

# Scholarly Communication in High-Energy Physics: Past, Present and Future Innovations

---

ROBERT AYMAR

CERN, European Organization for Nuclear Research, CH1211,  
Genève 23, Switzerland.

Unprecedented technological advancements have radically changed the way we communicate and, at the same time, are effectively transforming science into e-science. In turn, this transformation calls for an evolution in scholarly communication. This review describes several innovations, spanning the last decades of scholarly communication in High Energy Physics: the first repositories, their interaction with peer-reviewed journals, a proposed model for Open Access publishing and a next-generation repository for the field. We hope that some of these innovations, which are deeply rooted in the highly-interconnected and worldwide flavour of the High-Energy Physics community, can serve as an inspiration to other communities.

## 1. Introduction

The invention by CERN's Tim Berners-Lee in 1991 of what quickly became known as the World-Wide Web constitutes a turning point that often is compared to Gutenberg's development of the printing press.<sup>1</sup> This year marks the 15th anniversary of the release in the public domain of the World-Wide Web software by the CERN management.<sup>2</sup> The communication tools that we have at our fingertips today bear little resemblance to those of two decades ago. This is a dramatic social change, yet this revolution so far seems to have had only a limited impact on the patterns of scholarly communication. Notwithstanding the advances in electronic publishing, the core of scholarly communication has not really changed in itself: most of the changes we have seen so far limit themselves to electronic representations of what was earlier done on paper. The electronic information seems sometimes to be just a clone of the paper-based era.

Progress in communication technology has resulted in the opportunity for more researchers worldwide to potentially have access to more scientific results, enabling them to generate, in turn, more scientific results for a global progress. The next steps, which are currently being explored, will enable us to achieve scientific progress in totally new ways. In some branches of science we see already now examples of distributed research by geographically spread individual and research groups. The new way of working is often referred to as *e-science*, intended as both the solution of problems requiring distributed computing solutions or the invention of new research techniques based on data mining.

This unprecedented innovation makes the accessibility of scientific results a fundamental issue in scientific progress. At the same time the concept itself of scientific results has grown to encompass publications, algorithms, data, software, and all the elements of the knowledge-generation process. This is one of the main origins of the recent debate on Open Access, which has become mainstream, spreading to all areas and actors of scholarly communication and affecting its entire spectrum, from policy making to financial aspects.<sup>3</sup> Open Access models are actively being proposed by scholars, libraries and publishers alike, and Open Access definitions, of varying shades and colours, are actively – and frequently antagonistically – argued. A review of, or an insight into, this debate, is beyond the scope of this contribution, which will rather mention the historical and pragmatic perspective that has indissolubly linked the way High-Energy Physics (HEP) scientists communicate based on the Open Access principles.

This contribution, then, discusses the way the HEP community has faced, in the last few decades, challenges in scholarly communication, and how it proposed solutions that were ahead of their time, and that eventually became mainstream in the present evolution of the scientific process. The intention is to elaborate on a scientist-driven approach to the publishing and library landscape in this discipline, its tradition, its present evolution, and its possible future. We hope that this sectoral and possibly polarized viewpoint may serve as inspiration for a wider audience that will recognise some of the challenges and could consider some of the solutions promoted by the HEP community. This article is structured into seven parts. After the present introduction, Section 2 offers a brief contextualization of HEP as a scientific discipline. Section 3 discusses the history of innovations in scholarly communication in HEP. Section 4 traces the first steps of Open Access publishing in the field. Section 5 describes an ongoing project that is the HEP community's answer to the Open Access debate: SCOAP3. Section 6 traces the path of the next innovation in information provision in HEP as an indispensable tool for scientific process: INSPIRE. Section 7 offers a conclusion and some further remarks.

## 2. A short description of High-Energy Physics

The scientific goals of HEP are to unveil the intimate constituents of matter and to probe their interactions. This is a quest as old as science, which today aims to attain a fundamental description of the laws of physics and the evolution of the universe, to explain the origin of mass and to understand the dark matter in the universe. True to the scientific process, HEP is an experimental and a theoretical science, with a community split roughly into two halves: experimental physicists and theoretical physicists. Experimental HEP scientists join in thousand-strong collaborations to build the largest instruments ever, aiming to reproduce on Earth – through high-energy hadron collisions in vacuum – the energy densities of the universe at its birth. At the same time, theoretical particle physicists are linked in global networks through which they collaborate in formulating hypotheses and theories aimed to predict and interpret experimental findings.

HEP experimental research takes place mainly in international accelerator research centres based, for example, in Europe, such as the European Organization for Nuclear Research (CERN) in Geneva or the Deutsches Elektronen-Synchrotron (DESY) in Hamburg; in the United States, mainly at the Stanford Linear Accelerator Center (SLAC) in California and the Fermi National Accelerator Laboratory (Fermilab) in Illinois; and in Japan at the High Energy Accelerator Research Organization (KEK) in Tsukuba. HEP theoretical research takes place in hundreds of universities and institutes worldwide, which also host the experimental teams building parts of the large detectors that are used at the large accelerator laboratories, and analyses the data these detectors collect.

The crown jewel in HEP research is CERN's Large Hadron Collider (LHC), which started accelerating particles in 2008, after more than a decade of construction. This 27 km-long accelerator, hosted in a tunnel as deep as 100 m underground, and which operates at a temperature of  $-271^{\circ}\text{C}$ , will collide two 7 TeV proton bunches 40 million times a second. These collisions will be observed by large detectors, up to the size of a five-storey building, crammed with electronic sensors: think of a 100 megapixel digital camera taking 40 million pictures a second.

