

Texas A&M University Libraries' Purchased Data Value Statement

For purposes of this value statement and the management of collections and services in the Texas A&M University Libraries, the following definitions will be used:

- Data are "facts, figures, or instructions represented in a form that can be comprehended, interpreted by a human being or processed by a computer".
- A dataset is "a logically meaningful collection of similar or related data." Data are often numerical, although they can also be in other forms and are intended to be analyzed to generate new knowledge.
- Big data are "large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, and video, click streams, and/or all other digital sources available today and in the future."

Note: Random, unrelated sound bites would not be considered a dataset, but a collection of related or similar sound bites (such as samples of bird songs) that can be analyzed and interpreted would.

Value Statement Scope

While big data poses unique challenges, this data value statement should be applied to the purchase and license of all datasets that have a demonstrated or justifiable value to research and instruction at Texas A&M University. In some instances, datasets may be purchased as part of a mixed collection. For example, data may be included in online handbooks, exemplified by Knovel, or in recognized disciplinary databases such as Datastream, and the social science collection ICPSR. In these cases, the TAMU purchased data value statement may be superseded by the value statement for the format under which the resources acquired, or which represents the primary material type.

Discovery and Access

1. Data must be locally hosted or available both on and off-campus via IP or username/password authentication.
2. Vendor must allow multiple simultaneous users to access the product.
3. Data must contain sufficient descriptive information, either through the resource itself or through documentation, to enable the library to provide bibliographic records that would facilitate discovery through the catalog and Primo.
4. Data must be accessible and findable through discovery tools, such as the library's catalog, Primo, or other library implemented search mechanism.
5. If special, non-commercially available software is required for use, the vendor must supply this software along with documentation.
6. The interface must be user-friendly.
7. Vendor must provide documentation, or a guide to the dataset, which might include: logical labels; field names and definitions/explanations; list of abbreviations and what they mean (if applicable), etc.
8. Vendor must provide notifications when content is deleted, added, or updated.
9. Purchased/licensed data, including software used to access or manipulate the data, must be ADA compliant in accordance with federal and state law.
10. Open access is highly valued

Research & Instruction

1. License must accommodate fair use.
2. Vendor must allow users to download and export all or portions of datasets.
3. Scholar must be allowed to combine data or datasets that can be analyzed and used to generate new knowledge.
4. Scholars retain the rights to data derivatives.
5. Vendor will allow sharing among research teams, even if not all participants are A&M affiliated.

Sustainable & Fair Use Models

1. Vendor must provide perpetual access. If the company intends to withdraw access or goes out of business, the vendor will give the Texas A&M University Libraries the option to retrieve and store datasets and any necessary software.
2. In the interest of preservation, purchased data sets must be available in vendor neutral platforms, such as csv or tab delimited files in place of MS Excel workbooks.
3. If vendor specific software is used for access to data, the vendor must make the most current working version of the software available to the library to preserve access.
4. If the software is no longer supported by the vendor, the vendor must assist in the migration of data into a vendor neutral format that will preserve access.
5. If the vendor hosts the data and provides the interface, the vendor must make usage statistics available to the library.