# Center for Research Libraries

# Model Licensing Terms and Specifications for Global Data Resources

This report documents a set of licensing terms and specifications recommended for use when negotiating purchase agreements for large, global datasets and databases provided by commercial vendors in the areas of business, financial, and geospatial data.

## Background

In November 2016, CRL hosted the inaugural eDesiderata Forum on the topic of licensing very large databases and datasets (the event description, related resources and documentation can be found [here](https://www.crl.edu/events/licensing-big-data-edesiderata-forum)). The Forum invited experts in the fields of geospatial, business, census and public opinion data to discuss the challenges libraries face in licensing and acquiring such resources with the CRL community. Challenges identified include: limitations on use of licensed materials; limitations on local storage of materials; inconsistencies in vendors’ delivery and access mechanisms; and lack of established rights for long-term use and retention of the data. The full Forum summary report can be found [here](https://www.crl.edu/sites/default/files/event_materials/eDesiderata_Forum_Summary_Report_Jan_10_2017.pdf).

In acquiring datasets and statistical databases, libraries must often deal with commercial vendors who are unfamiliar with the academic market and whose standard licensing terms are ill suited for academic use and their specific requirements. It is imperative for libraries to establish clear terms for how data can be used in teaching and research as well as pricing models for data appropriate to the academic environment. These negotiations are most successful when academic use cases are presented clearly and effectively.

Guided by the findings of the 2016 Forum, and in and consultation of experts in data resource licensing at academic libraries, CRL has developed a set of terms and specifications recommended for inclusion in data acquisition licenses to facilitate more successful, informative and mutually constructive agreements with vendors of global data resources.

## Acknowledgements

CRL acknowledges and thanks the following individuals and organizations for their assistance and contributions to the Model Licensing Terms and Specifications.

* Amber Leahey, Data and GIS Librarian, Scholars Portal, Ontario Council of University Libraries
* Marcel Fortin, Head, Map and Data Library University of Toronto
* Carol Stephenson, Electronic Resources Licensing Consultant, Ontario Council of University Libraries
* Alex Carracuzzo, Senior Manager, Baker Research Services at Harvard Business School
* Susan Powell, GIS & Map Librarian, University of California Berkley
* Lara Cleveland, Senior Research Scientist, IPUMS Center for Data Integration

## International Research Consortia Model Licenses

CRL works with other national consortia and research institutions to promote high standards for academic resource licensing across the globe. In the use of the recommended terms and specifications suggested herein, CRL references relevant sections of the [LIBLICENSE Model License Agreement](http://liblicense.crl.edu/licensing-information/model-license/) to which the suggested language may be applied.

In addition to LIBLICENSE, CRL drew upon the following model licenses in construction of these recommendations:

* Canadian Research Knowledge Network (CRKN) [Model License](https://www.crkn-rcdr.ca/sites/crkn/files/2016-09/crkn_model_license_2016_final.pdf)
* California Digital Library (CDL) [Model License](http://cdlib.org/gateways/vendors/docs/CDL_Model_License_2016_public_version_final.docx)
* JISC [Database Model Licenses](http://www.jisc-collections.ac.uk/model_licence)
* Ontario Council of University Libraries (OCUL) Model Data License (*draft*)
* Fachinformationsdienst (FID) [Musterlizenzvertrag](https://www.fid-lizenzen.de/dateien/musterlizenz-2016-10-27) (Model License)

## Licensing Terms - Definitions

*To provide clarity in data licenses, it is advisable to include definitions for terms such as Authorized Users, Perpetual Access, Secure Network, and Server. Additionally, the following terms are used in the licensing language suggested below. The definitions supplied here may be added to license “definitions” or within clauses as appropriate.*

**Derivative**: a work based on or derived from already existing material. A derivative work must incorporate some or all of a preexisting material and add new original authorship to that work. [[1]](#footnote-1)

**Text and Data Mining (TDM)**: A machine process by which information may be derived from the Licensed Materials by identifying patterns and trends within natural language and/or numeric data through methods such as text categorization, statistical pattern recognition, concept or sentiment extraction, and the association of natural language with indexing terms.