The LHC programme epitomizes the spirit and the challenge of HEP, beyond its scientific goals and achievement. The LHC programme is at the technological frontier, and has required the invention, design and deployment of tools in engineering and information technology that did not exist at the time of the proposal of the scientific goals of the project. This has only been possible through international collaboration: tens of thousands of scientists and engineers from over 80 countries have contributed to the design and the construction of the LHC machine and its detectors, in what is possibly the largest scientific collaborative effort in history. This progress has been made possible by a powerful synergy between academia and industry for a cutting-edge R&D programme.

And it has consequences in scholarly communication: in order to honour the effort of each of the collaborators, every single contributor will be included in the list of authors of the articles that will derive from the experiments. Some articles might include over 2000 names, affiliated to a range of institutes. Such a large number of co-authors might appear absurd to other communities. However, in a competitive research world it is of extreme importance that the academic investment is recognized and that all participating scientists, along with their corresponding institutions, can be linked individually to the results achieved by their respective collaborations.

The LHC is only the latest example of the technological innovations that have made possible the success of HEP, and which are solidly based on international collaboration. Many of those innovations have spread to other areas of research and industry: from our daily communication tools, thanks to the world-wide web, to medical imaging; from particle accelerators now used in cancer therapy, to computing grids now used for crucial research in the life sciences. True to its spirit, this contribution will identify some innovations in scholarly communication that originated in the same collaborative and trans-national matrix of HEP research and which have then spread to other disciplines, as described in the next section.

### **3. A tradition of innovation in scholarly communication**

The leitmotif of HEP innovation in scholarly communication is its *preprint* culture.<sup>4</sup> For decades, theoretical physicists and scientific collaborations, eager to disseminate their results in a faster way than the distribution of conventional scholarly publications, took to printing and mailing hundreds of copies of their manuscripts at the same time as submitting them to peer-reviewed journals. The very first preprint repository in the world was set up at CERN in the late 1950s. It included working papers and reports submitted to CERN by authors from institutions across the world. It can still be seen today, as shown in Figure 1. In its 40 years of existence it grew to occupy several dozens filing cabinets, with a ‘traditional’ index cabinet for searching for author and title. Its growth stopped more than a decade ago, when authors turned to electronic submissions at the dawn of the arXiv era. The paper copies are now gradually being scanned and put online, providing Open Access to some documents that would not be accessible otherwise. This *ante-litteram* form of ‘author-pays’ or rather ‘institute-pays’ Open Access assured the broadest possible dissemination of scientific results, albeit privileging scientists working in affluent institutions. These researchers could afford the mass mailings and were most likely to receive a copy of preprints from other scientists eager to advertise their results. At the same time, for research-intensive institutions, preprint dissemination came at a cost: in the



**Figure 1.** The CERN preprint catalogue and the corresponding repository, maintained from 1954 to 1994.

early 1990s, CERN used to spend over 1 million Swiss francs a year for printing and mailing expenses.

Against this background, three innovations mark crucial advances in scholarly communication in HEP.

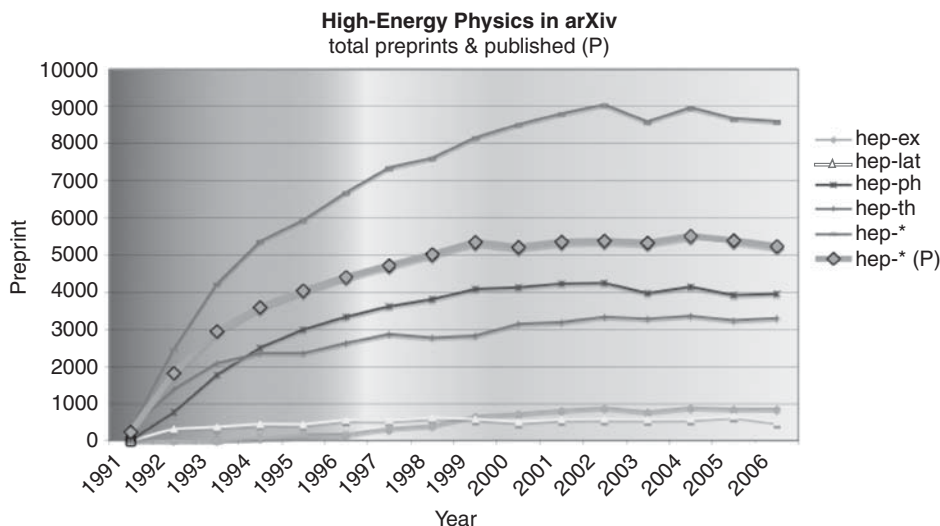
- (1) The SPIRES database, the first grey-literature electronic catalogue, saw the light at the SLAC (Stanford Linear Accelerator Center) HEP laboratory in Stanford, California, in 1974. It listed preprints, reports, journal articles, theses, conference talks and books and it now contains metadata for about 760,000 HEP articles.<sup>5</sup> A recent poll of HEP scholars has shown that SPIRES, in symbiosis with arXiv, is an indispensable tool in their daily research workflow.<sup>6</sup>
- (2) arXiv, the archetypal subject repository, was conceived in 1991 by theoretical physicist Paul Ginsparg, then at LANL (Los Alamos National Laboratory) in New Mexico, USA.<sup>7</sup> It evolved the four-decade old preprint culture into an electronic system, offering all scholars a level playing field from which to access and disseminate information. With  $\frac{1}{2}$  million articles, today arXiv has grown outside the field of HEP, becoming the reference repository for many disciplines: from mathematics to some areas of biology.

- (3) The invention of the web at CERN is a household story.<sup>1,2</sup> What is less known is that the first web server outside Europe was installed at SLAC in December 1991 to provide access to the SPIRES database – an example of the potential of the web.<sup>8</sup> HEP scholars imagined the web from its inception as a tool for scholarly communication. The interlinking of arXiv and SPIRES in summer 1992 eventually offered the first web-based Open Access application.

