Climate Information Portal

Management of Intellectual Property of Data Resources

"Gate-keeping" vs. "Open Door"

Martin Juckes, October 2014

Table of Contents

1 Introduction

This paper addresses the protocols for managing Intellectual Property associated with data. There are two aspects of intellectual property: restrictions imposed by the nature of the IPR and restrictions associated with how it is managed. In many cases the latter restrictions, which are associated with the protocols used, are a greater barrier to data exploitation that the restrictions associated with the nature of the IPR.

The term "Open Door" used here is not synonymous with "Open Access": the latter implies a specific licensing policy, whereas "Open Door" here refers to an approach to implementation. The distinction between these two is discussed further in section 3 below.

Data is covered by existing international copyright laws. The relevant paragraph of the World Intellectual Property Organization (WIPO) Copyright Treaty states that "Compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations, are protected as such. This protection does not extend to the data or the material itself and is without prejudice to any copyright subsisting in the data or material contained in the compilation."[1] The copyright laws apply automatically, unless exemptions are granted by putting information in the public domain, without the need for explicit licensing or acceptance of terms by people who have access to the material. The Copyright Treaty obliges partners to enact laws which prevent removal or alteration of electronic rights management information without authority.

There is, naturally, a mechanism for electronic management of rights information in the European Union INSPIRE directive. The system described, however, only deals with the encoding of rights information in an XML document.

Much of electronic rights management technology is aimed at the production of tamper-proof systems for enforcement of rights in a commercial environment. This is rather different from the requirements associated with environmental data: here we would like a robust system based on well known standards to ensure that appropriate rights information is passed on through any processing chain.

2 "Gate-keeping" versus "Open Door"

Data portals often use an approach to IPR management which interrupts access. Typically, a user navigates to a resource and when trying to access the resource is presented with a legal document which they must agree to before proceeding. This approach, referred to as "gate-keeping" here, ensures that the user has had sight of the relevant document before getting at the data, provided the user complies with the terms which often prohibit sharing of data in order to ensure that all users must come through the designated access route.

An alternative approach is to declare the owner of the data and the license conditions clearly. This approach is followed by the printed media and has worked for centuries. It has also been adopted by the UK Government for

1 WIPO Copyright Treaty: http://www.wipo.int/treaties/en/text.jsp?file_id=295166#P59_6206

all official data in the UK.

With the “Access-interrupt” approach some form of session-memory is usually employed to ensure that the inconvenience to the user is limited. However, for every new access method a new method of interrupting access needs to be defined. The problem becomes more severe when data is accessed indirectly through a chain of services. There are ways of dealing with these problems, but they carry a cost. The “Open Door” approach, coupled with a declared license, gives far greater flexibility. There are still technical issues to be resolved when data is being passed through chained services: a simple link to a license in a web site will not be enough if users do not have to visit the site, and information in the file will not be enough if users can get at the data without downloading the whole file. A cross domain standard for inclusion of licensing information in files would be beneficial.

For example, the OPeNDAP protocol allows users to access specific components of a file without accessing the complete file. This protocol does not, however, support data discovery (neither browsing of directories nor searching of catalogues). Arguably, the appropriate place to ensure that the copyright and license information is visible is the discovery stage. Additional steps might be taken to ensure that community tools which process NetCDF files will propagate copyright and license information as appropriate (i.e. use appropriate defaults – it will not be possible to prevent users from overwriting such information, and the objective is to give users the opportunity to develop new products which will carry their own intellectual property).

3 “Open Access” versus “Open Door”

The discussion of how to implement licensing policy is often overshadowed by discussion of the content of the terms of use. The two are related, but also have a degree of independence. The term “Open Access” refers to an approach of provided unrestricted work-flows to access data and unrestrictive terms of use. Typical “Open Access” terms of use refer to an obligation to provide proper credit to the producers of the data. Restrictive work-flows may be imposed for a number of reasons: (1) to prevent un-authorised access where restrictive license conditions apply; (2) to enable detailed analysis of usage patterns; (3) to enable “data recall” and (4) to manage load on services constrained by resources.

Restrictive license conditions

If the data really needs restrictive license conditions, a work-flow with interruptions for authentication and authorisation of users is unavoidable. Such data is likely to be distributed through a limited number of carefully controlled channels.

