

Article

Open Access and the Changing Landscape of Research Impact Indicators: New Roles for Repositories

Isabel Bernal

Spanish National Research Council, Unit of Information Resources for Research, DIGITAL.CSIC Technical Office, C/Joaquin Costa 22, Madrid 28002, Spain; E-Mail: isabel.bernal@bib.csic.es; Tel.: +34-915-681-667.

Received: 6 May 2013; in revised form: 2 July 2013 / Accepted: 10 July 2013 /

Published: 19 July 2013

Abstract: The debate about the need to revise metrics that evaluate research excellence has been ongoing for years, and a number of studies have identified important issues that have yet to be addressed. Internet and other technological developments have enabled the collection of richer data and new approaches to research assessment exercises. Open access strongly advocates for maximizing research impact by enhancing seamless accessibility. In addition, new tools and strategies have been used by open access journals and repositories to showcase how science can benefit from free online dissemination. Latest players in the debate include initiatives based on alt-metrics, which enrich the landscape with promising indicators. To start with, the article gives a brief overview of the debate and the role of open access in advancing a new frame to assess science. Next, the work focuses on the strategy that the Spanish National Research Council's repository DIGITAL.CSIC is implementing to collect a rich set of statistics and other metrics that are useful for repository administrators, researchers and the institution alike. A preliminary analysis of data hints at correlations between free dissemination of research through DIGITAL.CSIC and enhanced impact, reusability and sharing of CSIC science on the web.

Keywords: open access; research assessment systems; new metrics and indicators; expanded roles for repositories; repositories and research visibility optimization

1. Scholarly Communication Online and the Crisis of Traditional Measures

Coined by Eugene Garfield in 1955, the Journal Impact Factor (JIF) has been a predominant measure used to evaluate scientific literature and identify researchers' prestige, promotion and tenure world-over. Originally created to assist libraries in their journal purchases, this indicator continues to exert a large influence on scientific strategic plans, funding decisions and scholarly communication.

However, JIF misuse has often been reported and the need to establish a better correlation between this indicator and other research performance metrics is all but a novelty [1], all the more since Internet and other technological developments unleashed their potential. Critiques of the Thomson Scientific monopoly on impact factor game abound. Similarly, its secretive and unscientific processes to rank scholarly journals in the Index have been put into question [2]. New indicators like the h-index (2005) have emerged to focus on the productivity and impact of authors' works more accurately, whereas journal's rankings, like Eigenfactor (2006), weigh the significance of citations depending on the relevance of journals to the scientific community. What is more, Elsevier-driven Scimago (2007) and SNIP (2010) indicators also take into account online reputation for their scores. Nonetheless, scholarly citation counts and journal prestige amongst peers continues to underpin these models.

In practice, the system based on citations and JIF suffers from limitations. In particular, it has been observed that articles published in prominent journals cite works published in other prestigious journals only. It is the so-called "Matthew effect" [3], meaning that not all relevant sources of an article are referred or, at the same time, other valuable works are ignored. As a recent study reports, approximately 30% of influential sources are typically cited [4]. In addition, the concentration of academic citations is quite strong, and studies estimate that only the top 20% of articles receive approximately 80% of all citations in journals [5]. Does this mean that the rest of scholarly papers lack interest and impact? Accordingly, a PLoS analyzed its published articles and concluded that citations reflect just 1% of usage [6]. As another well-known critique argues, self-citations and certain unethical editorial policies tend to inflate JIF artificially [7].

The advent of electronic information technology, particularly that of the Internet, gave way to new journals, new data and new approaches for assessing research. Indeed, since 1990 the relation between JIF and paper citations has been gradually decreasing across disciplines. In other words, the proportion of highly cited papers published in highly cited journals has diminished. At the same time, the rate of highly cited papers published in not highly cited journals has increased [8].

Open access has made the debate of revaluating scientific impact indicators all the more evident, as wider accessibility and easy exchange of scholarly information has paved the way for further and deeper considerations—at the scientific level and beyond. The need for new evaluation systems was addressed at the 2012 BOAI Conference [9] and it soon became a hot topic amongst funders, publishers, open access advocates, repository administrators, bibliometricians as well as in the research and innovation sector, and software community. The Future of Academic Impact Conference organized by the London School of Economics Public Policy Group in December 2012 is just one example of the many initiatives that have since been launched to examine both the benefits of broader criteria to measure research impact and the role open access can play in the years to come.

2. New Tools and Venues in an Open Access Scenario

The beginning of the digital age witnessed a gradual breakdown in the correlation between highly cited journals and highly cited papers. Nevertheless, except for specific research areas, such as physics, the leap from print to online publishing did not cause immediate curiosity towards novel approaches to sharing and assessing science. Despite the exponential growth of data sets including usage statistics and webometrics, JIF has succeeded in keeping its prevalence second to none across most research evaluation systems in the world.

The uptake of open access and web 2.0 tools in the last decade has contributed to producing deep transformations in the foundations of the model. Today, researchers are actively involved in platforms such as Mendeley, Research Gate, and academia.edu so as to make their works more easily located and discussed and to establish quicker connections with peers. They also self-archive their postprint and preprint articles in institutional and subject repositories, publish in open access journals and make their scientific raw data public more and more often.

Lifting economic and technical barriers to searching, retrieving and accessing full-text scientific literature promotes greater readership. As a corollary, open access may also increase scientific impact. In 2001 it was first reported that freely available online science proceedings accumulated more than three times the average number of citations received by print articles [10]. This "citation advantage" has been found across many disciplines and a 2010 study summarized the main conclusions and methodologies by principal studies [11]. OpCit project [12] keeps an exhaustive track of new related bibliographies.

On average, freely available articles tend to be more highly cited and this trend has sometimes been related to the so-called self-selection postulate, meaning that authors self-archive their works that have been published in highly cited journals [13]. However, recent studies [14] stress that open access citation advantage takes place independently from other variables like article age, journal impact factor, number of co-authors, references or pages, research field, article type or country of publishing. And, above all, such an advantage peaks for the most highly cited articles.

In addition, open access is testing alternative models for scientific journals rankings. For instance, by considering relative openness of publishers' self-archiving policies, a MIT team is working on the so-called "accessibility quotient". This measure is calculated combining Ted Bergstrom's Relative Price Index that assesses affordability and quality, and data from SHERPA/RoMEO that assesses the right to share peer-reviewed version of an article [15]. Another tool, Google Scholar Metrics, attempts to rank journals by gathering citations across a wide set of sources, including repositories and non-traditional platforms [16].

Both gold and green open access routes have developed new approaches and tools to gauge the attention that open access research outputs arouse. A good paper may take anywhere from 2 to 5 years to start gathering scholarly citations. In the meantime, however, other instruments can quantify its impact and help make estimations on the potential growth of citations in the future. Hence, measuring short-term impact through counts of HTML views, downloads, social shares and bookmarking in scientific reference managers may be indicative of potential impact in the medium and longer term.