Thanks to its decade-old preprint culture, HEP is today an almost entirely ‘green’ Open Access discipline, i.e. a discipline where authors self-archive their research results in repositories that guarantee their unlimited circulation. Posting an article on arXiv, even before submitting it to a journal, is common practice. Even revised versions incorporating the changes due to the peer-review process are routinely uploaded.

It is interesting to remark that this success of ‘green’ Open Access in HEP originated without mandates and without debates: very few HEP scientists would not take advantage of the formidable opportunities offered by the repository of information in the discipline. This is a very interesting observation in recent times, when institutional repositories are multiplying and many institutions envisage mandates for the submission of documents. These are powerful tools to ensure capturing content in repositories, but clear added value for researchers is a more enticing way to ensure the pervasive spread of ‘green’ Open Access. The speed of adoption of arXiv in the field is presented in Figure 2, which plots the evolution in time of submissions to arXiv in the four categories in which HEP results are conventionally divided. The number of preprints that are subsequently published in peer-review journals is also indicated. The difference between the numbers of submissions and the published articles is mostly due to conference proceedings and other grey-literature material that is routinely submitted to arXiv, but which does not usually generate peer-reviewed publications.

With hindsight, it is interesting to look at the discussions in the early 1990s covering the role of arXiv in scholarly communication in HEP, and its potentially disruptive consequences for journals. Doomsday predictions of the time did not materialize, the two information outlets thrive side by side, each serving a different scope: the immediacy of dissemination without barriers in the first; quality certification in the second. Today, publishers of HEP journals not only no longer oppose the arXiv, or object to researchers posting post-peer reviewed author-formatted versions of their publications, but even host mirrors of this popular system. In addition, arXiv has facilitated the journal editorial workflow: in some cases the only information that is required by a journal upon submission is the arXiv number, automatic systems then recover and reformat the relevant files from arXiv.



**Figure 2.** HEP preprints submitted to arXiv in four different categories (hep-ex, hep-lat, hep-ph and hep-th) as well as total numbers (hep-\*). Preprints subsequently published in peer-reviewed journals are indicated with a ‘P’. After a phase of adoption of the arXiv system, corresponding to the rise of all curves, present outputs are constant. Data from the SPIRES database.

It cannot be denied that, as a consequence of the widespread role of arXiv, journals have, to a large extent, lost their century-old role as vehicles of scholarly communication. However, at the same time, they continue to play a crucial part in the HEP community. Evaluation of research institutes and (young) researchers is largely based on publications in prestigious peer-reviewed journals. The main role of journals in HEP is mostly perceived as that of ‘keeper-of-the-records’, by guaranteeing a high-quality peer-review process. In short, it can be argued that the HEP community needs high-quality journals as its ‘interface with officialdom’. At the same time, the coexistence of arXiv and the peer-reviewed journals in HEP has accelerated the debate on Open Access publishing in this field, which is described in the next section.

#### 4. Open Access publishing

Open Access journals existed in HEP over a decade ago. In 1997, the *Journal of High Energy Physics (JHEP)*, published by the International School of Advanced Studies (SISSA) in Trieste, Italy, was launched. The journal was free to read online and without publication fees for authors. The income was meant to come from libraries purchasing bound annual volumes intended as archival copies. However, this business model turned out not to be sustainable, and the journal

then became a low-cost subscription journal. Recently, it acceded to yet another Open Access model, with a very successful institutional membership scheme where, for a small fee, all articles originating from a contributing institution are Open Access. *JHEP* was followed in 1998 by *Physical Review Special Topics Accelerators and Beams*, published by the American Physical Society (APS), which operates under a sponsorship scheme, with 14 research institutions sponsoring the operation of this niche journal. The *New Journal of Physics*, published by the Institute of Physics Publishing (IOPP), which carries HEP content in a broader spectrum that covers many branches of physics, also started in 1998. It is financed by author fees, under the so-called ‘author-pays’ model.

After preprints, arXiv, the web and a few experimental journals published under an Open Access scheme, a full transition to Open Access journals appears to be the next logical step in the natural evolution of HEP scholarly communication. This vision is obviously shared with the publishing industry: most HEP publishers, Springer as of 2004 and, more recently, APS and Elsevier, now offer authors the option of paying an additional fee on top of the library subscription to make their single articles Open Access, under the so-called ‘hybrid model’. As from 2007 a couple of new HEP journals, *Physics A*, published by PhysMathCentral, a spin-off of BioMedCentral, and *Advances in High-Energy Physics*, published by Hindawi Publishing, have entered the market, fully based on publication fees to be met by the author.

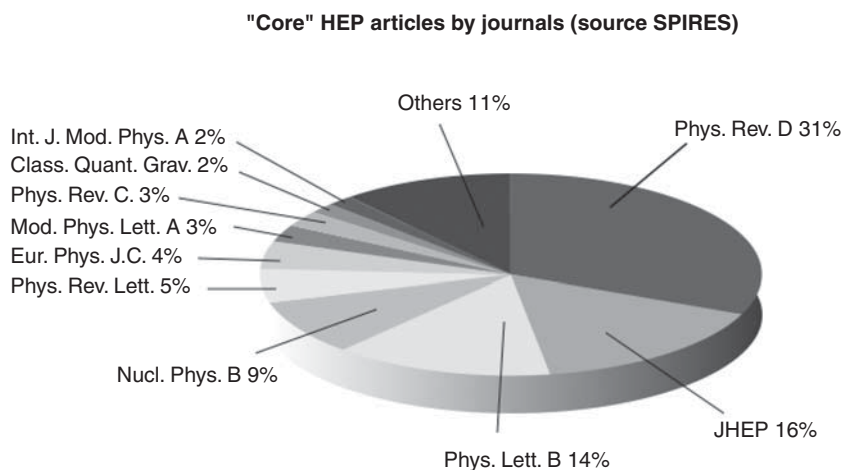
The ‘author-pays’ and ‘hybrid’ schemes, however, are not very popular: the total number of HEP articles that appear as Open Access under these two schemes is below 1% of the yearly HEP literature. In comparison, the volume of Open Access articles financed by the institutional membership fee in *JHEP* is about 20% of this journal, corresponding to about 4% of the total volume of HEP articles.

