.g. Elsevier, Springer, and Wiley), while forging collaborations with the top society publishers in each discipline: the American Physical Society and the American Institute of Physics for physics (the ADS already has an existing relationship with these societies), the American Geophysical Union and the American Meteorological Society for earth science, and the American Statistical Association for mathematics. Using these collaborations for leverage would likely encourage other smaller disciplinary publishers to collaborate with the ADS as well.

Data providers

The key to building a comprehensive linked database such as the ADS is developing relationships with the organizations that create, control, curate, and provide the data to users. There are several potential types of organizations for which these collaborations would be necessary:

- **Object information databases**. Similar to the way the current database is connected with the SIMBAD database of astronomical objects, various information providers could be linked to the ADS in these target fields. For physics, this could include targets such as the National Nuclear Data Center, at Brookhaven National Laboratory[53] or the physical data stored at the NIST Physical Reference Data site.[54] For earth science, this could include data from the USGS, including maps, locations, or geologic names from the National Geologic Map Database.[55] For

[51] Committee on Planning a Global Library of the Mathematical Sciences, "Developing a 21st Century Global Library for Mathematics Research."
[52] Kurtz et al., "The NASA Astrophysics Data System."
[53] http://www.nndc.bnl.gov/
[54] http://www.nist.gov/pml/data/
[55] http://ngmdb.usgs.gov/ngmdb/ngmdb_home.html

mathematics, this could include the Digital Library of Mathematical Functions[56] or other catalogs of concepts and objects discussed in the NRC report (pages 24-25)[57].

- **National laboratories and other research organizations**. Part of the power of ADS currently comes from the bibliographies of research products from NASA projects and observatories, which are generated and curated by researchers or librarians at each location. It will be necessary to build a network of similar types of organizations to create a relevant disciplinary set of bibliographies including Department of Energy and other governmental labs, USGS studies, etc.
- **Data archives**. Each discipline has various data archives available to provide research data from studies to the public. The ADS currently indexes research data products from Vizier,[58] improving the visibility of those datasets and the ability of astronomers to cite them in their work. Similar data archives can be located for these additional disciplines, especially since the NSF, a major funder in all of these areas, has instituted a data management plan requirement for their grant applications. None of these fields maintain a source as complete as Vizier, but locations like NASA's Earth Observing System Data and Information System[59] are a start. More in-depth analysis of data archiving practices in these fields could identify targets for indexing.

Funding partners

An important question for this potential expansion is how it might be funded and hosted. As mentioned above, the current funding from NASA is provided particularly to support NASA Astrophysics and would almost certainly not be increased to support an expanded disciplinary scope (it may even be contracted if the infrastructure were used to support more than just astronomy and astrophysics).

This expansion has been conceptually described by the ADS administration as a "Pubmed for non-medical sciences." As such, the natural funding source for such an endeavor would be the National Science Foundation, the federal agency tasked to promote the progress of science in the U.S. The NSF has taken on digital library initiatives in the past.[60] Additionally, because they are a major funder of research in the target disciplines, they would have a vested interest in supporting a tool that improves the work of their grant recipients; recall the quote of Christopher Stubbs: "taxpayers and granting agencies win because my work is more efficient." Testimonials such as this, coupled with the usage analyses described in section 2.5.1 of the programmatic review,[61] offer a convincing argument that support from the NSF would be of substantial financial gain to the agency.

The Department of Energy may also serve as a possible funding source; this agency oversees several national laboratories which would derive significant benefit from the ADS expanding into physics and earth science (e.g. SLAC National Accelerator Laboratory, Princeton Plasma Physics Laboratory, and Brookhaven National Laboratory). The DOE's Office of Science oversees a program known as the Office

---

[56] http://dlmf.nist.gov/
[57] Committee on Planning a Global Library of the Mathematical Sciences, "Developing a 21st Century Global Library for Mathematics Research."
[58] http://vizier.u-strasbg.fr/
[59] https://earthdata.nasa.gov/about
[60] http://www.nsf.gov/news/special_reports/cyber/digitallibraries.jsp
[61] Murray, Accomazzi, and Kurtz, "Astrophysics Archives Programmatic Review 2015: The NASA Astrophysics Data System."

of Scientific and Technical Information (OSTI); OSTI already operates a database called SciTech Connect[62] disseminating DOE-sponsored R&D results, with significant contents in physics and earth sciences. An expanded ADS database could be in line with the OSTI's commitment to disseminating the results of scientific endeavors, and including the resources indexed in SciTech Connect within the ADS would greatly increase the visibility of that information. However, the OSTI's mission is more focused on the dissemination of publicly available information, specifically results of DOE research. And so it is possible that OSTI would not be interested in supporting a tool which indexes a wider range of science in these fields.