The uptake of repositories has facilitated the implementation of usage statistics modules, like those developed by Eprints, DSpace and Bepress softwares. Some are outstanding in their level of detail, as

they provide information on the demographic distribution related to usage, top search keywords, aggregated data on end-user behavior, ratio of traffic generated within the institution and elsewhere, etcetera. Due to their popularity and usability, most repositories supply data concerning page views and free full text downloads, either publicly or to repository administrators (and even authors upon request). Researchers make use of such data to get a better insight into the reach of their works that are in open access, whereas repository administrators draw on the data to envisage usage-driven content development strategies, promote collections similar to most downloaded resources, and provide evidence of enhanced impact of OA contents to institutional policymakers. However, a pending issue in the realm of usage statistics refers to the lack of an international standard that make possible the collection of usage counts across publishers' platforms, middlewares, open access repositories and other sites in a consensual way. IRUS-UK [17] is still a developing project, but it appears as a positive effort to reach such consensus amongst repositories.

To a lesser extent, repositories configure tools to track citation counts in an open environment. In this regard, subject repositories have taken the lead and, for instance, CiteSeer, RePEC and arXiv maintain metrics for works housed in their platforms and cited by others available in open access, too. Institutional repositories have followed suit, and most of the times they have added APIs that display scholarly citations as gathered by WoS, Scopus and Google Scholar, if any. The Open Citations Project explored how to publish bibliographic and citation information about biomedical literature in RDF in order to facilitate citation links. Its extension is working on a prototype for wider usage [18].

By focusing on single item metrics, open access has placed the emphasis on the importance of impact of research outputs *per se*, not for the prestige of journals where they are published. A model that has being largely adopted by open access initiatives and subscription-based publishers is the one developed by PLoS. Since citation counts assess scholarly impact in the medium and long terms, works' immediate effect across multiple venues and fora used to be rather neglected. To fill the gap, PLoS article level metrics include, in addition to traditional benchmarks like scholarly citations, other data such as non scholarly citations (science blogging like ScienceSeeker or Wikipedia), usage statistics and altmetrics (social shares and academic bookmarks).

Plum Analytics, a forerunner in the race to consolidate the newest indicators and methodologies, works along similar lines to PLoS, It classifies the impact of research artifacts (from articles to videos) within 5 different categories [19]: usage, captures, mentions, social media and citations. Due to their infancy, some skepticism exists around these tools, but enthusiasm from players that call for new impact frameworks somehow redresses the balance. Indeed, a growing number of publishers (Cambridge University Press, BioMed, Nature, Taylor&Francis), bibliographic databases (Scopus), and institutional repositories have started to display such data, which proves that scholarly communication community sees a potential for development.

Preliminary studies on the impact of open access research and the measurement of its citation counts and immediate attention on the web have identified the following trends:

(1) Open access citation advantage ranges between 25% and 250% depending on the discipline. This is a quality advantage driven by the intrinsic value of works and not by the mere fact of being freely available [20];

- (2) Open access, either through voluntary self-archiving or mandatory deposit, has a positive effect on the scientific impact of new articles [14];
- (3) There may be a correlation between research dissemination through social web and downloads from repositories [21];
- (4) HTML/PDF paper downloads, coupled with social media shares, are a better indicator of public interest than absolute usage statistics [22];
- (5) As supplementary indicators, scholarly altmetrics can show value added for open access content, and this fact could act as a new incentive for authors to increase their repository self-archiving rates [23];
- (6) For open access content, there often seems to be a correlation between page views and full text downloads with scholarly citations and social impact in social venues [24,25];
- (7) Academic citations are often closely related to social reference management bookmarks [26];
- (8) Bookmarking services (Mendeley, Research Gate, CiteuLike) are increasingly used by scholars and are a better measure of scholarly interest than social media (Twitter, Facebook) where public interest may be stronger [22];
- (9) Author blog posts are especially meaningful for the scientific community insofar as they provide commentary and critique of newly published articles, thus enriching original works.

3. Analyzing Impact Indicators at DIGITAL.CSIC: An Institutional Repository Case

A corollary of the Berlin Declaration signed by CSIC Presidency in 2006, DIGITAL.CSIC [27] was launched in January 2008 in order to store, describe, enable open access and preserve outputs by scientific community of the Spanish National Research Council. The repository is as multidisciplinary as the Council is and its collections embrace scientific outputs, educational resources and non-traditional research artifacts. However, its main collections are those made of journal articles, conference papers and book chapters. With more than 73,000 items, it is the largest open access repository of scientific resources in Spain.

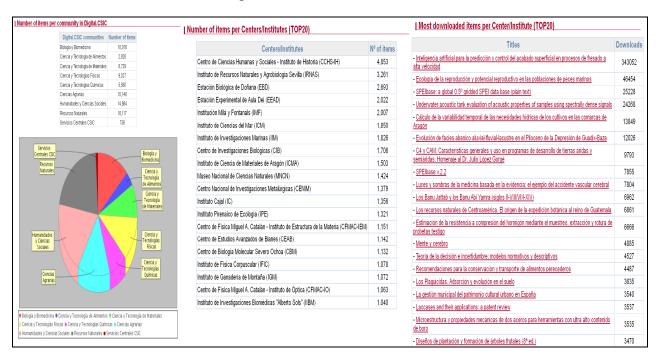
Since its inception, DIGITAL.CSIC has aspired to play a relevant role in maximizing the visibility and impact of CSIC science online. As a result, metrics have been steadily incorporated into the platform in the attempt to showcase this impact and gather other data for institutional purposes. What follows is a brief account of metrics available in DIGITAL.CSIC and how repository administrators, the CSIC scientific community and the institution itself can benefit from them.

3.1. Statistics

Since 2010, DIGITAL.CSIC has added functionalities to better analyze and report online impact generated by its contents. To achieve this goal, both traditional and new tools have been considered. Showcasing the diversity of CSIC research areas and lines of excellence and tracking usage on the web in a comprehensive fashion necessarily implies a focus on scholarly attention and public interest.

Usage statistics have been a most valued service by CSIC researchers, and an agenda of improvements has been designed in this respect [28]. Although DIGITAL.CSIC runs on DSpace, a module developed by Tasmania University was chosen to retrieve data on views and full text downloads at a time when DSpace statistics were less rich as they are today. On top of this module, in Spring 2010 the repository's Technical Office built an in-house application to generate more granular statistics [29]. DIGITAL.CSIC statistics collect rich data at the CSIC center/institute level (Figure 1). CSIC is a complex research performing organization comprising over 130 institutes and it was deemed essential to provide aggregated numbers as well as more specific information as regards scientific areas and institutes at equal footing. Further, the module does not limit itself to usage statistics, but also gives an overview of item distribution across scientific areas, typologies and a set of TOP20 rankings (institutes and authors with highest number of deposited items; authors -or mediating libraries- that self-archive the most per month/year -data with restricted access-; and most viewed and downloaded items per institute, month and year).

Figure 1. From left to right: Content distribution across all CSIC scientific areas; ranking of CSIC institutes with highest volume of items (totals), and 20 most downloaded items in 2013. Source: DIGITAL.CSIC general statistics.