The next two sections of this contribution will discuss the future of Open Access publishing in HEP, through the proposed SCOAP3 initiative, and the status and plans for the deployment of a platform to grant improved access and functionalities to information in HEP, through a successor of the SPIRES system built on the popularity of the arXiv content: INSPIRE.

## 5. The SCOAP3 project

The aim of the SCOAP3 project is to convert the entire HEP literature to Open Access. This is the first project of this kind, as recent Open Access business models have consisted either of single journals or suites of journals from one publisher migrating to Open Access, or being born Open Access, or achieving Open Access for the entire output of a single institution or group of institutions, possibly in the journal portfolio of a given publisher.

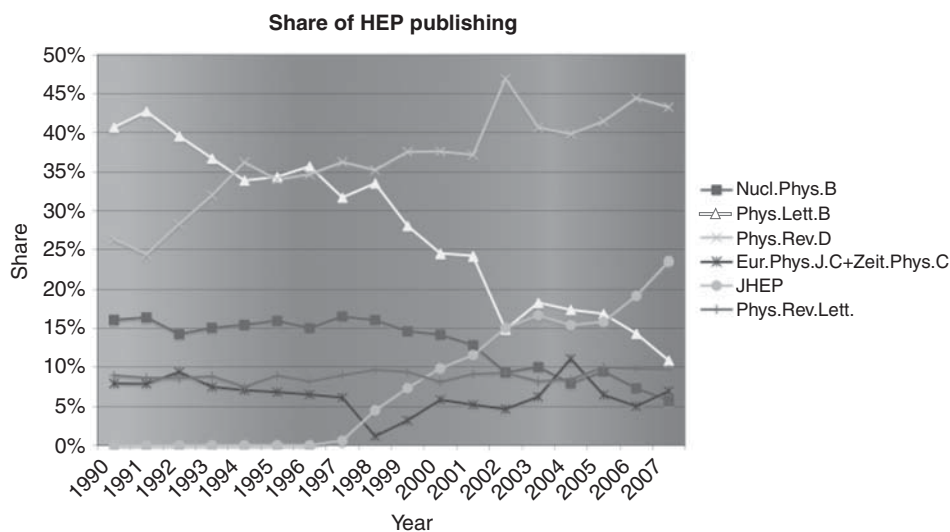




**Figure 3.** Journals favoured by HEP scientists in 2006. Journals that attracted less than 75 HEP articles are grouped in the slice named ‘Others’. Data from the SPIRES database.

Some background details are useful to put the initiative in context, as unveiled by in-depth studies of the HEP publication landscape that have informed the design of this model<sup>9–11</sup> and which are discussed in the following. As shown in Figure 2, yearly about 20,000 scholars publish about 6000 HEP articles. Of these, about 80% are produced by theoretical physicists and 20% by large collaborations of experimental physicists.

Figure 3 presents the journals favoured by HEP authors in 2006. About 80% of HEP articles are published in just six peer-reviewed journals from four publishers. Five of those six journals carry a majority of HEP content. These are *Physical Review D* (published by the APS), *Physics Letters B* and *Nuclear Physics B* (Elsevier), *JHEP* (SISSA/IOPP) and the *European Physical Journal C* (Springer). The sixth journal, *Physical Review Letters* (APS), is a ‘broadband’ journal that carries only about 10% of HEP content. These journals have long been favoured by HEP scholars, albeit with varying fortunes. Figure 4 presents the percentage of HEP articles published in each of these six journals in the last 17 years. Only the articles published in these journals are considered in this graph. Periods of stability are followed by a fast increase of some titles and a corresponding decline of others. The origins of these changes can be traced to the capacity of the different publishers to respond to the changing expectations of the authors. Some examples from a past before the period presented in Figure 4 are discussed in an essay<sup>12,13</sup> describing the emergence of North Holland, now Elsevier, as a successful publisher in HEP thanks to its introduction of a letter journal (*Physics Letters*) and a review journal (*Physics Reports*) in the late 1960s and early 1970s – editorial solutions which, at the time, were perfectly in tune