## Licensing Terms - Authorized Uses

*Academic use of datasets is inherently different than that of “traditional’ library e-resources for which most standard model licenses have been designed. Beyond citation, data and datasets may be used as raw material or components of scholarly products. GIS datasets are often incorporated into locally hosted software and systems, for example; and financial data incorporated into economic modeling applications. Licenses should address the full range of use cases for academic teaching and research, as well as terms for library hosting, user access, and long-term retention. The following rights should be incorporated into “Authorized Uses” clauses as appropriate.* [[2]](#footnote-2)

The Licensee and Authorized Users may access or use the Licensed Materials for the purposes of research, education, distance learning, or other non-commercial use as follows:

a. **Display.** Licensee and Authorized Users may electronically access the Licensed Material by Secure Authentication in order to search, retrieve, display and view the Licensed Material.

b. **Digitally Copy.** Licensee and Authorized Users may download and digitally copy a reasonable portion of the Licensed Materials.

c. **Print Copy**. Licensee and Authorized Users may print a reasonable portion of the Licensed Materials. Licensee may charge a reasonable fee to recover costs of copying or printing portions of Licensed Materials for Authorized Users.

d. **Caching.** Licensee and Authorized Users may make local digital copies of Licensed Materials to the extent necessary to ensure efficient use by Authorized Users in connection with their work. For the avoidance of doubt, the cached copy is not a Derivative work.

e. **Backup Copy**. Licensee may create one (1) copy of the entire set of Licensed Materials to be maintained as a backup copy. In the event that the Agreement is terminated, Licensee may use the backup copy to exercise any perpetual license rights granted in this Agreement.

f. **Use in Education, Teaching and Research.** Licensee and Authorized Users may extract and use excerpts from the Licensed Materials for academic research, scholarship, and other educational purposes, including extraction and manipulation for the purpose of illustration, explanation, example, comment, criticism, teaching, research, and analysis.

Reasonably small but representative samples of the data may be included in scholarly research and educational materials and may be displayed/presented as permitted by the limitations and exceptions of the U.S. Copyright Act (17 U.S.C. §101, et seq.).

d. **Scholarly Citation.** Authorized Users may use, with appropriate credit, figures, tables, and brief excerpts from the Licensed Materials in the Authorized User’s own scientific, scholarly, and educational works. Authorized Users may self-archive or deposit in perpetuity in a relevant data repository excerpts from the Licensed Materials used in connection with the Authorized User’s work, and to comply with all grant or institutional requirements for depositing data associated with the work.

e. **Electronic Learning Environment.** Authorized Users may incorporate parts of the Licensed Materials to be used in virtual and management environments (including but not limited to virtual learning environment, managed learning environments, virtual research environments, library environments, and courseware technologies) hosted on a Secure Network. Each item shall carry appropriate acknowledgement of the source, listing title and copyright owner.

f. **Text and Data Mining**. Authorized Users may use the Licensed Materials to perform and engage in Text and Data Mining (TDM) activities for academic research, scholarship, and other educational purposes, utilize and share the results of TDM in their scholarly work, and make the results available for use by others, so long as the purpose is not to create a product for use by third parties that would substitute for the Licensed Materials.

g. **Derivatives.** Authorized Users may selectively extract datasets from the Licensed Materials; translate, adapt, alter, transform, modify or arrange data; and electronically converge and merge additional datasets in the most appropriate form for their work. Broadly distributed derivative works must be created in such a way that they cannot be converted back to original source data. Authorized Users will retain ownership of all derivative works provided that derivatives are subject to restrictions regarding confidentiality, use, and transfer covered herein. Derivative works shall include proper citation of the original Licensed Material as a source.