The data has been extremely helpful for the repository's administrators to study growth patterns through years, learn how involved CSIC institutes and libraries are in the project, and start measuring impact of contents. For instance, we have observed libraries' consolidation in their intermediary role by looking at monthly and yearly ratios of deposits made on behalf of researchers in their institutes (Figures 2 and 3). We have also been able to identify open access advocates across all research areas and institutes and narrow down themes and research lines that attract the biggest interest on the web. On very limited occasions, extremely high volume of downloads in a short time has hinted at upload of invalid objects or at irregular traffic from specific sites.

Figure 2. The graph shows how self-archiving and mediated archiving services have evolved in the last three years. The Mediated archiving is a service offered by DIGITAL.CSIC Technical Office and CSIC libraries network to researchers and overall more than 95% of manual uploads is carried out by CSIC library community.

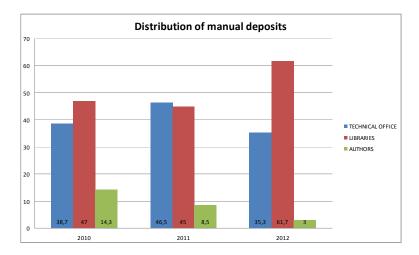
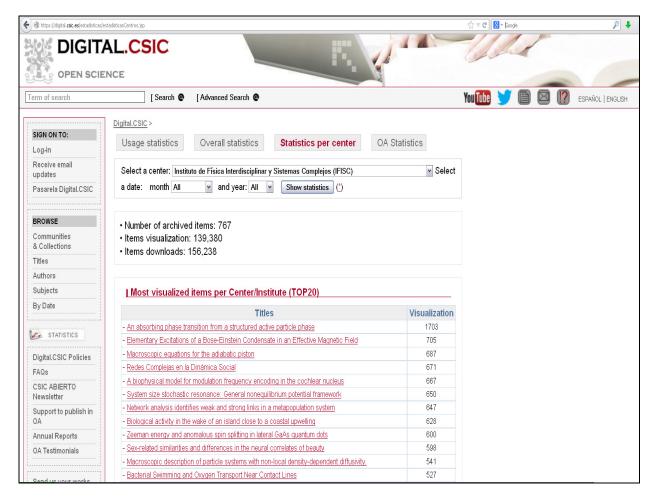


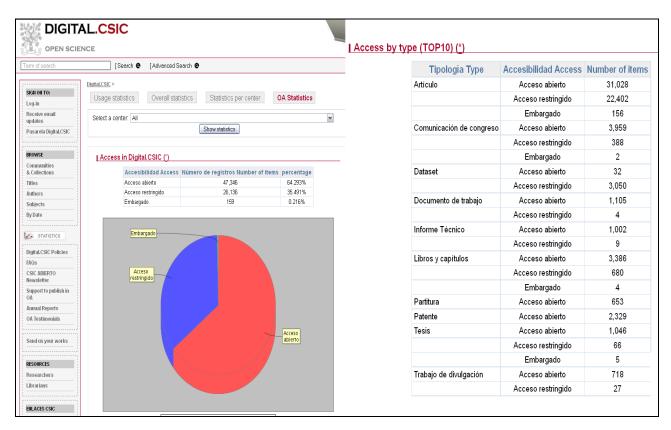
Figure 3. Statistics provided at CSIC institute level. Details include total number of items; monthly and yearly deposits; monthly, yearly and total downloads and views; CSIC staff (researchers, librarians) who upload the most per month and year; 20 most downloaded and viewed works per institute. Source: DIGITAL.CSIC Institute level statistics.



Equally, CSIC institutes and authors can catch sight of their deposited works' immediate and long-term interest. The module has acted as a powerful catalyst for institutes to revise their dissemination strategies and give room for DIGITAL.CSIC –this is the case of the research institute *Estación Experimental de Aula Dei EEAD-CSIC* [30]. Likewise, authors can complement traditional indicators (citations counts) with HTML views and PDF downloads of their works, classified by end-user countries. As a whole, DIGITAL.CSIC items have been downloaded 15 million times (a rough 25% accounts for robot action) since 2008.

In Spring 2013, the DIGITAL.CSIC statistics module got enriched with open access data, thus allowing a step further in content analysis. The repository now displays aggregated percentages of openness and the application enables a closer look into accessibility degrees by object typologies, year of deposit, scientific areas and CSIC centers and institutes (Figure 4). Thanks to additional data, it is possible to monitor items growth filtering by full text files, restricted and embargoed access options and to identify typologies (articles, conference papers, patents, musical compositions, theses, teaching material, *etc.*) that contain the highest volume of OA resources.

Figure 4. Openness degree at DIGITAL.CSIC and by item typologies. Source: DIGITAL.CSIC OA statistics.



Deeper insight into these results shows a relatively high correlation between institutes that boast the highest number of items in the repository, institutes that enjoy the most highly downloaded items and their free full text accessibility through DIGITAL.CSIC. In particular, three of the five institutes with the highest total number of items rank amongst the five institutes that accumulate the highest number of full text downloads (Tables 1 and Table 2).

Table 1. Openness degree in collections of CSIC institutes with the highest number of downloads at DIGITAL.CSIC (totals).

CSIC institute	Percentage of OA items at DIGITAL.CSIC	Percentage of restricted access items at DIGITAL.CSIC	Full text downloads from DIGITAL.CSIC
Estación Experimental de Aula Dei (EEAD)	95% (1846)	5% (87)	744,277
Centro de Ciencias Humanas y Sociales—Instituto de Historia (CCHS-IH)	36% (3067)	64% (1,715)	648,021
Instituto de Ciencias del Patrimonio (INCIPIT)	99.5% (626)	0.4% (3)	388,877
Institución Milá y Fontanals (IMF)	95% (1877)	4.7% (93)	388,325
Instituto de Ciencias de la Construcción Eduardo Torroja (IETCC)	58% (354)	41.5% (254)	410,977

Table 2. Correlation between openness and full text downloads for CSIC institutes with the highest number of items at DIGITAL.CSIC.

CSIC institute	Number of items in DIGITAL.CSIC	Downloads from DIGITAL.CSIC	Open access items in DIGITAL.CSIC	
Centro de Ciencias Humanas				
y Sociales-Instituto de	4784	648,021	36% (3,067)	
Historia (CCHS-IH)				
Instituto de Recursos				
Naturales y Agrobiología	3082	196,327	87% (2,084)	
Sevilla (IRNAS)				
Estación Biológica de	2633	178,604	84.4% (2,223)	
Doñana (EBD)	2033	1 / 0,004	04.470 (2,223)	
Institución Milá y	1971	388,325	95% (1877)	
Fontanals (IMF)	19/1	300,323	93/0 (10/7)	
Estación Experimental de	1937	744,277	95% (1,846)	
Aula Dei (EEAD)	1/3/	777,277	7570 (1,040)	

In addition, for internal purposes, DIGITAL.CSIC administrators make use of Google Analytics to complete the picture. Most attention is paid to search terms, ratio of external and institutional users, search engines users preferences, and total document views. Monthly reports summarize the most remarkable data. Amongst relevant findings it is worth mentioning that an important percentage of traffic originates from within CSIC sites, while most used external sites include Google, scholarly aggregators Google Scholar, Scirus and Worldwidescience, Spanish academic harvester Recolecta, social media platforms like Wikipedia, Facebook, Twitter and so on (Figure 5).