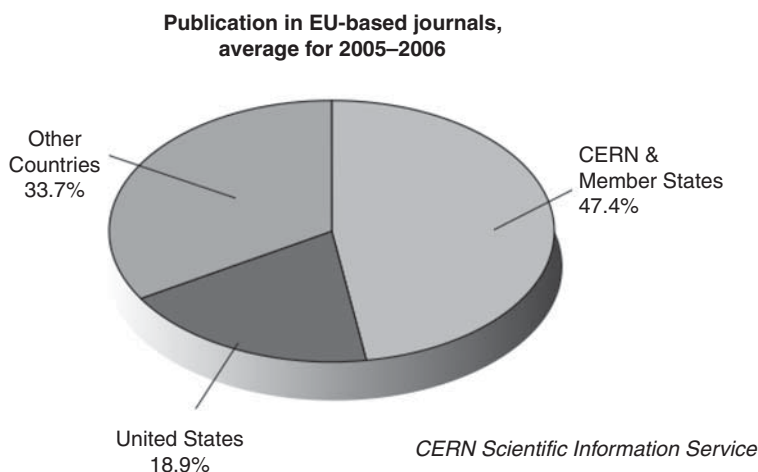
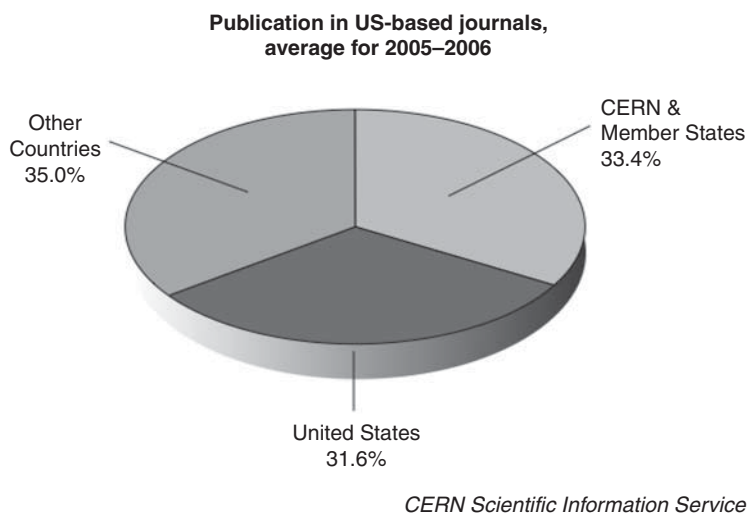
**Figure 4.** Journals favoured by HEP scientists in the last 18 years. For each year, only articles published in these six journals are considered, and the relative fractions are displayed. Articles published in *Zeitschrift für Physik C* and the *European Physical Journal C* are aggregated, as the latter is a successor of the former. Data from the SPIRES database.

with the needs of the authors. The trends in the period covered in Figure 4 can be attributed, from left to right, to:

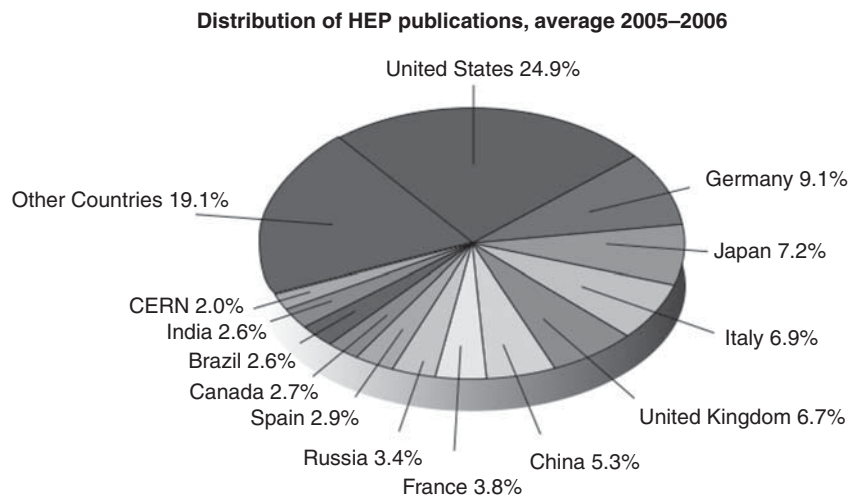
- An increase in the number of submissions to *Physical Review D* following the stabilization of its policy of page charges.
- A decrease in the number of submissions to *Physics Letters* following the emergence of *JHEP*, aggressively marketing itself as an ‘author-friendly’ publication outlet.
- An increase in the number of submissions to *JHEP* following its recent, popular, Open Access scheme.

It is interesting to remark that in a discipline such as HEP, with traditionally strong cross-border collaborative links, journals published in the United States or in Europe attract contributions from all geographical regions, as presented in Figure 5. Therefore any Open Access initiative can only succeed if it is truly global in scope.

Figure 6 presents the contribution by country to the HEP scientific literature. Co-authorship is taken into account on a pro-rata basis, assigning fractions of each article to the countries in which the authors are affiliated. In the case of authors with a double affiliation, large laboratories and the countries with the larger GDP per capita are used as a country of affiliation. This study is based on



**Figure 5.** Geographical origin of publications in HEP journals based in the United States and in Europe. Co-authorship is taken into account on a pro-rata basis, assigning fractions of each article to the countries in which the authors are affiliated. This study is based on all articles published in the years 2005 and 2006 in five HEP ‘core’ journals: *Physical Review D* (US), *Physics Letters B* (EU), *Nuclear Physics B* (EU), *Journal of High Energy Physics* (EU) and the *European Physical Journal C* (EU), and the HEP articles published in two ‘broadband’ journals: *Physical Review Letters* (US) and *Nuclear Instruments and Methods in Physics Research A* (EU).<sup>11</sup> The European contribution is well represented by CERN and its Member States, which are: Austria, Belgium, Bulgaria, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, the Netherlands, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland and the United Kingdom.



*CERN Scientific Information Service*

**Figure 6.** Contributions by country to the HEP scientific literature published in the largest journals in the field. Co-authorship is taken into account on a pro-rata basis, assigning fractions of each article to the countries in which the authors are affiliated. Countries with individual contributions less than 2% are aggregated in the ‘Other countries’ category.<sup>10,11</sup>

all articles published in the years 2005 and 2006 in five HEP ‘core’ journals: *Physical Review D*, *Physics Letters B*, *Nuclear Physics B*, *JHEP* and the *European Physical Journal C*; and the HEP articles published in two ‘broadband’ journals: *Physical Review Letters* and *Nuclear Instruments and Methods in Physics Research A*. A total sample of almost 11,300 articles is considered.<sup>10,11</sup>