Professional societies in these disciplines may also be interested in supporting expansion of the ADS into their disciplines. This is not unfamiliar territory for many of these organizations. Societies such as the American Mathematical Society (MathSciNet) and the American Geological Institute (GeoRef) already operate in this space; other societies such as the American Physical Society, International Mathematical Union, and American Geophysical Union would see great benefit to their communities; and more general organizations such as the American Association for the Advancement of Science and the National Academy of Sciences would likely see great value in an interdisciplinary database of this type. A consolidated effort across societies could potentially yield the support necessary for expansion. The observed and measured efficiency gain for researchers, as well as the ability to have a voice in the direction of development of the expanded database, would be particularly compelling arguments to encourage these societies to lend support. However, the societies currently operating subject-specific databases do so via subscriptions which at least subsidize the cost of operating the database (if not make a profit to support other society activities). Organizations supporting the ADS may be interested in a similar model. However, a pay model may be at odds with the current philosophy of the ADS management and its existing funder, NASA (who pays for the operation of the database on behalf of its researchers). A pay model may or may not be in alignment with the goals of the Harvard Library in taking on this project.

Promotion partners

While not all of the above organizations may be willing or able to contribute financial support to the ADS were it to expand its scope to include these additional disciplines, many of them would still be targets for memoranda of understanding or other non-financial agreements by which the organizations still lend support to ADS.

This type of agreement would be particularly powerful with any of the scholarly societies discussed in this report. These societies are valuable resources to their members, encouraging professional communication and collaboration, organizing conferences, and distributing various publications. Undoubtedly, positive publicity via word-of-mouth through any of these societies would go a long way towards connecting with the user community in that field.

The National Academy of Sciences, and as a part of it, the National Research Council, could serve as strong advocacy partners for an expanded, interdisciplinary ADS. The NRC is already interested in working in this space: they composed the report, supported by the International Mathematical Union, which described the need for a 21st century digital mathematics library. While they do not serve as a granting organization and therefore may not be able to help financially, they do employ a strong voice at

[62] https://www.osti.gov/scitech/

the federal level which could help raise awareness of the availability of a linked-data system like the ADS.

Finally, library organizations would be important allies to help an expanded ADS reach the research communities it needs to be successful. Groups such as the Special Libraries Association (SLA), the American Library Association (ALA), and the International Federation of Library Associations and Institutions (IFLA) could all help reach subject librarians serving researchers in these target communities. Subject librarians can couple their connections to library-engaged researchers with their subject-specific knowledge to effectively demonstrate to prospective users the power of a linked-data digital library such as the ADS, and can also maintain a consistent engagement with the ADS itself to act as a liaison between the management at the database and its users. Successfully cultivating a strong community of librarian supporters who are willing to advocate to and assist users at their institutions could also significantly reduce the projected user support costs discussed in the next section.

Cost estimates

For each discipline, cost estimates are made based on the budget data presented in Table 2, the discipline-specific information discussed above, and the anticipated logistical needs for expansion to additional fields. Three estimates are given for each discipline: a best-case (or low-cost) scenario, a medium scenario, and a worst-case (or high-cost) scenario.

Below is a summary of pertinent data points which estimates in all disciplines will make use of:

- The ADS currently ingests about 442K documents per year, including 60K from astronomy (from Table 1).
- The ADS currently curates about 251K documents per year (from Table 1); this assumes that astronomy documents are fully curated, and all other content being ingested is curated at approximately one-half effort.
- Based on these numbers and budget data in Table 2, it currently costs the ADS $1.08 to ingest and $1.09 to curate a single document.
- The ADS currently contains approximately 11M documents.

Physics

Up-front costs

As mentioned above, the major up-front costs will be ingest and curation of past literature, as well as any development needs. In the case of physics, Table 6 shows that Inspec contains approximately 1 million more total physics documents than the ADS currently; this is a reasonable middle estimate. The low estimate of half of a million documents could be realized if much of the Inspec content is deemed out of scope, or if much of the content is classified by the ADS as astronomy rather than physics. The high estimate of 2M documents could be realized if a significant amount of content is identified beyond what Inspec may cover. Curation is more difficult: physics content in ADS is likely partially curated already, but may need additional curation effort depending on their state and potential services being developed. The medium estimate of 4M documents assumes that the collection is about half-curated; the low and high estimates assume higher and lower levels of curation accordingly.

For development effort, in the low estimate, no up-front service development is performed and instead new services are developed as needed by permanent staff while running the database. In the high

estimate, a year-long single developer project is undertaken to identify and create additional user services to support the physics community. In the medium estimate, a shorter, three-month single developer project is undertaken to implement any critical services required before rollout to the physics community.