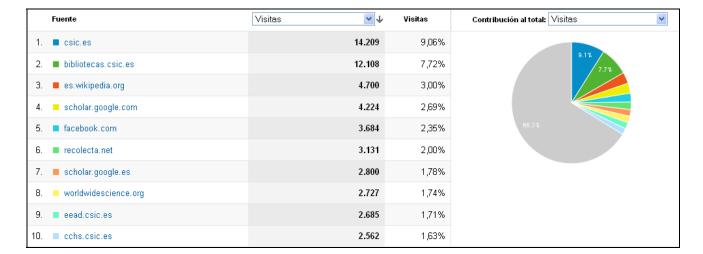


Figure 5. Main search sources (July 2011–April 2013) Source: Google Analytics.

3.2. Impact Data at Item Level

Building bridges with society, reaching out to different audiences and measuring public interest in its research outputs have gained rising importance at CSIC Action Plans. Institutional dissemination policies and inclusion of a few types of dissemination efforts into annual research assessment exercises have followed as a direct consequence.

Since 2011 DIGITAL.CSIC has been adding tools and applications to capture and display scholarly citations and facilitate wider sharing in social and researcher-oriented platforms. From the very outset, relying on a single data source was avoided, given the multidisciplinary nature of the repository and its broad content policy. Hence, commercial databases and freely available tools are used to gather impact data for as much content as possible. Thus far, scholarly citations counts are retrieved from Scopus, WoK, PMC and Google Scholar.

On another front, altmetrics have been added recently for their potential to supplement traditional citation counts and to enrich usage statistics data with details on how and where works are shared and commented on the web. These alternative indicators assist researchers in getting a better insight into readership of their content available through DIGITAL.CSIC, monitoring impact of their non-traditional research artifacts such as grey literature and working papers, and helping build a richer CV. For their part, repository administrators can persuade potential depositors more easily by showcasing how open access works can track additional impact data, whereas institutional policymakers and funders can consider enhanced criteria for evaluation activities.

The analysis of how all these metrics perform together on a selection of items available at DIGITAL.CSIC often finds correlation between free full text accessibility on the web and immediate interest and enhanced impact amongst scientific and lay communities. Without meaning to assert direct correlations, these findings hint at certain trends, which are worth of further study:

Digital.CSIC as gateway to fuller CVs and the role of mediated archiving service: CSIC does have an open access institutional policy in place but does not mandate submission of scientific outputs to the repository. Several studies have observed that, in general, voluntary self-archiving accounts for less than 20% of deposits to open access institutional repositories, while institutions with open access mandates see the figure rocket up to 90% [31]. Along with the self-selection postulate, we would

therefore expect self-archiving to be close to ground and, if any, involve works mostly published in highly cited journals.

In fact, CSIC authors' self-archiving rates are very low as they range from 3% to 4% per year, but we also witness a monthly increase in deposit rates through the Mediated Archiving Service provided by CSIC libraries to their researchers. The growing intermediation in upload activity (Figure 2) has freed authors from dealing with copyright verification issues and information management activities, considered a work typical of libraries. With libraries acting as knowledgeable mediators, institutes and researchers are motivated by their own engagement with open access for they can make their outputs freely available in a more efficient way (Table 3).

Table 3. Selected OA friendly authors across all CSIC scientific areas and presence of their outputs in Scopus and DIGITAL.CSIC.

CSIC	CSIC scientific area and	Works indexed	Works at
author	institute affiliation	by Scopus	DIGITAL.CSIC
Jordi Figuerola	Natural Resources, EBD	126 items	129 items
José Luis Chiara	Chemistry, CENQUIOR	56 items	66 items
Angel Mantecón	Agricultural Sciences, IGM	77 items	472 items
Antonio Pich	Physics, IFIC	147 items	188 items
Antonio Almagro	Humanities, EEA	6 items	229 items
Francisca Puertas	Materials, IETCC	111 items	112 items

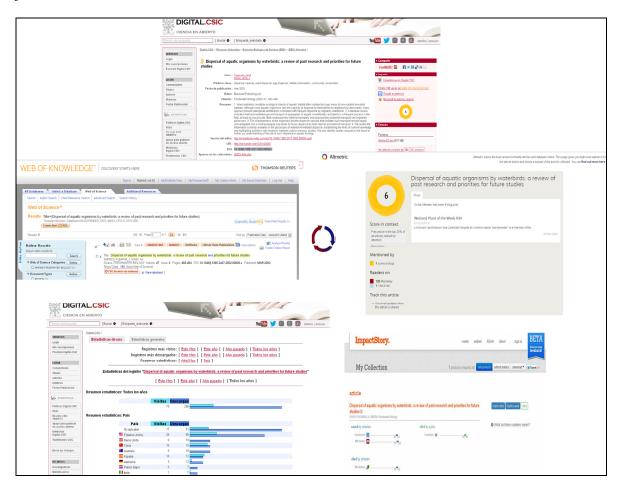
The authors above most reflect the CSIC community that benefits from the repository to enhance visibility of all their research outputs, regardless of their format, year of publication, nature of work and number of scholarly citations. In all instances, DIGITAL.CSIC serves as a platform to give a more complete overview of their careers, and data referred to the Humanities researcher is a telling example of how partially commercial bibliographic databases cover research in the area, thus distorting scientific productivity and authors performance greatly.

The value of usage statistics: Repository usage statistics report a fraction of impact which scholarly citation counts do not necessarily reflect. Reasons may hint at a relatively extended tendency amongst researchers to avoid citing "informal" works. Usage statistics may also accumulate interest generated by audiences like science amateurs, students, scientific dissemination communities, professional sectors, practitioners, R&D companies and so on. However, such data reveal attention and shareability of works that are in open access. On other occasions, correlation between highly cited papers and high usage statistics may exist, and this is mostly the case for high quality works.

As an exercise, we will consider the most highly cited paper by Jordi Figuerola, a CSIC Natural Resources researcher who uploads all his outputs at the institutional repository, to observe whether there may be a correlation between traditional metrics, the repository's usage statistics and traffic retrieved by altmetrics. Published in "Freshwater Biology" in 2002, his work *Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies* [32] has gathered 196 academic citations in Scopus, while DIGITAL.CSIC usage statistics have totaled 258 downloads (mainly from the US, Australia and the UK) of its freely available copy since it was uploaded in December 2011. In parallel, altmetrics have identified sites where the work has been mentioned, discussed and shared. As a result, the Altmetric score displays 106 bookmarks at Mendeley and one

mention in a researcher blog. ImpactStory retrieves additional data like its inclusion as a reference in a Wikipedia entry about Bird migration (Figure 6).

