Snapshot of the uses of science

DOI: 10.4000/books.oep.1640
Publisher: OpenEdition Press
Place of publication: Marseille
Year of publication: 2016
Published on OpenEdition Books: 29 avril 2016
Serie: Laboratoire d'idées
Electronic ISBN: 9782821868700

http://books.openedition.org

Electronic reference

This text was automatically generated on 16 septembre 2019.
1 Science is in transition towards a new system of rights and major changes in its practices.
2 The new digital STI practices followed by researchers and laboratories are out of phase with the existing legal framework, which is lacking in some areas and has shortcomings in others.
3 Science is also grappling “with the new risk of appropriation of its data, mainly by scientific publishers demanding licence transfers for the datasets integrated or associated with the research articles they publish”. The explanatory statement for the Digital Republic Bill affirms these findings, which were confirmed in particular by the national consultation, by the French Digital Council, as well as by the impact assessment for the Bill.

**Snapshot of the uses of science**

4 The practices of science place it at the heart of the digital transition.
5 This section offers a snapshot of the practices in force in France and abroad concerning:
   - the use of STI:
     - by researchers according to the scientific community to which they belong, in particular the use of STI as a research tool;
     - by laboratories, institutes and agencies dealing in STI, with regard to organisation of access to STI;
   - scientific publishing contracts with regard to the conditions of publication, in particular in public–private partnerships.
Science at the heart of the digital transition

The digital transition is marked by an explosion in the quantity of data that are available and accessible at any time and any place in the world. This age of Big Data is characterised by:

- the absence of borders and the globalisation of information;
- a spirit of cooperation and sharing;
- the automation of certain activities, especially professional ones;
- the creation of value.

The absence of borders

The Internet is the flagship tool of the digital transition. This global network is notable for the absence of physical borders and entry visas, being instead characterised by cross-border flows and exchanges through the world of information.

Thanks to the IP protocol and the interconnection between the various operators, the Internet constitutes a seamless end-to-end network.

Despite flaws and inequalities such as the digital divide or equipment with differential flow rates according to a country’s public policies, this global network provides everyone with access to globalised information.

The right to cooperation and sharing

The digital transition is marked by the passage from print to digital, in other words a shift from a single physical property to a multiplication of that property.

The Web 2.0 generation is characterised in particular by the ease of placing content online and the ability of users to interact; it is based on users themselves being able to generate, disseminate and consult multimedia content directly.

The principle of the collectivisation of content has fundamentally transformed uses. Sharing and a spirit of cooperation characterise the Web model. Its archetypes include YouTube, Wikipedia and social networks.

The right to automation

Numerous technological applications have resulted in authorities, research organisations and businesses holding large quantities of data and metadata.

Semantic and lexicographical analysis tools need to be developed in order to process these data. Dematerialised or automated work processes enable the creation of content by the user: user-generated content.

The right to creation of value

The digital transition is generating a new so-called digital economy. It is based on online trade as well as on non-commercial exchanges such as the sharing of knowledge and user-generated content (usage data, wikis, discussion forums, blogs, tweets, etc.).
The OECD’s report of 12 April 2007 had already highlighted the very significant economic impact of user-generated content, and the creation of the resulting value and innovation.

**New uses by researchers**

The CNRS survey and the hearings revealed the practices of researchers and institutes regarding the use of digital STI.

**Hearing at the University of Strasbourg: Paul-Antoine Hervieux, 10 July 2015**

“Uses are changing and the paradigm for research data (BSN10 – research data) is undergoing a transformation. Researchers and academics are beginning to realise the value of their data. Once we start talking about value, we need to start thinking about rights. We are now in a world where private businesses predominate and are increasingly interested in data with a view to commercialising uses.”

The following findings emerged regarding the way researchers use STI as part of their research activity:

- STI is a working tool that can be used, shared and exploited freely as part of research work;
- databases in the form of open archives are used in a heterogeneous manner by the different communities;
- scientific communities lack sufficient knowledge of copyright and the **sui generis** right for databases to effectively manage the use of STI.