SCOAP3, the Sponsoring Consortium for Open Access Publishing in Particle Physics, aims to convert to Open Access the HEP peer-reviewed literature in a way that is transparent to authors,<sup>10,14</sup> meeting the expectations of the HEP community for peer-review of the highest standard, and administered from the journals that have served the field for decades, while leaving room for new players. The SCOAP3 business model originates from a two-year debate involving the scientific community, libraries and publishers.<sup>10,15</sup> The essence of this model is the formation of a consortium to sponsor HEP publications and make them Open Access by redirecting funds that are currently used for subscriptions to HEP journals. Today, libraries and the funding bodies behind them purchase journal subscriptions to implicitly support the peer-review and other editorial services and to allow their users to read articles, even though – in HEP – scientists mostly access their information by reading preprints on arXiv. The SCOAP3 vision for tomorrow is that funding bodies and libraries worldwide would federate in a consortium that will pay centrally for the peer-review and

other editorial services, through a re-direction of funds currently used for journal subscriptions, and, as a consequence, articles will be free to read for everyone. This evolution of the current ‘author-pays’ Open Access models will make the transition to Open Access transparent for authors, by removing any financial barriers.

The SCOAP3 model offers another advantage for libraries and funding bodies over the present ‘author-pays’ model. Indeed, disciplines with successful ‘author-pays’ journals often see publication costs met either by libraries or by funding bodies. At the same time, the costs of the subscriptions to ‘traditional’ journals do not decrease following the reduced volume of articles that these publish, due to the drain towards ‘author-pays’ Open Access journals, resulting in a global larger expenditure. Some parties in the Open Access debate often present this as a scarecrow. Conversely, the SCOAP3 model aims to convert to Open Access all the literature in a field, keeping therefore the total expenditure under control.

In practice, the Open Access transition proposed by the SCOAP3 model will be facilitated by the fact that the large majority of HEP articles are published in just six peer-reviewed journals from four publishers, as presented in Figure 3. Five of those six journals carry a majority of HEP content and the aim of the SCOAP3 model is to assist publishers to convert these ‘core’ HEP journals entirely to Open Access. It is expected that the vast majority of the SCOAP3 budget will be spent to achieve this target. In addition, SCOAP3 will sponsor the conversion to Open Access of the HEP fraction of ‘broadband’ journals. Of course, the SCOAP3 model is open to any other, present or future, ‘core’ or ‘broadband’, high-quality journals carrying HEP content, beyond those highlighted here.

The price of an electronic journal is mainly driven by the costs of running the peer-review system and editorial processing. Most publishers quote a price in the range of €1000–2000 per published article. On this basis, given that the total number of HEP publications in high-quality journals is between 5000 and 10,000 (according to how one defines HEP and its overlap with cognate disciplines), the annual SCOAP3 budget for the transition of HEP publishing to Open Access would amount to a maximum of €10 million per year.<sup>10</sup> The costs of SCOAP3 will be distributed among all countries according to a fair-share model based on the distribution of HEP articles per country, as shown in Figure 6. In practice, this is an evolution of the ‘author-pays’ concept: countries will be asked to contribute to SCOAP3, whose ultimate targets are Open Access and peer-review, according to their use of the latter. To cover publications from scientists from countries that cannot reasonably be expected to make a contribution to the consortium at this time, an allowance of not more than 10% of the SCOAP3 budget is foreseen.

SCOAP3 will sponsor articles through contracts with publishers of high-quality HEP journals. The conditions of these contracts will be established through a tendering procedure: publishers will be invited to bid for their

peer-review and other editorial services. The consortium will then evaluate these offers as a function of indicators, such as the journal quality and price, and attribute contracts within its capped budget envelope. SCOAP3 therefore has the potential to contain the overall cost of journal publishing by linking price, volume, and quality, and injecting competition into the market. These aspects are not present in today's subscription model.

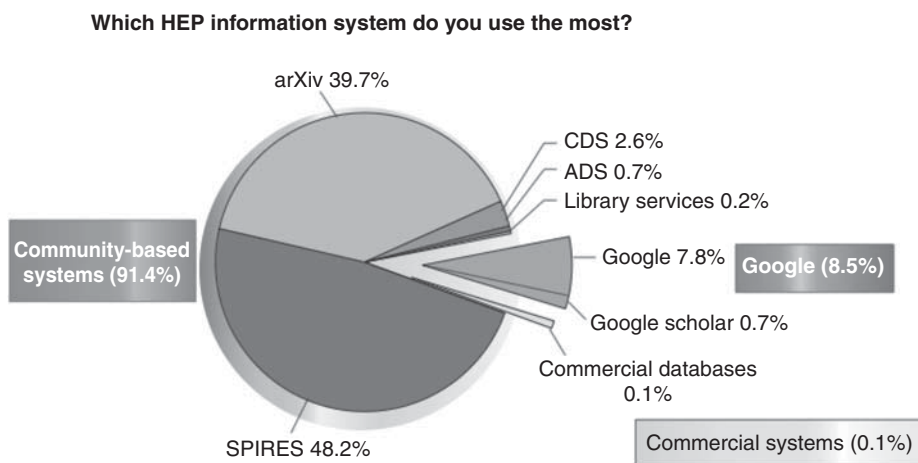
In the SCOAP3 model, libraries will not be paying twice for the journals to be converted to Open Access: the principle of the model is to finance the peer-review and the other editorial services, achieving Open Access, through the re-direction of subscription funds. In the case where the journals that will receive a contract from SCOAP3 are part of a large journal licence package, the publishers will be contractually required to extract these titles from the corresponding packages and to reduce the subscription cost for the remainder of the package.

It appears at first glance to be a formidable enterprise to organize a worldwide consortium of research institutes, libraries and funding bodies that cooperates with publishers in converting the most important HEP journals to Open Access. At the same time, HEP is a perfect environment to try such an experiment given its track record in international cooperation.