Table 11: Physics up-front expansion costs

	Estimated rate, FY16	Startup needs			Startup cost estimate		
		Low	Med	High	Low	Med	High
Ingest	$1.08/doc	0.5M docs	1M docs	2M docs	$540K	$1,080K	$2,160K
Curation	$1.09/doc	2M docs	4M docs	6M docs	$2,160K	$4,360K	$6,540K
Development effort	$189.6K/FTE	0 FTE	0.25 FTE	1 FTE	$0K	$47.4K	$189.6K
Total					**$2,800K**	**$5,487.4K**	**$8,889.6K**

Steady-state support

Because much of the physics literature is already covered by the ADS, the staffing increase for bibliographic ingest will be minimal; Table 6 identified approximately 24K more physics documents per year in Inspec than in ADS, indicating an estimated increase of about 5% over the current ingest rate. Indexing, will likely be more intensive for all physics literature, in order to support the expanded services built upon the database for physics. However, this increase will be tempered by the fact that some of this work is already being done on existing physics content.

Additional development effort will also be somewhat limited for physics. A search interface for physics already exists, and would likely only need limited additional ongoing development. However, the development of user tools will increase somewhat substantially; even though the identified need for user tools in physics is perhaps less than other fields, the much larger community size (~300K physicists vs. ~12.5K astronomers) means that it is likely at least as much development support will be needed for physicists as for astronomers. Software maintenance increases as a result, because even though the database size will be only slightly larger, the added user services will require upkeep.

User support and management will both need to increase significantly to support an expansion into physics, due to the large jump in community size (of about 24 times). For user support, the needed effort will increase proportionally to the community size, but at a lesser rate to account for the efficiencies gained by having a stronger user support structure in place. A reasonable estimate would be at the rate of a fractional power; the medium estimate is therefore an increase of 4.9x the current level of effort (the square root of 24). The high estimate uses half of the community size increase to obtain a 12x increase over the current level of effort, while the low estimate assumes that support is based more on database operation, thereby needing to invest equal effort to support physics as to support astronomy.

Management will similarly need to increase its investment proportional to community size, exerting more effort to collaborate and interact with, and guide database development and direction for, a larger scientific community. However, the organization of academic communities into professional societies

and organizations should give management even greater efficiency in connecting with the community. Therefore, a high-end estimate for increase managerial effort is the fractional power of the increase in community size. The low estimate assumes that the existing management group continues overseeing the general development of the database while new managerial staff take on limited roles in working directly with the community. The medium estimate falls in between the two.

Table 12: Physics ongoing staffing estimate

	FY16 $/FTE	**Fully staffed current ADS**	**Rate of increase over current ADS**			**FTE increase over current ADS**			**Cost increase over current ADS**		
			Low	**Med**	**High**	**Low**	**Med**	**High**	**Low**	**Med**	**High**
Bibliography Support	$212.0	3.95				0.49	1.70	2.58	$92.0	$330.7	$509.3
Bibliographic ingest	*$223.3*	*2.39*	*0.03*	*0.05*	*0.10*	*0.07*	*0.12*	*0.24*	*$16.0*	*$26.7*	*$53.4*
Indexing/archiving/databases	*$194.9*	*1.56*	*0.25*	*1.00*	*1.50*	*0.39*	*1.56*	*2.34*	*$76.0*	*$304.0*	*$456.0*
Development	$189.6	4.25				0.83	1.80	3.59	$151.5	$328.6	$657.2
search functionality	*$216.8*	*0.8*	*0.05*	*0.10*	*0.20*	*0.04*	*0.08*	*0.16*	*$8.7*	*$17.3*	*$34.7*
user tools	*$180.1*	*0.68*	*0.75*	*1.50*	*3.00*	*0.51*	*1.02*	*2.04*	*$91.9*	*$183.8*	*$367.5*
software maintenance	*$183.5*	*2.78*	*0.10*	*0.25*	*0.50*	*0.28*	*0.70*	*1.39*	*$51.0*	*$127.5*	*$255.0*
User Support	$344.0	1.07	1.00	4.90	12.0	1.07	5.24	12.84	$368.1	$1,803.5	$4,416.7
Management	$303.6	1.98	0.75	2.00	4.90	1.49	3.96	9.70	$450.9	$1,202.4	$2,945.9
Staffing Total	**$231.6**	**11.25**				**3.87**	**12.70**	**28.71**	**$1,062.5**	**$3,665.1**	**$8,529.1**

Notes: budget values are in thousands of US dollars. Predicted values are highlighted in gray.

Hardware and web hosting costs will see very small increases due to the added storage space needed, but larger increases due to the added computation and bandwidth needed to support a larger community. These costs will also increase proportionally to these values, but less than linearly; a fractional power estimate is used here as well. Other non-affected costs are presented for comprehensiveness; these costs include contributions to the CDS for SIMBAD and Vizier support as well as other membership fees (ORCiD, CrossRef).

Table 13: Physics ongoing non-staff cost estimate

	FY16 Cost	Relevant metric	Expected increase factor	Fractional power estimate	Expected total cost	Total cost increase over FY16
Storage	$43	Corpus size	1.1	1.05	$45.2	$2.2
Computation/ bandwidth	$43	Community size	25	5.0	$215.0	$172
Other Hardware and Licenses costs	$169.7	N/A	1	1	$169.7	$0.0
Total	**$255.7**				**$429.9**	**$174.2**

Notes: budget values are in thousands of US dollars. Predicted values are highlighted in gray.