**Digital STI: A tool for exploration and analysis**

STI is primarily seen by researchers as a working instrument to which they must have access, which they can share, exchange, reuse or reprocess for the needs of their research subject. As a raw material, STI is part of the scientific process.

Scientific communities distinguish two main categories of data in STI:

- research data;
- publications.

Other categories of data are specific to certain communities:

- material: especially in biology;
- third-party data: in particular in the human and social sciences (HSS).

**Access to STI.** The contribution of the CNRS Scientific Board helped identify the main practices of researchers in the area of access to STI: access to publications, access to research data and, for some communities, access to the research material and to the third-party data used in the framework of their research.
| Human and social sciences | Access to publications: “In the human and social sciences, with regard to recent scientific publications, while more and more French-language journals offer free access immediately or after a few years (mainly through HumaNum, BSN or OpenEdition), English-language journals are often confined to rather expensive platforms.” |
| Access to numerical data or the data produced by research: “Platforms for exchanging numerical data ... have also been established (Quetelet Network, DIMEHS, etc.): they provide better circulation of data, compliance with the necessary constraints, such as anonymisation, and the documentation (‘metadata’) without which the figures would be unusable.” |
| Access to third-party data: “The problem here is that some of the data used by scientists in HSS are not produced by them (this may concern a song, a company’s annual report or the architecture of a monument): other natural or legal entities have rights over them.” |
| Science of the universe | Access to research data: “The data are freely accessible in astronomy for the entire community following expiration of a proprietary period.” |
| Biology | Access to publications: “Digital publishing is widespread. Academic institutions have developed platforms to help researchers find articles, and access to abstracts is free. The most important of these (PubMed) is managed by the US National Institutes of Health (NIH).” |
| PubMed also provides access to an unformatted version of any article published commercially describing work funded by the NIH. |
| Access to research data: “Many publishers, including Nature, also make publication of an article conditional on the depositing of mass data associated with a publication on a platform that is accessible to all, free of charge.” |
| Access to material: “It should be noted that this requirement goes beyond digital data and also concerns material produced within the framework of a publication.” |
| Physics | Access to research data: “Free access to raw data is not yet very widespread.” |
| “Many digital libraries have been formed and made freely accessible by groups of researchers.” |
| Chemistry | Access to publications: “The rule is still that papers appear in paid journals published by learned societies (American Chemical Society, Royal Society) or commercial companies (Wiley, Elsevier, etc.), with the timid development of ‘Gold’-type open access, paid for by the authors.” |
| “Freely accessible databases are developing, especially the Cambridge Structural Database, which contains all the published molecular structures.” |
Access to publications: “Databases relating to publications are very important for both individual and community work. A unique feature of this discipline is the importance of easy access to both recent and ‘old’ publications (dating from several years, decades or even centuries ago). Long-term access to these publications is therefore crucial for research.”

Access to these categories of data is obtained by means of several technical tools:

- publisher portals;
- digitised archives;
- voluntary submission platforms;
- paper (question PAP 1-6 from the Shared Action Plan in the CNRS survey notes “the historical attachment of mathematics and the HSS to paper documentation resulting from the needs for/uses of old documents”).

Processing of STI. The scientific communities have different approaches to the sharing of data as well as to the techniques of text and data mining.

<table>
<thead>
<tr>
<th>Community</th>
<th>Processing of STI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human and social sciences</td>
<td>“Data sharing and ‘text and data mining’ techniques are thus unevenly spread according to the types of data, mainly due to legal obstacles [third-party data], and a lack of human resources for the production and maintenance of quality metadata.”</td>
</tr>
<tr>
<td>Science of the universe</td>
<td>“The formats, descriptions and modes of access to archive data, metadata and the applications likely to be used to process them should be harmonised and standardised, in order to achieve interoperability.”</td>
</tr>
<tr>
<td>Biology</td>
<td>“While text-mining techniques are not a priority for most fields of biology as a discovery tool (but rather in terms of documentary collection), ‘data mining’ itself is playing an increasingly important role”.</td>
</tr>
<tr>
<td>Computing and mathematics</td>
<td>“Furthermore, mathematics and computing play an important role in the analysis, management and exploitation of masses of data (questions surrounding Big Data). It is certainly very important for the data to be accessible, but when they become more and more massive, it must also be possible to exploit them effectively.”</td>
</tr>
</tbody>
</table>

Sharing of knowledge: Depositing of scientific results in open digital archives

Disparate practices. The CNRS survey and the contribution of its Scientific Board revealed different uses of open archives according to the given scientific community.

