

CU Data--Background Information

The CU Enterprise System runs ~2 PetaBytes of storage across 3 sites. This is comparable to operating over 11,000 laptops with full hard drives. This only relates to the Enterprise Business System for the University of Colorado; adding each campus' data to this requirement raises the amount of data impacted, by an order of magnitude.

The data impacted by this bill is stored in multiple different formats, very few of which are as simple as an excel spreadsheet or PDF file. Most data is kept in databases, accessed by the specific business system tasked with tracking it or in proprietary flat file formats, not accessible by standardized commercial means. This means we have to convert the data to an accessible format before we can even give it to the requestor.

A "record" in reality means "a finished product generated by data retrieval and compilation of this data into a usable and meaningful format and presented to the consumer." Raw data inside the business system is not technically accessible by consumer level tools without prior conversion and concatenation. Turning it into a "record", is really the product of multiple sophisticated systems which assemble and present it.

Native access to this data by external entities is made much more complex by the need to a) assemble all the various data into a coherent unit necessary for retrieval, and b) the requirement to anonymize the data extracted to protect individual privacy. Direct access to current production systems would negatively impact reliability and data integrity, as the systems are designed both for data retrieval and manipulation, and are not designed to address data masking or the random load generated by external entity searches.

All of these competing requirements would require us to deploy a "mirror" system that can be accessed by the requestor and at the same time "scrub" or "mask" personal data to protect personally identifiable information from being compromised. That system must be isolated from the production systems in use, in order to assure the production systems reliability, and would require an additional copy of unique data be "scrubbed" and separate from extant data inside the system. Even then, additional labor will be required to present the data in a way that is understandable or useful in context to the external requesting entity.

A "record" as outlined in the amendment, assumes quite a simplistic data architecture, which is not reflected in current electronic data systems design or usage. Due to the scale of the data held and manipulated by organizations, the simplistic "record" referenced in the amendment is in actuality, the result of a sophisticated process used to extract the requested data from larger data sets utilized by enterprise business tools.