Mathematics

Up-front costs

Based on Table 8, MathSciNet and Web of Science approximately agree that the total corpus of mathematics needing both ingest and curation is around 3.3M documents. This makes a reasonable medium estimate for both ingest and curation. The agreement between these two databases indicates that this is likely quite close to the actual number. A low estimate, representing that some of these

documents are out of scope or otherwise do not need coverage, could be around 2.5M, while a high estimate, representing that more documents are identified which need to be indexed, could be around 4M.

Up-front development effort is likely moderately higher for mathematics. Desirable functionality has already been identified by the IMU and the National Research Council report; it may be necessary to build these user services into the database to start. A medium estimate here would be a six-month project to develop the necessary services, while a high estimate would be two developers working on a year-long project to build a larger suite of user services. Again, the low estimate assumes that no up-front service development is performed and instead new services are developed as needed by permanent staff while running the database.

Table 14: Mathematics up-front expansion costs

	Estimated rate, FY16	Startup needs			Startup cost estimate		
		Low	Med	High	Low	Med	High
Ingest	$1.08/doc	2.5M docs	3.3M docs	4.0M docs	$2,700K	$3,564K	$4,320K
Curation	$1.09/doc	2.5M docs	3.3M docs	4.0M docs	$2,725K	$3,597K	$4,360K
Development effort	$189.6K/FTE	0 FTE	0.5 FTE	2.0 FTE	$0K	$94.8K	$379.2K
Total					**$5,425K**	**$7,255.8K**	**$9,059.2K**

Steady-state support

Based on Table 8, Web of Science has covers about 125K mathematics documents per year over the last five years; this is higher than MathSciNet's rate of about 104K, and so it makes a reasonable medium estimate for the ADS's ingest rate. An increase of 125K documents would represent a 28% increase in the current volume of ingest; low and high estimates are set closely at 15% and 40%, respectively, to signify the similar volume estimates between Web of Science and MathSciNet. 125K documents would represent about a 50% increase in the current volume of curation; the low and high ingest estimates would calculate to a 27% and 70% increase in curation volume.

As discussed above, the development effort for mathematics is likely to be reasonably large. While the mathematics community size is estimated to be at least 50K researchers, or four times the size of astronomy, user tools have already been identified which will need to be developed and maintained; this will likely require approximately as much development effort and maintenance as astronomy. Additionally, search functionality does not currently exist for mathematics; this separate database and interface will need to be developed and maintained.

Similar to physics, staff requirements for user support and management are projected based on the community size, which is calculated to be at least 4-5 times that of astronomy. For user support, the low estimate sets staff needs to be exactly the same as for astronomy, the high estimate assumes that staff needs grow at half the rate of the community size, and the medium estimate assumes that staff needs grow at a fractional power (square root) of the community size. For management staff, the low

estimate assumes that existing management retains their overall administrative role, while new managerial staff equal their level of effort in subject-specific strategy and interacting with the community; the high estimate assumes that managerial needs grow at a fractional power of the community size, and the medium estimate is approximately between the two.

Table 15: Mathematics ongoing staffing estimate

	FY16 $/FTE	**Fully staffed current ADS**	**Rate of increase over current ADS**			**FTE increase over current ADS**			**Cost increase over current ADS**		
			Low	**Med**	**High**	**Low**	**Med**	**High**	**Low**	**Med**	**High**
Bibliography Support	$212.0	3.95				2.52	4.32	7.67	$162.1	$301.4	$426.3
Bibliographic ingest	*$223.3*	*2.39*	*0.15*	*0.28*	*0.40*	*0.96*	*1.20*	*1.43*	*$80.1*	*$149.5*	*$213.5*
Indexing/archiving/databases	*$194.9*	*1.56*	*0.27*	*0.50*	*0.70*	*1.56*	*3.12*	*6.24*	*$82.1*	*$152.0*	*$212.8*
Development	$189.6	4.25				0.78	1.78	3.35	$146.9	$336.7	$630.1
search functionality	*$216.8*	*0.8*	*0.2*	*0.5*	*0.75*	*0.16*	*0.40*	*0.60*	*$34.7*	*$86.7*	*$130.1*
user tools	*$180.1*	*0.68*	*0.5*	*1*	*2*	*0.34*	*0.68*	*1.36*	*$61.3*	*$122.5*	*$245.0*
software maintenance	*$183.5*	*2.78*	*0.1*	*0.25*	*0.5*	*0.28*	*0.70*	*1.39*	*$51.0*	*$127.5*	*$255.0*
User Support	$344.0	1.07	1.0	2.2	2.5	1.07	1.61	2.68	$368.1	$809.7	$920.1
Management	$303.6	1.98	0.75	1.5	2.2	1.49	1.98	2.97	$450.9	$901.8	$1,322.6
Staffing Subtotal	**$231.6**	**11.25**				**5.85**	**9.68**	**16.67**	**$1,128.0**	**$2,349.7**	**$3,299.1**

Note: budget values are in thousands of US dollars.