SCOAP3 is now collecting Expressions of Interest from partners worldwide to join the consortium. Once a critical mass is reached and a global consensus demonstrated, the consortium will be formally established and its international governance defined. SCOAP3 will then issue a call for tender to publishers, aimed at assessing the exact cost of the operation, and it will then move quickly forward with negotiating and placing contracts with publishers. SCOAP3 is rapidly gaining momentum. At the time of writing, most countries in Europe have pledged their contribution to the project. In the United States, leading libraries and library consortia have pledged a redirection of their current expenditures for HEP journal subscription to SCOAP3. In total, SCOAP3 has already received, or is about to receive, pledges for about half of its budget envelope, with another considerable fraction having the potential to be pledged in the short-term future. This consensus basis is not restricted to Europe and North America: Australia is part of the consortium and advanced negotiations are in progress in Asia and in Latin America.

## **6. The INSPIRE project**

In 2007, a user survey was carried out to assess the use of HEP information resources, and learn how to meet users' needs better.<sup>6</sup> HEP has a long history of providing electronic access to scientific information, pioneering the development of tools such as the SPIRES database and the arXiv repository. Since then, the



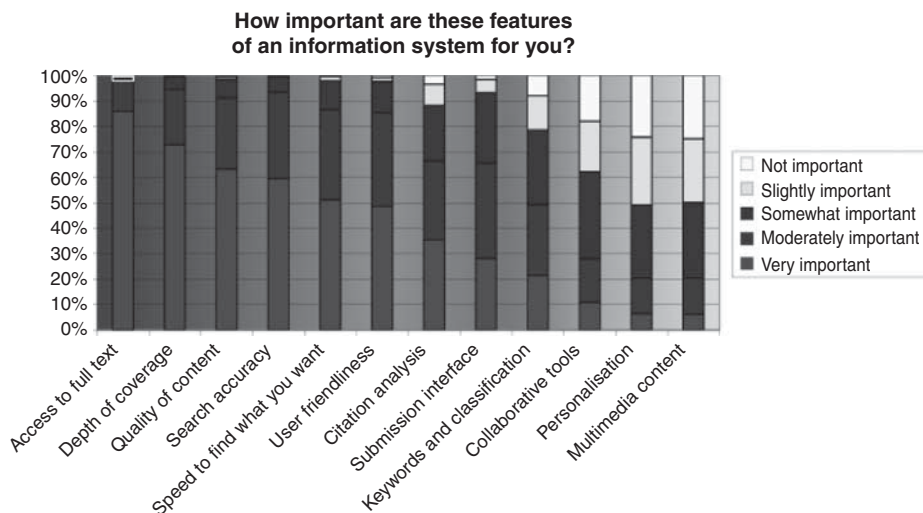
**Figure 7.** Information resources favoured by HEP scientists. Community-based systems dominate the landscape, even though among younger scholars there is an onset of Google. The usage of commercial systems (SCOPUS, INSPEC, the Web of Science and similar products) is negligible.

scientific information landscape has changed radically: electronic versions are available for most journals, commercial on-line databases add value by providing metadata, and the world wide web is easily searchable. For all these reasons, users' expectations and requirements are changing, and the HEP-developed systems may no longer be sufficient.

The survey met an overwhelming response and was completed by about 10% of the scholars active in the field. It provides an in-depth case study on the use of discipline-based information resources, as opposed to institution-based or commercial ones. The survey sheds light on strategies to adapt commercial products to discipline-specific environments, by assessing the penetration of commercial platforms in a field where community-produced resources have long been the only source of information. And information about use of institution-based information resources is particularly relevant as these are at the centre of recent worldwide moves towards self-archiving of research results.

Figure 7 shows responses to the main question: which information system do you use the most? For over 91% of respondents it was the SPIRES database, the arXiv repository or another community-based service. The use of commercial services, at 0.1%, is negligible. Survey questions were further broken down to investigate preferences in varying circumstances, for example where the authors or references were known, or in the case of theses.

Figure 8 shows the importance accorded by users to various aspects of information resources. Four features predominate: access to full-text, depth of coverage, quality of content and search accuracy.



**Figure 8.** Features of an information system most relevant for HEP scientists.

Questions to ascertain how far HEP scientists expected information resources to change over the next five years revealed that 75% expected ‘some’ to ‘a lot of’ change. In particular, they wanted to see:

- linking of all instances of a result, including grey literature, conference slides, letter articles, long reviews, etc;
- the possibility of publishing ancillary material in the HEP information resources:
  - numerical data corresponding to tables and figures;
  - correlation matrices and additional information beyond those presented in tables, to allow effective re-use of scientific results;
  - fragments of computer code accompanying complex equations in articles, to improve the research workflow and reduce the possibility of errors;
  - primary research data in the form of higher-level objects;
- smarter search tools, allowing access to related material cited or tagged by others in the HEP community.

The results of this survey have been studied and are now being acted upon. The four leading HEP laboratories (CERN, DESY, Fermilab and SLAC) are combining their resources to develop INSPIRE – a fully integrated HEP information platform for the future. Work is already underway to transfer the functionality of SPIRES to a modern platform: CERN’s open-source digital library software, Invenio.<sup>16</sup> The release of INSPIRE is expected at the end of 2008. Text- and data-mining



applications, citation analysis and other tools, and Web2.0 features will then be incorporated. Among other things, these will allow:

- detection of relations between similar documents;
- creation of datasets enabling new hybrid metrics to measure the impact of articles, authors and groups;
- extraction of numerical information from figures and tables;
- increased involvement of authors and readers in their information resource: tagging documents, modifying automatically generated classifications, community-based aggregation of related objects (articles, preprints, conferences, lectures), addition of links to other digital objects, etc;
- community discussion and review of articles.