Usage patterns

Data providers have a requirement for information about data usage. Gate-keeping at the data archive is a natural approach for archive managers to follow in order to generate clear objective information on data usage. Despite being clear and objective, the figures generated may not be very informative. Many users who take data from an archive will be browsing, the key information on what is actually done with the data cannot be collected at the archive gate. It makes sense to separate the question on usage into two: who is browsing the data and who is using the data? The answer to the first question is usually given, in the web context, in terms of usage patterns determined from service logs which record the identity of the machine to which resources are delivered. The number machines is taken as a surrogate for the number of users (e.g. illustration 1, on web site usage provided by the Google Analytics system). This approach does not rely on collecting personal information about users. For the second question, what is done with the data, we need to look not at the people who take the data out of the archive gate, but at the people who publish results. The appropriate system here is the internationally managed citation index which collects and manages the information provided by the authors about the resources they have exploited.

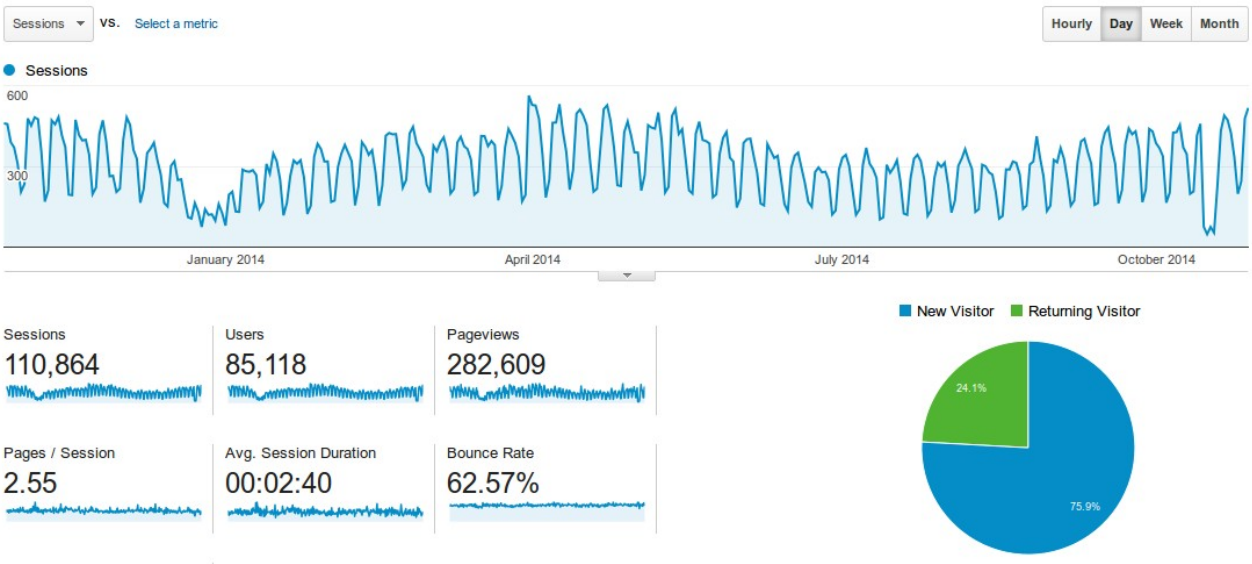

Illustration 1: Example web site usage report from Google analytics.

Data recall

From time to time errors may be found in data and there is consequently a need to be able to contact people using the data. With the gate-keeping system, details of everyone using data are actively collected and the terms of use must be phrased to ensure that data is either not passed on or only passed on to people who have registered.

Resource management

A system of registering users and actively controlling access provides system administrators with additional tools for managing the load on resources. Increasingly, however, resource management software is well separated from the user access control tools. Integration of these very different categories of software tools is not a priority, so cannot be expected to happen in the near future. Within ESGF, for instance, resource management was implemented through restrictions on the number of requests from a given IP address which could be handled simultaneously. The benefits of access restrictions for resource management appear to be more theoretical than practical.

4 What happens if we abandon access interruptions?

Can we guarantee that a user will only access the data through routes along which the declared license is clearly indicated? The answer is probably not. The Copyright Treaty cited above perhaps gives an indication as to how this can be dealt with: data should not be distributed through a protocol which does not support reliable propagation of license information unless the copyright holders have granted an exemption.

5 Case studies

For years the United States has set the standards for openness of data, but the UK Government has moved ahead recently with a new open data policy.