Non-staff infrastructure costs will increase less than for physics: the corpus only grows by slightly more to bump up storage costs, but the smaller community means that computation and bandwidth costs will increase notably less. As before, other non-affected costs are presented for comprehensiveness.

Table 16: Mathematics ongoing non-staff cost estimate

	FY16 Cost	Relevant metric	Expected increase factor	Fractional power estimate	Expected total cost	Total cost increase over FY16
Storage	$43	Corpus size	1.3	1.1	$47.3	$4.3
Computation/ bandwidth	$43	Community size	5	2.2	$94.6	$51.6
Other Hardware and Licenses costs	$169.7	N/A	1	1	$169.7	$0.0
Total	**$255.7**				**$311.6**	**$55.9**

Notes: budget values are in thousands of US dollars. Predicted values are highlighted in gray.

Earth Science

Up-front costs

The overall volume of historical publications in earth science is likely to be more closely estimated by GeoRef than by Web of Science; as shown in Table 9 and discussed thereafter, GeoRef has a more thorough collection of conference material and other document types than Web of Science. As such, GeoRef's corpus size of 3.9M documents is an appropriate medium estimate of the number of documents needing indexing. However, this estimate could be off significantly in either direction: on

the low-cost end, ADS already offers strong coverage of geophysics, and so many of these documents may already be in the database (although not necessarily fully curated). On the high-cost end, however, gray literature is an important component of earth science research, and while GeoRef's coverage is good it is possible that ADS will locate a significant volume of additional material to be ingested, curated, and digitized.

Development effort could be significant in the earth sciences; the task of building user services and tools which take advantage of the location or coordinate data that is inherent in much of the geoscience literature could prove to be a significant effort. While the low estimate here is still based on no startup development effort, earth science more than other fields would greatly benefit from developing and implementing user tools at the outset, as this effort could have a major impact on the adoption of the database by users.

Table 17: Earth science up-front expansion costs

	Estimated rate, FY16	Startup needs			Startup cost estimate		
		Low	Med	High	Low	Med	High
Ingest	$1.08/doc	2.0M docs	3.9M docs	5.0M docs	$2,160K	$4,212K	$5,400K
Curation	$1.09/doc	3.0M docs	3.9M docs	5.0M docs	$3,270K	$4,251K	$5,450K
Development effort	$189.6K/FTE	0 FTE	1.0 FTE	2.5 FTE	$0K	$189.6K	$474K
Total					**$5,430K**	**$8,652.6K**	**$11,324.0K**

Steady-state support

The yearly output of earth science researchers was estimated in Table 10 to be approximately 91K documents per year, which would represent an estimated 21% increase over the current volume of literature ingested into the ADS, and about a 36% increase over the current volume of literature curated by the ADS. Conservative estimates are made for the low and high costs, for the same reasons as above (potential overlap with current coverage on the low end, and large potential batches of gray literature being missed by GeoRef and Web of Science).

Similar to mathematics, the search functionality for earth science would need to be developed, implemented, and maintained; this would result in a moderate increase over current ADS staffing needs. Additionally, the added effort expected for developing and maintaining new user tools such as location and coordinate searching is likely to be notably higher than that needed for astronomy tools currently, especially given the significant size of the research community in earth science. Finally, due to the large community of users and the complex additional services, the added costs for software maintenance have the potential to be slightly higher than for physics or mathematics.

The earth science research community is comparable in size to physics; the community size, as estimated by membership in AGI-affiliated societies, is 250K researchers, an estimated 20 times greater than that of astronomy. Projected staffing needs for user support and management are therefore calculated by the same methods used for physics, above.

Table 18: Earth science ongoing staffing estimate

	FY16 $/FTE	Fully staffed current ADS	Rate of increase over current ADS			FTE increase over current ADS			Cost increase over current ADS		
			Low	Med	High	Low	Med	High	Low	Med	High
Bibliography Support	$212.0	3.95				0.52	1.06	1.80	$108.1	$221.5	$375.3
Bibliographic ingest	*$223.3*	*2.39*	*0.10*	*0.21*	*0.35*	*0.24*	*0.50*	*0.84*	*$53.4*	*$112.1*	*$186.8*
Indexing/archiving/databases	*$194.9*	*1.56*	*0.18*	*0.36*	*0.62*	*0.28*	*0.56*	*0.97*	*$54.7*	*$109.4*	*$188.5*
Development	$189.6	4.25				1.40	2.87	4.99	$259.2	$535.7	$926.1
search functionality	*$216.8*	*0.8*	*0.2*	*0.5*	*0.75*	*0.16*	*0.40*	*0.60*	*$34.7*	*$86.7*	*$130.1*
user tools	*$180.1*	*0.68*	*1*	*2*	*4*	*0.68*	*1.36*	*2.72*	*$122.5*	*$245.0*	*$490.0*
software maintenance	*$183.5*	*2.78*	*0.2*	*0.4*	*0.6*	*0.56*	*1.11*	*1.67*	*$102.0*	*$204.0*	*$306.0*
User Support	$344.0	1.07	1	4.5	10	1.07	4.82	10.70	$368.1	$1,656.3	$3,680.6
Management	$303.6	1.98	0.75	1.75	4.5	1.49	3.47	8.91	$450.9	$1,052.1	$2,705.4
Staffing Subtotal	**$231.6**	**11.25**				**4.47**	**12.22**	**26.40**	**$1,186.2**	**$3,465.6**	**$7,687.3**

