



# Rancho BioSciences Data Curation Services

Rancho BioSciences is a Scientific Data Consulting Company specializing in curation of clinical, preclinical, OMICs, NGS data as well as vendor catalogs, data sheets and scientific papers. We can work with all data sources including public repositories (TCGA, CCLE, GEO, Allen Brain Atlas, ADNI and more), internal data sources and third party repositories.

## Process

Rancho BioSciences has PhD and MD scientists available around the world for onsite or off-site data curation services. The projects our scientists work on range from a week to several years.

We take time to listen to our customers’ unique requirements and to understand their projects. Often, we will meet with customers to define the project scope and then submit an initial proposal for our customers to review. At the onset of project we schedule a kick off meeting to discuss how everyone wants the project to work and how communications will be handled. We work with a point person that sends the data to Rancho for data curation. This can be sent by email, in spreadsheets, thumb drives or a secure drop box. We then schedule regular conference calls and send update reports weekly until the project is complete and then we have a final on line meeting to wrap up the project, make sure everyone has what they need and understand what to do next.

## Data

Our scientists can work with almost all data types. Excel files, pdf documents, Word documents, SAS file, SPSS file can all be used as input data files. Once curated, the data can also be delivered in the format desired by customer. Specific data formats for various platforms are generated so that customers can load data directly into tranSMART, Array Server, IDBS, Oracle’s platforms or a custom platform.



## Data Curation Services

We really emphasize a manual, personalized approach to data curation with attention to detail. Some of the components for our data curation services are listed below:

**Data stewardship:** We coordinate the timely release of data to the curators.

**Creation of ontologies:** ontologies are created to reflect the study experimental design and can be compatible with platforms such as transSMART, OmicSoft and Oracle.

**Data curation:** Processing and formatting the data to platform specifications, cleanup of the data (for example, resolution of bogus data entries, data QC, correction of misspellings and so forth). Wherever possible, we apply CDISC, MeSH and MedDRA terms to the study and will retain mapping files to the original data.

**Data loading:** Rancho Biosciences can assist customers with data loading. We have a team of software engineers who can load the data themselves or assist customers with their ETL needs.

**Data quality control:** We inspect new datasets and check for data integrity and for performance of all workflows. We will solicit input from the scientists and/or project managers on the loaded studies.

**Ontology building:** Often we will create an ontology for the data to be curated. The ontology has to be convenient for researchers and lend itself to the intended analyses. We are also proficient in the applications of existing ontologies to data, both internal and public.

**Semantic alignment and data curation:** Inspection of the data variables from each study and semantic alignment for those terms that are overlapping. Data curation involves, for example, unifying and cleaning gene IDs, assigning universal IDs to genes/proteins/SNPs, correction of misspellings in categorical data, classification of the categorical data, as well as application of standard terminology (such as MeSH, MedDRA, CDISC, etc) to applicable terms and checking for data to be in correct range.

**Data formatting:** Curated and cleaned data is then exported in a format that is compatible with the database that will be selected to host the data.

**Supporting documentation for the project:** data curation rules, data acquisition guidelines, and a project summary are delivered to customer to ensure the reproducibility of the results.

## Questions?

Call Julie Bryant, **760-717-7881**

[Julie@RanchoBioSciences.com](mailto:Julie@RanchoBioSciences.com)

[www.RanchoBioSciences.com](http://www.RanchoBioSciences.com)

## Experience

Rancho BioSciences is working with most major Pharmaceutical companies and well known institutes globally. Here is a list of current Rancho projects:

- Manually curating TCGA, GEO, CCLE and ADNI data sets to add value by providing supplementary information and complementary data types with ontology and vocabularies.
- Manually curating internal sensitive data sets on site: Rancho BioSciences provides curation services for sensitive clinical and OMICs data on site. When data transmission is not an option we will send our team member to assist you with the project on site.
- Accelerated Cure Project (ACP) for Multiple Sclerosis ([www.acceleratedcure.org](http://www.acceleratedcure.org)): Rancho curated data and cleaned-up and recorded the rules by which curation was performed. The process becomes a transparent and reproducible workflow that can be repeated as needed. For this specific project, the curated dataset (complete with OMICs and clinical data) will be loaded into transSMART.
- Large research hospital data harmonization project: Researchers and MD's at a major research hospital have been conducting pioneering research that demonstrated a crucial link between the immune system and cancer development. Data from these experiments as well as other available data needs to be aggregated into one knowledgebase, allowing researchers and analysts to more easily find and mine it. With the laboratory and biomarker data curated and loaded in transSMART.