By combining the content of the current repositories and databases into INSPIRE it will be possible to provide access to the entire body of metadata and the full-text of all open access publications. To ensure that the functionality provided meets the requirements of the HEP community, development will be carried out in synergy with partners such as arXiv, the major publishers and, of course, the users. Some of the Web2.0 features mentioned above are already available as a proto-form of alternative peer-review on sites overlaid on arXiv, but they are not yet widely accepted as usage of these sites is relatively low. INSPIRE will almost certainly become the system used by the majority of HEP users to access literature on a daily basis, and it will be very interesting to see whether the provision of such tools here will lead naturally to the widespread adoption of these new means of communication in the mainstream research workflow.

## **7. Conclusions**

After half a century of circulation of preprints and almost two decades of inception and adoption of repositories, HEP has spearheaded (open) access to scientific information. The publishing and library landscapes in HEP are now in a new period of change, built on the tradition of successful, user-driven innovations: HEP is at the crossroads of open access and peer-reviewed literature and the inception of a next-generation repository that is adapting the current technological advances to the research workflow of HEP scientists.

SCOAP3 is a unique experiment of ‘flipping’ from Toll Access to Open Access all the journals of a given discipline. Its success so far, and its eventual fate, are important to inform other initiatives in Open Access publishing for several reasons: the collaborative structure of HEP, its contained publication landscape, and its tradition of maintaining repositories, make SCOAP3 a unique laboratory to identify sustainable Open Access publishing models in the era of

widespread author self-archiving of research results. At the same time, some of the obstacles met by Open Access publishing so far are related to authors' justified concerns about financial barriers for the payment of Open Access fees and their reluctance to submit articles to new, Open Access, journals. The SCOAP<sup>3</sup> initiative benefits from a strong consensus from the side of the researchers as it addresses both these points: it does not imply any direct financial contribution from authors and aims to convert to Open Access the high quality peer-reviewed journals that have served the community for decades.

INSPIRE, a new e-infrastructure for information retrieval in HEP is being designed and will soon be deployed. It aims to answer to the evolving needs of HEP scholars, deploying novel text- and data-mining, as well as Web2.0 applications. This new e-infrastructure might provide an inspiration to many other communities that are currently exploring ways to improve the dissemination, discovery and organization of research results, primarily focusing on author self-archiving.

In conclusion, the HEP research community, in decades of partnership with its libraries and its publishers, has charted a route to navigate the intricate relationship between research dissemination and accessibility. Its next steps are an important beacon that could be followed by other communities facing similar needs for a change in scholarly communication.

## Notes and References

1. T. Berners-Lee (1999) *Weaving the Web* (San Francisco: Harper Collins).
2. J. Gillies (2008) *The World Wide Web turns 15 (again)*. <http://news.bbc.co.uk/2/hi/technology/7375703.stm> (Last visited 25 May 2008).
3. One of the most extensive sources of information on the Open Access movement is <http://www.earlham.edu/~peters/fos/overview.htm> (Last visited 25 May 2008).
4. L. Goldschmidt-Clermont (1965) Communication patterns in high-energy physics. *High Energy Physics Libraries Webzine*, issue 6, March 2002, <http://library.cern.ch/HEPLW/6/papers> (Last visited 25 May 2008).
5. L. Addis (2002) *Brief and Biased History of Preprint and Database Activities at the SLAC Library*, <http://www.slac.stanford.edu/spires/papers/history.html> (Last visited 25 May 2008); P. A. Kreitz and T. C. Brooks (2003) *Sci. Tech. Libraries* **24**, 153, arXiv:physics/0309027.
6. A. Gentil-Beccot *et al.* (2008) *Information Resources in High-Energy Physics: Surveying the Present Landscape and Charting the Future Course*, arXiv:0804.2701.
7. P. Ginsparg (1994) *Computers in Physics*, **8**, 390.
8. P. Kunz *et al.* (2006) *The Early World Wide Web at SLAC*, <http://www.slac.stanford.edu/history/earlyweb/history.shtml> (Last visited 25 May 2008).
9. S. Mele *et al.* (2006) *Journal of High Energy Physics*, **12**, S01, arXiv:cs.DL/0611130.

10. S. Bianco *et al.* (2007) Report of the SCOAP<sup>3</sup> Working Party, <http://www.scoap3.org/files/Scoap3WPReport.pdf> (Last visited 25 May 2008).
11. J. Krause *et al.* (2007) *Quantitative Study of the Geographical Distribution of the Authorship of High-Energy Physics Journals*, <http://scoap3.org/files/cer-002691702.pdf> (Last visited 25 May 2008).
12. M. Jacob (1995) *In the Wings of Physics* (Singapore: World Scientific).
13. M. Jacob (1990) *Publishing and Editing Physics with North Holland*, <http://doc.cern.ch/archive/electronic/other/preprints//CM-P/CM-P00052106.pdf> (Last visited 25 May 2008).
14. <http://scoap3.org> (Last visited 25 May 2008).
15. R. Voss *et al.* (2006) *Report of the Task Force on Open Access Publishing in Particle Physics*, <http://www.scoap3.org/files/cer-002632247.pdf>
16. <http://cdsware.cern.ch/invenio/index.html> (Last visited 25 May 2008).

### About the Author

**Robert Aymar's** career has been focused on fundamental research in plasma physics and its application in controlled thermonuclear fusion research. He was director of natural science at CEA (France) from 1990 to 1994 and Director-General of the ITER project from 1994 to 2003. Robert Aymar was then elected as Director-General of CERN as of 1 January 2004 for a period of five years. During his tenure, the laboratory completed the construction of the LHC, which circulated the first beams in the summer of 2008. Robert Aymar is an Open Access advocate and the driving force behind the CERN actions on Open Access over the recent years.