UK Government Data

UK has placed all official, public data under its control on http://data.gov.uk/ . All data is available without access interruptions (e.g. http://www.hscic.gov.uk/catalogue/PUB13648/Obes-phys-acti-diet-eng-2014-tab.xls). There is

a “Terms and Conditions” page (http://data.gov.uk/terms-and-conditions) linked from every page on the site, which explains basic terms and conditions and links to the license which applies to all data downloaded from the site: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/. This approach replaces the earlier “Click-Use” license (a form of gate-keeping in which the authorisation to use data is granted automatically once a user has clicked an appropriate acknowledgement): “The Open Government Licence replaces the Click-Use PSI licence and makes it faster and easier than ever before to freely re-use public sector information (including Crown copyright information)” (http://www.opsi.gov.uk/psi/).

NOAA Data Centres

The data in NOAA data centres is generally available with no access control. In some cases a user approaching the data through the data centre web pages will be invited to register, but this is not generally obligatory. Different centres have slightly different approaches .. for example the Climate Date Record Program has a document which urges users to use an appropriate citation (http://www1.ncdc.noaa.gov/pub/data/sds/cdr/CDRs/AMSU-A%20Ch7%20Mean%20Temperature/UseAgreement.pdf).

Copernicus land service

Box 1: Rights information for EU-DEM dataset (http://www.eea.europa.eu/data-and-maps/data/eu-dem)

Access to the data is governed by the draft delegated regulation on Copernicus data and information policy, as approved by the EC on 12th of July 2013, and in the process of decision making by the Council and European Parliament. This delegated act supplements regulation (EU) No 911/2010 of the European Parliament and of the Council on the European Earth monitoring programme (GMES). It establishes registration and licensing conditions for GMES/Copernicus users and defines criteria for restricting access to GMES/Copernicus dedicated data and GMES/Copernicus service information.

The following credit must be displayed when using these data: "Produced using Copernicus data and information funded by the European Union - EU-DEM layers."

Access and use of the data is made on the conditions that:
1. When distributing or communicating Copernicus data and information to the public, users shall inform the public of the source of that data and information.
2. Users shall make sure not to convey the impression to the public that the user's activities are officially endorsed by the Union.
3. Where that data or information has been adapted or modified, the user shall clearly state this.

The Copernicus land service (http://land.copernicus.eu) provides “Open Door” access to data through a browsable archive, with a “Rights” notice posted on the the data set landing page (see example in box 1). Data can be downloaded in GeoTIFF files[2] – users accessing these links do not need to come into contact with the rights declaration.

6 Ambiguity

There is some ambiguity about intellectual property and data, just as there is around copyright and creative works: when is a new digital collection sufficiently original to merit a new intellectual owner? There is a substantial history of past decisions to demarcate the boundary for creative works, less so for data collections. A peer reviewed publication is clearly a badge of new intellectual property, but the link between new datasets and associated publications is not always clear: it may be that the intellectual ideas leading to a publication come from theoretical developments and the new collection of data associated with the publication is a near trivial re-arrangement of existing data resources. Copyright legislation does allow for this situation, making provision for an author to claim copyright for the organisation of a collection of objects, while the component objects retain the

2 Sample EU-DEM file (300Mb): http://published-files.eea.europa.eu/eudem/entr_r_4258_1_arcsec_gsgrda-eudem-dem-europe_2012_rev1/eudem_tiles_5deg/eudem_dem_5deg_n50w015.tif

copyright of the original authors.

7 Conclusion

It appears clear that there is no general justification for a gate-keeping approach. Such restrictive practises may be needed for some datasets, but such datasets should be considered as special cases and not necessarily become available through a full range of services. There will be a distinction between Copernicus owned data, which will fall under an Open Access policy, and third party datasets. The policy should be to encourage providers of third party datasets to accept an Open Door approach, relying on implicit rights to protect their copyright and, where this not an option, to provided only limited access to data under a gate-keeping approach.

The Copernicus regulation, passed by the European Parliament in 2012[3] envisages full and open access to all Copernicus data for a user community defined as "*those comprising the European national, regional or local bodies entrusted with the definition, implementation, enforcement or monitoring of a public service or policy in areas referred to in point (1) of Article 4.*" CLIPC has additional requirements to provide a comprehensive knowledge base. The rather narrow user community referenced in the regulations do not generally make direct use of the data. In order of a Climate Information Portal to have sufficient credibility to support policy needs the data must be open to all users who have an interest in the data.

3 REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing the Copernicus Programme and repealing Regulation (EU) No 911/2010
http://www.copernicus.eu/pages-principales/library/policy-documents/?no_cache=1&cHash=b537b7aba03c64288a16f9e66140d6f2