Note: budget values are in thousands of US dollars.

Due to the similarity in community size, hardware and web hosting costs will likely show increases comparable to those estimated for physics; the expected increase represented by a corpus size of 3.9M documents is about 40%, but the expected increase represented by a community size of 250K is 21x.

Table 19: Earth science ongoing non-staff cost estimate

	FY16 Cost	Relevant metric	Expected increase factor	Fractional power estimate	Expected total cost	Total cost increase over FY16
Storage	$43	Corpus size	1.4	1.2	$51.6	$8.6
Computation/ bandwidth	$43	Community size	21	4.6	$197.8	$154.8
Other Hardware and Licenses costs	$169.7	N/A	1	1	$169.7	$0.0
Total	**$255.7**				**$419.1**	**$163.4**

Notes: budget values are in thousands of US dollars. Predicted values are highlighted in gray.

Summary

A summary table of the predicted costs of expanding to each of the target disciplines is below.

Table 20: Summary of cost estimates for expansion

	Startup cost estimates			Ongoing staff cost estimates			Ongoing non-staff cost increase estimate
	Low	Med	High	Low	Med	High	
Physics	$2,800.0	$5,487.4	$8,889.6	$1,067.8	$3,670.5	$8,529.1	$174.2
Mathematics	$5,425.0	$7,255.8	$9,059.2	$1,128.0	$2,349.7	$3,299.1	$55.9
Earth Science	$5,430.0	$8,652.6	$11,324.0	$1,186.2	$3,465.6	$7,687.3	$163.4

V. Conclusions

Each of the three target disciplines discussed would be a reasonable target for an expansion in scope of the ADS. While all three have existing subject-specific (subscription) databases with reasonable user bases operated by societies within the discipline, none of those databases offers the same efficiency gains to their users that ADS currently does for astronomers. Each discipline has unique characteristics that affect how a potential expansion of scope would impact and be received by that scientific community, as well as how much it would cost to undertake.

Physics

From a disciplinary standpoint, physics is the closest to astronomy and astrophysics, and as such, may represent the next logical discipline in which to expand the ADS. Additionally, because the two disciplines are so closely connected, the vast majority of the physics literature is already being indexed and included in the ADS, and collaborations have already been forged between the ADS and physics societies and information organizations. However, physicists may be a difficult community of researchers to engage with a new resource like an expanded ADS: they are busy and feel well-served by existing resources, and the discipline is widely distributed such that there are fewer immediate, highly impactful, obvious conversions in the area of linked data which the database could provide. Because the ADS already indexes much of the physics literature, the startup cost for the expansion would be low, but with such a large community of researchers worldwide, the ongoing costs of user support and relationship development would be larger than other fields.

Mathematics

Of the disciplines investigated in this project, mathematics currently has the least coverage by the ADS. Additionally, from a disciplinary standpoint, mathematics is notably separate from astronomy and fairly self-contained in its own right. While it may superficially seem like a less logical first choice for expansion than physics or earth science, the field of mathematics, particularly the IMU along with the NAS, is currently exploring what a next-generation linked-data mathematics digital library would look like, and has espoused many concepts and objectives currently implemented in the ADS. As such, more so than other fields, the ADS may find interest and cooperation from a societal and organizational perspective (and therefore indirectly from an individual researcher perspective) for an expansion into mathematics. Additionally, the investigations to date of the IMU and NAS have already enumerated areas of linked mathematical data into which the ADS could venture, including linking mathematical concepts and objects such as general ideas, specific theorems, and equations. They have also identified further areas of research, such as programmatically or computationally identifying these concepts and objects within the literature, an area that the ADS is already starting to work in with the new Bumblebee interface. While the startup cost for the ADS to move into mathematics would be rather high because the field is not well-covered by the ADS currently, the ongoing costs of support would likely be the lowest of the three target disciplines because of the relatively small, compact community.