The HAL (Hyper Articles OnLine) multidisciplinary open archive is the one used most widely by the national community. Its purpose is the depositing and dissemination of scientific documents in any field of research; these have not necessarily been validated by an editorial board or programme committee. These documents come from researchers in French or foreign, public or private teaching and research institutions.
Question PAP 2-26 from the Shared Action Plan shows that different scientific communities use HAL in different ways. Researchers at the National Institute for Mathematical Sciences (INSMI) and those at the National Institute of Nuclear and Particle Physics (IN2P3) use HAL extensively. These institutes use it for historical purposes, while researchers from the Institute of Chemistry (INC) or the Institute of Biological Sciences (INSB) use it very little, and use other archives rarely or not at all.

In total, 69% of those communities not using HAL as an open archive do not use other archives either (PAP 2-29). Many of the publications are therefore merely archived by the publishers and are not freely available to the scientific communities after an embargo period. Access to publications is possible only via the publisher’s platform, by subscription or the one-off purchase of the article.

PAP 2, question 26, results of the survey on the uses and needs of STI – CNRS – March 2015

Mixed use of HAL: INSMI/IN2P3 (historical and essential) versus INC/INSB (no dissemination of preprints)
Better adapt HAL to the practices of the communities
The contributions of the Scientific Board provided some clarification on the practice of depositing in open archives:

<table>
<thead>
<tr>
<th>Community</th>
<th>Submission practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human and social sciences</td>
<td>The &quot;parallel depositing of articles in open archives is not greatly developed&quot;.</td>
</tr>
<tr>
<td>Science of the universe</td>
<td>Very widespread practice.</td>
</tr>
<tr>
<td>Physics</td>
<td><strong>Pre-publication server (pre-print):</strong> “Digital technologies also play an essential role in the dissemination of results, with the almost systematic use of pre-publication servers. Articles are deposited on these servers at the same time as they are sent to a peer-reviewed scientific journal: this enables readers to take early notice, prior to publication.”</td>
</tr>
<tr>
<td>Chemistry</td>
<td>“There is no pre-publication archive like ArXiv.”</td>
</tr>
<tr>
<td>Computing and mathematics</td>
<td><strong>Archiving platform:</strong> “Publication archiving platforms such as HAL or ArXiv thus respond in part to this problem and should be supported, along with metadata platforms (MathSciNet, Zentralblatt, etc.).”</td>
</tr>
</tbody>
</table>

**Data deposited.** The data deposited by researchers in open archives differ from one community to another:

- pre-print publications (article, book, chapter of a book);
- post-print publications (article, book, chapter of a book);
• thesis/dissertation (PhD, master’s, etc.);
• lessons;
• images, videos, sounds, maps;
• bibliographic records only;
• bibliographic records accompanied by the publication;
• research data;
• metadata associated with these data.

The CNRS survey revealed that of the communities that deposit data in HAL:
• 63% deposit bibliographic records (PAP 2-27);
• 56% deposit the full text (PAP 2-28).

In total, 69% of the communities that do not deposit data in HAL do not use other archives either.

Rights on uses to be created from scratch

Publication, deposit in open archives, text and data mining, or operations relating to other data- or text-processing or mining techniques are carried out by researchers with an almost complete lack of awareness of intellectual property rights.

Hearing at the University of Strasbourg: Paul-Antoine Hervieux, 10 July 2015
“In academic communities there is an almost across-the-board lack of awareness of copyright and its implementation measures, regardless of the type of document or data.”

In the framework of the CNRS survey, two questions were used to assess the researchers’ level of knowledge about their rights.

To the question “Do you think that these data [the raw data] are copyright-free?”, the answers are divided between “Yes” (29%), “No” (39%) and “I don’t know” (32%).
PAP 3, question 59, results of the survey on the uses and needs of STI – CNRS – March 2015

Do you think that these data are copyright-free?