Figure 6. Impact data (immediate interest and medium term impact) collected for Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies.



Figuerola's next two most cited papers are also accessible from DIGITAL.CSIC. Both are characterized by high PDF downloads since they were uploaded in December 2011. Thus, *Greatly Enhanced Arsenic Shoot Assimilation in Rice Leads to Elevated Grain Levels Compared to Wheat and Barley* [33] (120 citing works in Scopus) accumulates 151 downloads at the repository, while *Implications of water ecology for the dispersal of aquatic organisms* [34] (110 citations in Scopus) hits 234 full text downloads. However, none shows any Altmetric score thus far.

The repository's usage statistics can give a sense of the interest generated by works other than journal articles. This is often the case for theses. For instance, food-related *Inteligencia artificial para la predicción y control del acabado superficial en procesos de fresado a alta velocidad* [35] is a highly popular item at DIGITAL.CSIC as it ranks first in the TOP20 classification of most downloaded items, with more than 300,000 downloads. Working papers, technical reports and conference papers may be the object of similar popularity (e.g., *Underwater acoustic tank evaluation of acoustic properties of samples using spectrally dense signals* [36], with almost 28,000 downloads, ranks ninth of TOP20 most downloaded works in DIGITAL.CSIC).

More players gathering scholarly citations and references: Despite all efforts to catch up with changes in scholarly communication and broaden content coverage by accepting more journal titles and other artifacts (i.e., books and datasets), commercial databases are challenged by competitors that provide similar services yet in a free and more open setting. Further, the latter tend to index a more diverse set of data sources including repositories, reference managers, alternative publishing platforms, etc.

In this sense, remarkable disparities in measuring books' impact (citation counts) is telling enough, although similar cases may be found with other research artifacts. Most books and book chapters available at DIGITAL.CSIC do not gather any scholarly counts in commercial databases, yet a closer look at Google Scholar or Microsoft Academic Search indexation may shed some more light. This is the case for the following CSIC-authored books: Google Scholar collects 2 citations *Automation for the Maritime Industries* [37] and 25 citations for *Cuatro viajes en la literatura del antiguo Egipto* [38], the most highly downloaded work under *Instituto de Lenguas del Mediterráneo y Oriente Próximo CCHS-ILC* collections.

As mentioned above, another useful service by competitors to traditional impact databases refers to a broader source of coverage. By way of illustration, the article *Polycarboxylate superplasticiser admixtures: effect on hydration, microstructure and rheological behaviour in cement pastes* [39] gathers 50 scholarly citations in Google Scholar. Plus, the ability to go through all data sources where citing works have been retrieved, ranging from reference managers to repositories and other open access platforms, adds a sense of a more transparent process.

On another front, analysis of citation performance as regards CSIC journal articles [40] brings about interesting outcomes. CSIC academic journals have undergone an open access conversion since 2007. Simultaneously many titles have been included in traditional bibliographic databases. However, a fuller picture in this regard may be provided by new indexation players. In fact, Google Scholar displays 135 citations for article *Fauna of the Mediterranean Hydrozoa* [41] (Figure 7). Further, a correlation between DIGITAL.CSIC usage statistics (with 8,212 full text downloads the article is the most downloaded item falling under *Instituto de Ciencias Marinas ICM* collections in the repository) and impact through traditional citation counts is observed.

Immediate interest versus long term impact? Altmetrics are best used to gauge immediate interest and attention towards recently published work and to identify channels and geographical locations where mentions and reference occur. Thus, the recently published article *Mitochondrial genome diversity and population structure of the giant squid Architeuthis: Genetics sheds new light on one of the most enigmatic marine species*, which is available in bibliography-only format in the repository [42], has already generated remarkable attention across social and researcher-oriented platforms including science blogs, Facebook and Twitter (Figure 8). However, it will be advisable to wait at least a couple of years before confirming whether such interest in the short-run will have any influence on citation counts.

Figure 7. DIGITAL.CSIC as a gateway to retrieve multi-level impact data. *Fauna of the Mediterranean Hydrozoa*.

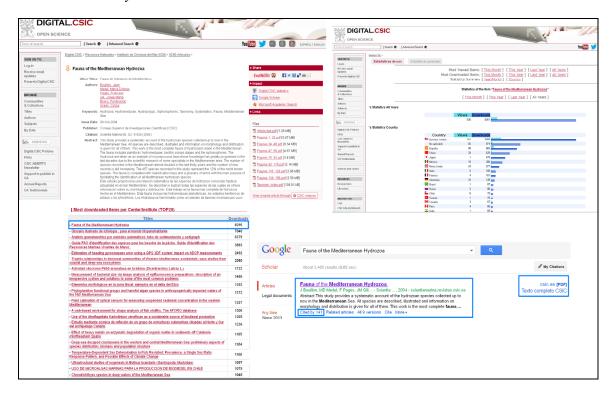


Figure 8. Altmetric API at work in DIGITAL.CSIC. Article *Mithocondrial genome diversity and population structure of the giant squid Architeuthis: Genetics sheds new light on one of the most enigmatic marine species* as an example with metrics on immediate attention.



Although not necessarily relevant in all circumstances, we would be tempted to assert that all these data collected and displayed at item level help us give a richer picture of impact according to audiences (general public, peer community), dimensions (immediate interest, self-promotion, academic impact),

and timepoints (pre- and post-publication, days, months and years). This approach may be especially useful for disciplines (Humanities, Social Sciences, Engineering) that have been poorly portrayed in traditional assessment tools, although benefits would spread across broad scientific community. Last but not least, these new data sources cover non-traditional research outputs in their analysis, such as conferences, scientific videos, teaching and educational materials. Increasingly used by many research communities, they have been largely neglected in evaluation exercises thus far, but the availability of impact data may expedite their incorporation.

3.3. Other Tools to Track Web Performance

DIGITAL.CSIC Technical Office usually resorts to SEO tools to retrieve additional data about web traffic generated by content in the repository. SEO strategies that improve web profiles in search engines and promote web prominence, link building, blogging *etc.*, have become all the rage amongst open source and commercial players lately.

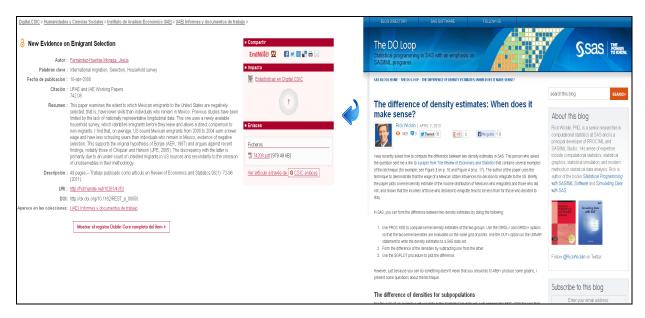
SEO tools can monitor range and nature of sites and domains that refer to DIGITAL.CSIC contents. Besides, they assist in understanding contexts where CSIC research gets traced back and reused the most, and what research lines in the repository are most frequently backlinked. Evidence of public engagement, media coverage and discussions within scholarly circles can be obtained through these tools and enhances DIGITAL.CSIC.