Earth Science

While the discipline of earth science is perhaps less connected to astronomy and astrophysics than the discipline of physics, there are still strong connections between the two fields, particularly in the area of planetary sciences. Additionally, NASA does support research in both areas. As such, the ADS already indexes a small amount of the earth science literature, and ADS staff maintain a moderate level of cooperation with societies in the field, regularly attending AGU meetings and the like. Additionally,

earth science offers some specific opportunities which would be especially impactful to researchers, including the indexing and potential scanning and archiving of the large collections of gray literature in the field as well as the implementation of search functionality based on locations or geographical coordinates. These potential improvements over current systems could win ADS quick support from researchers and societies in the discipline, perhaps more than physics or mathematics. However, because the literature is not well-covered in the ADS currently and the community is large and rather distributed, both the startup and ongoing costs for the ADS to move into earth science would likely be high.

Moving forward with expansion

Strategically speaking, were the ADS to move forward with expanding its scope, expanding to all three disciplines at once would likely be infeasible: the startup costs and staff hiring for each discipline would be compounded, the simultaneous outreach efforts would likely be unsustainable, and the rapid growth could destabilize the organization. Instead, a more reasonable approach would be to expand to one discipline at a time, using the lessons and successes of each discipline to inform the approach to the next, and to use as a proof of concept to stakeholders moving forward. The appropriate order in which to approach this expansion is dependent on the goals of the Harvard Library, the amount and type of effort and resources they are willing to expend, and the successes in negotiating with disciplinary stakeholders for support of expansion into each discipline.

VI. Bibliography

American Astronomical Society. “What Is the AAS?” Accessed July 1, 2016. https://aas.org/about/what-aas.

American Mathematical Society. “Announcing Mathematical Reviews,” 1940. http://www.ams.org/publications/mr-1940-promo.pdf.

“arXiv.org E-Print Archive.” Accessed June 28, 2016. https://arxiv.org/.

“arXiv.org Help - arXiv Submission Rate Statistics.” Accessed June 24, 2016. https://arxiv.org/help/stats/2015_by_area/index.

Bichteler, J, and D Ward. “Information-Seeking Behavior.” *Special Libraries* 80, no. 3 (1989): 169–78.

Bichteler, Julie. “Geologists and Gray Literature.” *Science & Technology Libraries* 11, no. 3 (May 14, 1991): 39–50. doi:10.1300/J122v11n03_04.

Brown, Cecelia M. “Information Seeking Behavior of Scientists in the Electronic Information Age: Astronomers, Chemists, Mathematicians, and Physicists.” *Journal of the American Society for Information Science* 50, no. 10 (July 19, 1999): 929–43.

Claspy, William P. “Information Use in Astronomy,” 153:177, 1998. http://adsabs.harvard.edu/abs/1998ASPC..153..177C.

Committee on Planning a Global Library of the Mathematical Sciences. “Developing a 21st Century Global Library for Mathematics Research.” *arXiv:1404.1905 [Math]*, April 4, 2014. http://arxiv.org/abs/1404.1905.

“Geologic Guidebooks of North America Database | American Geosciences Institute.” Accessed June 22, 2016. http://www.americangeosciences.org/georef/geologic-guidebooks-north-america-database.

Hemminger, Bradley M., Dihui Lu, K.t.l. Vaughan, and Stephanie J. Adams. “Information Seeking Behavior of Academic Scientists.” *Journal of the American Society for Information Science and Technology* 58, no. 14 (December 1, 2007): 2205–25. doi:10.1002/asi.20686.

International Astronomical Union. “Geographical Distribution of Individual Members.” Accessed July 1, 2016. https://www.iau.org/administration/membership/individual/distribution/.

Jamali, Hamid R., and David Nicholas. “Information-Seeking Behaviour of Physicists and Astronomers.” *Aslib Proceedings* 60, no. 5 (2008): 444–62. doi:http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1108/00012530810908184.

Kurtz, Michael J. Kurtz, Guenther Eichhorn, Alberto Accomazzi, Carolyn S. Grant, Stephen S. Murray, and Joyce M. Watson. “The NASA Astrophysics Data System: Overview.” *Astronomy and Astrophysics Supplement Series* 143, no. 1 (2000): 19. doi:10.1051/aas:2000170.

Mounts, M. “MathSciNet.” *Choice: Current Reviews for Academic Libraries* 48, no. 12 (August 2011): 2352–53.

Murray, Stephen S., Alberto Accomazzi, and Michael J. Kurtz. “Astrophysics Archives Programmatic Review 2015: The NASA Astrophysics Data System,” March 16, 2015.

National Science Foundation. “Science and Engineering Indicators: 2000.” Accessed June 29, 2016. http://wayback.archive-it.org/5902/20160210155318/http://www.nsf.gov/statistics/seind00/.

Tenopir, Carol, Donald W. King, Peter Boyce, Matt Grayson, and Keri-Lynn Paulson. “Relying on Electronic Journals: Reading Patterns of Astronomers.” *Journal of the American Society for Information Science and Technology* 56, no. 8 (June 1, 2005): 786–802. doi:10.1002/asi.20167.

Thomson Reuters. “2015 Journal Citation Reports®,” 2016.