Information is needed on data rights

The breakdown of responses to this question in three equivalent groups shows that the researchers:

- cannot distinguish between what is copyright-free and what is protected by an intellectual property right;
- do not know their rights over the raw data and by extension the protected data (the publications).

The second question revealing their ignorance is the following:

PAP 4, question 85, results of the survey on the uses and needs of STI – CNRS – March 2015

Have you already been faced with legal issues about the digitisation and posting online of content owned by your unit?
It can be seen that 63% of respondents have never been faced with legal issues concerning the digitisation and posting online of content. To be “faced with legal issues”, there must first be an awareness of the existence of problems raised by digitisation and posting online with regard to copyright.

This lack of awareness of rights concerning research data reveals that illegal practices can occur through ignorance, and indicates the need for clarification or even affirmation of the legal position with regard to these practices and the needs of researchers. Training initiatives will be necessary to support these desirable legislative developments.

Pressing demand from laboratories, institutes and agencies

The multiplication of platforms and the multiplicity of STI objects have generated a growing need for governance and the definition of a legal framework for science in the digital age.

Multiplication of platforms: A need for governance

Multiplication of platforms. A multiplication in the number of STI platforms has been observed in recent years:

• institutional platforms;
• thematic platforms;
• submission platforms;
• bibliographic platforms;
• archive platforms, etc.

This multiplication has resulted in a dilution of the information and a risk that it could lose its value. It also prevents cross-referenced information searches as well as multidisciplinary searches. Some platforms offer value-added services while others offer simple consultation of the texts, or simple access to the bibliographic data. This multiplication has a particularly high cost, whether regarding the quality of the resulting services or in human or financial terms.

Weakness of the legal framework of the platforms. These various platforms have more or less restrictive general terms and conditions of use with regard to the use of STI, which are not always in line with intellectual property rights and the publishing contracts entered into between researchers and publishers.

The Digital Republic Bill proposes to introduce a definition of the concept of “platform” and to associate it with a duty to act in good faith. These provisions are introduced in the French Consumer Code but could be extended, as “good practices”, to science platforms.

* The multiplication of the platforms and the weakness of their contractual framework have generated a need for new governance
Multiplicity of STI objects (data, analyses, articles): A need for consistency for their exploitation

Exploiting the fruits of public research is an essential concern, and the filing of patents is at the centre of the exploitation process.

According to the 2014 SIR (Scimago Institutions Rankings) classification, the CNRS is the world’s leading filer of patents from institutions’ scientific publications. Most of these patents are jointly owned with universities. This achievement is the result of cutting-edge research but also an incentivising exploitation policy. French research is generally active and produces many patentable innovations.

Public research has established a policy to incentivise the research units from which the inventions and their inventors come. Staff members thus receive a lump sum bonus for patents, whose amount is set by joint order of the ministers responsible for the budget, public service and research. A coefficient reflecting their contribution to the invention is assigned to each staff member involved.

Patents are not, however, the only way to exploit results, and may prove to be unsuitable for some innovations, for which the filing of a patent is impossible or inappropriate.

STI covers multiple elements: data, analyses, results, processed results, articles, search queries, user-generated content, etc., all of which are potential sources of exploitation that should be taken advantage of by French research.

Open access to scientific platforms should not compromise the exploitation of STI in any of its different components.

The search for a balance between exploitation and open access requires clarification of the rights and the legal regime applicable to the STI objects, and a dividing line to be drawn between the common good and the protected innovation, as well as between freedom of access and private reservation.

+ The multiplicity of STI objects calls for a clarification of the law and a balance to be struck between access to scientific knowledge and preservation of the potential for STI exploitation in all its components.

NOTES

3. http://www.bibliothequescientifiquesnumerique.fr/bsn-10-donnees-de-la-recherche/
5. **Version submitted** ("Pre-print"): version submitted by the author(s) to the journal before any process of revision (peer-reviewing by publishers and referees).

6. **Accepted version** ("Post-print"): version after the author(s) have taken into account the remarks of the evaluators and the article is accepted by the editorial board.