## Licensing Terms - Licensor Performance Obligations

*Vendors may use a variety of formats and delivery methods that are unfamiliar or unacceptable for storage and management in library systems. Data to be acquired must be evaluated with regard to compatibility with existing university technology. Vendors may need to be explicitly instructed on acceptable delivery formats for data, metadata, and documentation required by the library to facilitate access and local hosting.*

*Additionally, in some cases vendors may expect that previous versions of data will be deleted as updates are provided (affecting scholarly citations to earlier versions of the data). The following clauses should be incorporated into “Licensor Performance Obligations” as appropriate (Schedules are further described in Section 5).*[[3]](#footnote-3)

a. **Availability of Licensed Materials.** The Licensor will make the Licensed Materials available to the Licensee and Authorized Users according to the terms set forth in Schedule [X] below. The Licensor will make the Licensed Materials available to the Licensee in specified delivery formats for data, metadata, and documentation required by the library as set forth in Schedule [X], below.

*[In cases of local hosting, include:]* Licensee will load and host the Licensed Materials on a Server on its Secure Network for use and access by Authorized Users.

b. **Updates.** The Licensor may from time to time provide updated data to the Licensee; all updates will be covered by and subject to the provisions of this agreement. The Licensor will specify the changes made to the data. Unless specifically prohibited in this agreement, Licensee retains the right to archive and continue to provide access to prior versions of the data, explicitly identifying those versions as superseded, obsolete, or otherwise non-current.

## Licensing Terms - Perpetual License and Archives

*Licenses must be vetted to consider what rights to the data the library will require in the future as well as in the near term, such as the right to preserve the acquired data in a local repository and geospatial data infrastructure. The Agreement should provide for how the Licensed Materials are to be handled on termination or expiration of the agreement, in particular, whether locally hosted data may be retained or must be destroyed. Libraries should negotiate for perpetual access and/or the right to make archival copies of the material to support long-term research, peer review, and scholarly sharing. Clauses regarding Perpetual License and Archives should be incorporated into “Licensor Performance Obligations” or other sections of the agreement as appropriate.* [[4]](#footnote-4)

## Licensing Terms – Description, Delivery, and Format Specifications

*Any agreement for the purchase or licenses of data to be locally hosted should specify the data's content, and how, when and in what format the data will be delivered. CRL recommends the inclusion of a Schedule or multiple Schedules to the agreement specifying:*

a. Description of the Licensed Materials subject to the Agreement.
*Describe licensed materials with reasonable specificity.*

b. Fees and delivery terms covered by the Agreement.
*Include total fees, payment schedule, delivery schedule, frequency of data delivery, protocols for communicating receipt and acceptance of delivery.*

c. Specification of format types to be delivered and specific hardware or software configuration requirements.
*Examples of acceptable format types:
GIS data standard and acceptable format types: KML (Keyhole), GML, Shapefile, AutoCAD, MrSID, JPEG 2000, ECW, TIF.
Financial data standard and acceptable data format types: PDFs (reports), tab delimited files, Excel, TXT.
For archival copies, promote use of formats that will be stable and open over time, for example: TIFF, SHP, XML. Refer to Library of Congress’* [*Sustainability of Digital Formats*](http://www.loc.gov/preservation/digital/formats/)*.*

1. Examples of data derivatives are included in the article “Where are the Lightning Hotspots on Earth?” (Albrecht, *American Meteorological Society*, November, 2016). The article includes imagery derived from LIS data (lightning imaging sensor) provided by the NASA EOSDIS Global Hydrology Resource Center Data Analysis and Archive Center (DAAC). Using LIS data from DAAC, the Authors of this paper identified Earth’s lightning hotspots and other regional features revealed by detailed high-resolution satellite-derived climatology; the data used is described by the authors on pg. 2052-53; one derivative example can be seen on Pg. 2059, Fig. 3. <https://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-14-00193.1> [↑](#footnote-ref-1)
2. See LIBLICENSE Model License Section 3. for additional terms regarding Authorized Uses. [↑](#footnote-ref-2)
3. See LIBLICENSE Model License Section 5.1. for additional terms regarding Licensor Performance Obligations. [↑](#footnote-ref-3)
4. See LIBLICENSE Model License Section 8. for terms regarding Perpetual License and Archives. [↑](#footnote-ref-4)