“USGS National Geologic Map Database.” Accessed June 22, 2016. http://ngmdb.usgs.gov/ngmdb/ngmdb_home.html.

van Leeuwen, Thed, and Robert Tijssen. “Interdisciplinary Dynamics of Modern Science: Analysis of Cross-Disciplinary Citation Flows.” *Research Evaluation* 9, no. 3 (December 1, 2000): 183–87. doi:10.3152/147154400781777241.

VII. Appendices

Appendix 1: Web of Science Subject Categories

In evaluating publication output volume and interdisciplinary nature of each subject using the Web of Science, the following mapping was used from the Web of Science subject category to the more general discipline referred to in this report.

Discipline described above	WoS Subject Categories
Astronomy and Astrophysics	Astronomy & Astrophysics
Physics	Biophysics Optics Physics, Applied Physics, Atomic, Molecular & Chemical Physics, Condensed Matter Physics, Fluids & Plasmas Physics, Mathematical Physics, Multidisciplinary Physics, Nuclear Physics, Particles & Fields
Mathematics	Logic Mathematical & Computational Biology Mathematics Mathematics, Applied Mathematics, Interdisciplinary Applications Statistics & Probability
Earth Science	Ecology Energy & Fuels Environmental Sciences Environmental Studies Geochemistry & Geophysics Geology Geosciences, Multidisciplinary Meteorology & Atmospheric Sciences Oceanography Water Resources
Computer Science (for comparative purposes only)	Computer Science, Artificial Intelligence Computer Science, Cybernetics Computer Science, Hardware & Architecture Computer Science, Information Systems Computer Science, Interdisciplinary Applications Computer Science, Software Engineering Computer Science, Theory & Methods
Engineering (for comparative purposes only)	Acoustics Engineering, Aerospace Engineering, Biomedical Engineering, Chemical Engineering, Civil Engineering, Electrical & Electronic

	Engineering, Environmental Engineering, Geological Engineering, Industrial Engineering, Manufacturing Engineering, Marine Engineering, Mechanical Engineering, Multidisciplinary Engineering, Ocean Engineering, Petroleum Instruments & Instrumentation Mechanics Metallurgy & Metallurgical Engineering Mineralogy Mining & Mineral Processing Nanoscience & Nanotechnology Robotics Thermodynamics Transportation Transportation Science & Technology
Multidisciplinary (for comparative purposes only)	Multidisciplinary Sciences
Nuclear Science (for comparative purposes only)	Nuclear Science & Technology

Appendix 2: Discipline Data Sheet

	Astronomy/ Astrophysics	Physics	Mathematics	Earth Science
Overall Corpus Size	2.2M (ADS) 0.7M (WoS)	8.0M (ADS) 6.0M (WoS) 9.1M (Inspec)	3.3M (MSN) 3.4M (WoS)	3.9M (GeoRef) 1.5M (WoS)
Growth per year (2010-2013)	61.1K (ADS) 24.0K (WoS)	310K (ADS) 211K (WoS) 334K (Inspec)	104K (MSN) 69.8K (WoS)	82.2K (GeoRef) 56.6K (WoS)
arXiv overall corpus[63]	179K (astro-ph)	564K (hep + physics + gr-qc + quant-ph + cond-mat + nucl)	238K (math + math-ph)	N/A
arXiv growth per year (~90K/year total)	12.9K	41.9K	28.8K	N/A
Community size	IAU – 12.5K AAS – 7K	APS – 51K OSA – 19K	AMS – 30K ASA – 18K SIAM – 14K	AGI – 250K GSA – 26K AGU – 60K AAPG – 40K
Interdisciplinarity factor – WoS (2010-2014)	High w/ Physics (30%); Low w/ all others (<10%)	High w/ Engineering (26%); Low w/ all others (<10%)	Moderate w/ CS (14%), Engineering (12%); Low w/ all others (<10%)	Moderate w/ Engineering (17%); Low w/ all others
Interdisciplinarity factor – citation (van Leeuwen paper)[64]	14.3	59.9	26.1	>60
Gray lit factor (lit/db review)	**Moderately High** Observing proposals; source code; data tables;	**Moderate** Strong pre-print culture, but already maintained in arXiv. Reports well-maintained in INSPIRE. Lab report series.	**Low** Not overly dependent on conferences; Math. Rev. well cared-for. Heavier reliance on personal communication	**High** OFRs from surveys; field trip guidebooks; maps; technical reports
WoS Conference % (2010-2014)	18.7%	22.7%	11.9%	9.7%
Subject-specific conference %		18% (Inspec)	9.2% (MSN)	45.5% (GeoRef)
Data linkage factor (lit/db review)	**High** Astronomical objects; observatory bibliographies;	**Moderate** Lab bibliographies; materials data	**Moderate** Potential for linkage of researchers; connection of mathematical concepts and objects	**High** Location data; local material; newsletter references

[63] "arXiv.org Help - arXiv Submission Rate Statistics."

[64] van Leeuwen and Tijssen, "Interdisciplinary Dynamics of Modern Science."