More and more scholars maintain blogs to share, comment on and disseminate works of interest, be they their own or authored by peers. Scientific blogs have rapidly spread, and for the repository's managers they are interesting sources to measure the reusability of open access contents. In addition, they allow authors to keep aware of discussion around their works in scientific non-traditional fora, which may end up having a positive impact on academic citations. The examples below evidence these functionalities (Figures 9 and Figure 10).

Figure 9. Article *A short account of Leonardo Torres' endless spindle* [43] featuring engineer Torres Quevedo is backlinked by Google blog entry Celebrating Leonardo Torres Quevedo as the inventor of the first computer game in the world.



Figure 10. Working paper *New Evidence on Emigrant Selection* [44] is highlighted as a useful resource at the blog entry *The difference of density estimates: When does it make sense?* by a computational statistics researcher.



To conclude, enhanced research visibility is beneficial for the institution itself as it helps raise its profile on the web. Ranking Webometrics is a CSIC project [45] which classifies research centers, universities, repositories and hospital centers in the world by looking at web-related criteria, including external backlinks, number of pages extracted from Google, Google Scholar and other search engines and rich files. Based on the theoretical approach by Almind and Ingwersen who first proposed a Web Impact Factor (WIF), Ranking webometric methodology establishes ratios by combining these criteria. The number of documents (measured from rich files in a web domain) and number of publications indexed by Google Scholar outstand in the weight. In the January 2013 edition, CSIC occupies the 7th position in world ranking for research institutions while DIGITAL.CSIC ranks on the 22nd position in world repositories classification.

4. Open Access and Evaluation Frameworks

In times of general cuts to research budgets, analyzing social and scientific impacts and outcomes of research projects has become even more pressing for funding agencies. Rapid transformations are taking place in how research can be measured in an open access environment, and institutional repositories are playing a bigger role in maximizing the impact of research outputs by universities and research performing institutions. However, few evaluation frameworks have been seriously revised with concrete measures to accommodate such changes so far.

Nevertheless, pioneering initiatives hint at what the near future may look like. Active engagement to reform tenure and promotion systems has been mostly observed in the Anglo-Saxon world. First of all, impact definition is being revisited and a many-layer approach promoted. According to the UK Research Evaluation Framework (REF), impact means "any effect on, change or benefit to the economy, society culture, public policy or services, health, the environment or quality of life, beyond academia" [46], thus going beyond the traditional interpretation of impact as scholarly excellence.

Concrete policies are being tested. To name but a few, Wellcome Trust requires funded researchers to make their outputs available in open access and its revised guidelines put the stress on the intrinsic merit of the work and not the journal where it is published as main criteria for assessment [47]. Recent statements by Higher Education Funding Council for England (HEFCE) and Research Councils UK (RCUK) seem to second Wellcome Trust policy with JIF no longer considered as a criterion in evaluation exercises. Further, a proposed HEFCE/REF mandate in the UK argues that "To be eligible for REF, all articles need to be deposited in the author's institutional repository immediately upon acceptance for publication", which opens the door for institutional repositories to play an influential role in assessment activities [48]. The so-called Liège model rests on similar lines since researchers are asked to upload their outputs into ORBI institutional repository as a prior condition for annual assessment [49]. In 2012, Indiana University revised guidelines for tenure and promotion and, as a result, reviewers are requested to keep in mind that the best new/creative research may not necessarily appear in traditionally prestigious journals or books, and researchers must provide evidence of value of their publication outlets [50].

In Spain, potential ground for building a new system exists although implementation has been rather slow. First of all, impact is still defined along scholarly parameters only. Tenure and promotion systems by ANECA (national agency that evaluates university professorship) and ANEP (national agency that assesses excellence of research outputs) are dominated by traditional quality benchmarks such as JIF Indexes and citation counts. Related criteria encompass indexation by renowned commercial databases (Scopus and WoK), publishers' standards and quality indicators, reviews, internationalization of Editorial Board committees and so on. In the Social Sciences and Humanities domain, local and alternative indexes like IN-RECS, Latindex, DICES, ERIH still prevail as tools to rank Spanish-speaking journals and those in European languages. Institutional assessment exercises greatly mirror these national frames.

Spain's Law of Science (2011) cleared the way to the possibility of using public versions of research outputs (postprint articles) for evaluation purposes [51], but this implementation is yet to happen. By the same token, regional open access mandates like those issued by Community of Madrid (2008–2009) [52] and Principality of Asturias (2009) [53] have been monitored irregularly, without aiming to stir the foundations of the system. At the Spanish National Research Council, DIGITAL.CSIC is linked to institutional CRIS ConCiencia since 2011 so that the repository can benefit from automated data ingestion and import full text files uploaded by the research community [54]. Given this integration, in the last year the repository has enjoyed a rapid content growth and a more productive research information management system has been put in place, allowing researchers to give their input to the institution once for multiple purposes. In any case, open access is not a criterion in the institutional assessment system, which revolves around indicators like JIF, Scimago rankings and citation counts. Promising enough, some additional parameters including dissemination, transfer of knowledge and social impact of science already show at secondary level [55].

5. Conclusions

The monopoly of traditional research benchmarks has been challenged by new players and alternative metrics gaining from the rich variety of data that can be collected and measured in a digital environment. Open access is also contributing to changing the scenario and fresh initiatives are being explored by journals and repositories. Driving forces for change rest on an enhanced definition of impact, which takes into account returns on investment and social benefits, with attention placed on the very value of works *per se*. This new approach makes essential the ability to access works under evaluation easily.

Despite great dynamism in the area, a few shortcomings must be overcome before new metrics consolidate as completely reliable indicators for research excellence and impact. Headways are already reported in some respects. Requiring availability of research outcome in open access is gaining momentum through funding agencies and institutional mandates and it is hoped that the European Commission Open Access policy at the upcoming Horizon2020 will set the example for other funders and stakeholders.

Increasing the role of institutional repositories within assessment exercises, institutional CRIS and national research networks is becoming more clear, partially thanks to new technical solutions and international standards. Citation gathering in an open access environment poses one important challenge, namely figuring out the best way to retrieve counts generated by different versions of the same paper (those available in publisher sites, repositories, reference managers, preprint servers) without leading to distorted numbers. Equally, new approaches to rank academic journals may be enriching for the research community. The Eigenfactor Journal Cost Effectiveness model has added an inspiring nuance in subscription-based and open access journals classifications [56]. Similarly, it may be worth considering their rankings by their overall attention on the Internet.

Altmetrics face the urgent challenge of being able to distinguish quality from non-quality data sources. In addition, rapid changes in the web 2.0 environment may cast a doubt on the durability of some of the new indicators and potential misuse to inflate attention and interest through sites crawled is possible. Other barriers for broader implementation by institutional repositories may involve technical support resources, limited DOI identification in many repositories and author disambiguation. Last but not least, the need to identify datasets through an agreed international standard that eases retrieval of scientific and public interest on the web remains compelling.

Conflict of Interest

The author declares no conflict of interest.

References

- 1. Finardi, U. Correlation between journal impact factor and citation performance: An experimental study. *J. Inf.* **2013**, *7*, 357–370.
- 2. The PLoS Medicine Editors. The impact factor game. *PLoS Med.* **2006**, *3*, e291.

3. The Matthew Effect in Science—Citing the Most Cited. Available online: http://blogs.bbsrc.ac.uk/index.php/2009/03/the-matthew-effect-in-science/ (accessed on 23 April 2013).

- 4. MacRoberts, M.H.; MacRoberts, B.R. Problems of citation analysis: A study of uncited and seldom-cited influences. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 1–12.
- 5. Davis, P.M. Access, Readership, Citations: A Randomized Controlled Trial of Scientific Journal Publishing. Dissertation Presented in 2010 to the Faculty of the Graduate School of Cornell University. Available online: http://ecommons.cornell.edu/bitstream/1813/17788/1/Davis,% 20Philip.pdf (accessed on 27 April 2013).
- 6. Buschman, M.; Michalek, A. Are Alternative Metrics Still Alternative? 2013, ASI&ST Bulletin April/May. Available online: http://www.asis.org/Bulletin/Apr-13/AprMay13_Buschman_Michalek.html (accessed on 25 April 2013).
- 7. Van Noorden, R. Researchers feel pressure to cite superfluous papers, Nature News 2012. Available online: http://www.nature.com/news/researchers-feel-pressure-to-cite-superfluous-papers-1.9968 (accessed on 25 April 2013).
- 8. Lozano, G.; Larivière, V.; Gingras, Y. The weakening relationship between the impact factor and papers' citations in the digital age. *J. Am. Soc. Inf. Sci. Technol.* **2012**, *63*, 2140–2145.
- 9. Joseph, H. The impact of open access on research and scholarship. Reflections on the Berlin 9 Open Access Conference. *Coll. Res. Libr. News* **2012**, *73*, 83–87.
- 10. Davis, P.; Lewenstein, M.; Bruce V.; Simon, D.H.; Booth, J.G.; Connolly, M.J.L. Open access publishing, article downloads, and citations: Randomised controlled trial. *BMJ* **2008**, *337*, a568. Available online: http://www.bmj.com/content/337/bmj.a568 (accessed on 22 April 2013).
- 11. Swan, A. The Open Access citation advantage: Studies and results to date 2010. Available online: http://eprints.soton.ac.uk/268516/ (accessed on 25 April 2013).
- 12. OPCIT. The effect of open access and downloads ("hits") on citation impact: A bibliography of studies Available online: http://opcit.eprints.org/oacitation-biblio.html (accessed on 24 April 2013).
- 13. Antelman, K. Do open access articles have a greater research impact? *Coll. Res. Libr.* **2004**, *65*, 372–382.
- 14. Gargouri, Y.; Hajjem, C.H.; Lariviere, V.; Gingras, Y.; Carr, L.; Brody, T.; Harnad, S. Self-selected or mandated, open access increases citation impact for higher quality research. *PLOS One* **2010**, *5*, e13636.
- 15. Willmott, M.A.; Dunn, K.H.; Duranceau, E.F. The accessibility quotient: A new measure of open access. *J. Libr. Sch. Commun.* **2012**, *I*, eP1025. Available online: http://jlsc-pub.org/jlsc/vol1/iss1/7/ (accessed on 25 April 2013).
- 16. Google Scholar Metrics. Available online: http://scholar.google.com/citations?view_op=top_venues (accessed on 28 April 2013).
- 17. IRUS-UK. Available online: http://www.irus.mimas.ac.uk/ (accessed on 28 April 2013).
- 18. JISC Open Citations. Available online: http://opencitations.net/ (accessed on 27 April 2013).
- 19. Plum Analytics Current List of Metrics. Available online: http://www.plumanalytics.com/metrics.html (accessed on 27 April 2013).

20. Wagner, A.B. Open access citation advantage: An annotated bibliography. *Issues Sci. Technol. Libr.* **2010**. Available online: http://www.istl.org/10-winter/article2.html (accessed on 27 April 2013).

- 21. Terras, M. The impact of social media on the dissemination of research: Results of an experiment. *J. Digit. Hum.* **2012**, *1*. Available online: http://journalofdigitalhumanities.org/1–3/the-impact-of-social-media-on-the-dissemination-of-research-by-melissa-terras/ (accessed on 28 April 2013).
- 22. Lin, J.; Fenner, M. The many faces of article-level metrics. Available online: http://www.asis.org/Bulletin/Apr-13/AprMay13_Lin_Fenner.html (accessed on 23 April 2013).
- 23. Konkie, S.; Scherer, D. New opportunities for repositories in the age of altmetrics. Available online: http://www.asis.org/Bulletin/Apr-13/AprMay13_Konkiel_Scherer.html (accessed on 23 April 2013).
- 24. Priem, J.; Piwowar, H.A.; Hemminger, B.M. Altmetrics in the wild: Using social media to explore scholarly impact. Available online: http://arxiv.org/abs/1203.4745 (accessed on 13 July 2013).
- 25. Brody, T.; Harnad, S.; Carr, L. Earlier web usage statistics as predictors of later citation impact. *J. Am. Soc. Inf. Sci. Technol.* **2006**, *57*, 1060–1072.
- 26. Haustein, S.; Siebenlist, T. Applying social bookmarking data to evaluate journal usage. *J. Inf.* **2011**, *5*, 446–457.
- 27. Institutional repository of Spanish National Research Council, Digital.CSIC. Available online: http://digital.csic.es (accessed on 29 April 2013).
- 28. Bernal, I. Percepciones y participación en el acceso abierto en el CSIC: Informe sobre la Encuesta de Digital.CSIC para investigadores. Report, CSIC 2010. Available online: https://digital.csic.es/handle/10261/28543 (Spanish), https://digital.csic.es/handle/10261/28547 (English) (accessed on 28 April 2013).
- 29. Bernal, I.; Pemau-Alonso, J. Estadísticas para repositorios. Sistema métrico de datos de Digital.CSIC. *El Professional de la Información* **2010**, *19*, 534–543. Available online: http://digital.csic.es/10261/27913 (accessed on 24 April 2013).
- 30. Martínez-Giménez, C.; Albiñana, C. Colaboración de la Biblioteca en el control y la difusión de la producción científica del Instituto: integrando Digital.CSIC, www, Conciencia y Memoria desde el diseño de un Servicio bibliotecario complementario con gran valor potencial y visibilidad. Report, CSIC 2012. Available online: https://digital.csic.es/handle/10261/44840 (accessed on 30 April 2013).
- 31. Swan, A.; Sheridan, B. Open access self-archiving: An author study. Available online: http://www.jisc.ac.uk/uploaded_documents/Open%20Access%20Self%20Archiving-an%20author%20study.pdf (accessed on 3 May 2013).
- 32. Figuerola, J.; Green, A.J. Dispersal of aquatic organisms by waterbirds: A review of past research and priorities for future studies. *Freshw. Biol.* **2002**, *47*, 483–494. Available online: http://digital.csic.es/handle/10261/43045 (accessed on 12 July 2013).
- 33. Williams, P.N.; Villada, A.; Deacon, C.; Raab, A.; Figuerola, J.; Green, A.J.; Feldmann, J.; Meharg, A. Greatly Enhanced Arsenic Shoot Assimilation in Rice Leads to Elevated Grain Levels Compared to Wheat and Barley. *Environ. Sci. Technol.* **2007**, *41*, 6854–6859.

34. Green, A.J: Figuerola, J.; Sánchez, M.I. Implications of waterbird ecology for the dispersal of aquatic organisms. *Acta Oecologica* **2002**, *23*, 177–189. Available online: http://digital.csic.es/handle/10261/43043 (accessed on 12 July 2013).

- 35. Correa Valencia, M. Inteligencia artificial para la predicción y control del acabado superficial en procesos de fresado a alta velocidad, PhD Dissertation Presented in 2010 to the Faculty of Informatics of Madrid Politechnic University. Available online: http://digital.csic.es/handle/10261/33107 (accessed on 12 July 2013).
- 36. Ranz Guerra, C.; Cobo, P. Underwater acoustic tank evaluation of acoustic properties of samples using spectrally dense signals. *J. Acoust. Soc. Am.* **1999**, *105*, 1054. Available online: http://digital.csic.es/handle/10261/7145 (accessed on 12 July 2013).
- 37. Aranda, J.; Armada, M.; Cruz, J.M. Automation for the Maritime Industries; Instituto de Automática Industrial: Madrid, Spain, 2004. Available online: http://digital.csic.es/handle/10261/2906 (accessed on 12 July 2013).
- 38. Galán Allué, J.M. Cuatro viajes en la literatura del antiguo Egipto; Consejo Superior de Investigaciones Científicas: Madrid, Spain, 2000. Available online: http://digital.csic.es/handle/10261/36807 (accessed on 12 July 2013).
- 39. Puertas, F.; Santos, H.; Palacios, M.; Martínez-Ramírez, S. Polycarboxylate superplasticiser admixtures: Effect on hydration, microstructure and rheological behaviour in cement pastes. *Adv. Cem. Res.* **2005**, *17*, 77–89.
- 40. CSIC Open Access Journals. Available online: http://revistas.csic.es/index_en.html (accessed on 2 May 2013).
- 41. Bouillon, J.; Medel, M.D.; Pagès, F.; Gili, J.M.; Boero, F.; Gravili, C. Fauna of the Mediterranean Hydrozoa. *Sci. Mar.* **2004**, *68*, 5–282. Available online: http://digital.csic.es/handle/10261/2366 (accessed on 12 July 2013).
- 42. Winkelmann, I.; Campos, P.F.; Strugnell, J.; Cherel, Y.; Smith, P.J.; Kubodera, T.; Allcock, L.; Kampmann, M.L.; Schroeder, H.; Guerra, A.; Norman, M.; Finn, J.; Ingrao, D.; Clarke, M.; Gilbert, M.T.P. Mitochondrial genome diversity and population structure of the giant squid Architeuthis: Genetics sheds new light on one of the most enigmatic marine species. *Proc. R. Soc. B* **2013**, *280*, 1759. Available online: http://digital.csic.es/handle/10261/73739 (accessed on 12 July 2013).
- 43. Thomas, F. A short account of Leonardo Torres' endless spindle. *Mech. Mach. Theory* **2008**, *43*, 1055–1063. Available online: http://digital.csic.es/handle/10261/30460 (accessed on 12 July 2013).
- 44. Moraga, J.F.H. New Evidence on Emigrant Selection. *Rev. Econ. Stat.* **2011**, *93*, 72–96. Available online: http://digital.csic.es/handle/10261/4353 (accessed on 12 July 2013).
- 45. Ranking Webometrics. Available online: http://research.webometrics.info/ (accessed on 2 May 2013).
- 46. REF. Assessment Framework and Guidance on Submissions. Available online: http://www.ref.ac.uk/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/ 02_11.pdf (accessed on 24 April 2013).

47. Wellcome. Trust Position statement in support of open and unrestricted access to published research. Available online: http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Policy/index.htm (accessed on 23 April 2013).

- 48. HEFCE/REF Open Access and submission to the REF Post-2014, 2013 Letter. Available online: http://www.hefce.ac.uk/media/hefce/content/news/news/2013/open_access_letter.pdf (accessed on 24 April 2013).
- 49. Rentier, B.; Thirion, P. The Liège ORBi model: Mandatory policy without rights retention but linked to assessment processes. Presentation at Berlin 9 Pre-Conference Workshop 2011. Available online: http://www.berlin9.org/bm~doc/berlin9-rentier.pdf (accessed on April 26 2013).
- 50. Guidelines for tenure and promotion reviews at Indiana University Bloomington 2012. Available online: http://www.indiana.edu/~vpfaa/WRAP_SDDU/BL-PROV-COMM/VPFAA%20-%20New/docs/promotion tenure reappointment/pt-revised-review-guidelines.pdf (accessed on 26 April 2013).
- 51. Spain's Law of Science, Technology and Innovation 2011, Article 37 "Dissemination in open access". Available online: http://www.boe.es/boe/dias/2011/06/02/pdfs/BOE-A-2011-9617.pdf (accessed on 12 July 2013).
- 52. Open Access policy by Community of Madrid with regard to funded CSIC research (2008) and to general funding calls (2009). Available online: https://www.madrimasd.org/quadrivium/convocatorias/Portals/13/Documentacion/ordencsic2008. pdf and http://www.madrimasd.org/informacionidi/convocatorias/2009/documentos/Orden_679-2009_19-02-09_Convocatoria_Ayuda_Programas_Actividades_Tecnonologia.pdf (accessed on 12 July 2013).
- 53. Open access policy by Principality of Asturias 2009. Available online: https://sede.asturias.es/portal/site/Asturias/menuitem.1003733838db7342ebc4e191100000f7/?vgn extoid=d7d79d16b61ee010VgnVCM1000000100007fRCRD&fecha=03/02/2009&refArticulo=20 09-03201 (accessed on 12 July 2013).
- 54. Bernal, I.; Román-Molina, J.; Pemau-Alonso, J.; Chacón, A. CSIC Bridge: Linking Digital.CSIC to Institutional CRIS, Poster OR2012, Edinburgh. Available online: https://digital.csic.es/handle/10261/54304 (accessed on 3 May 2013).
- 55. CSIC Institutional Action Plan CSIC 2010–2013. Available online: http://www.csic.es/web/guest/plan-de-actuacion-2010–2013 (accessed on 29 April 2013).
- 56. Eigenfactor Journal Cost Effectiveness for Open Access Journals. Available online: http://www.eigenfactor.org/openaccess/ (accessed on 25 April 2013).
- © 2013 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).