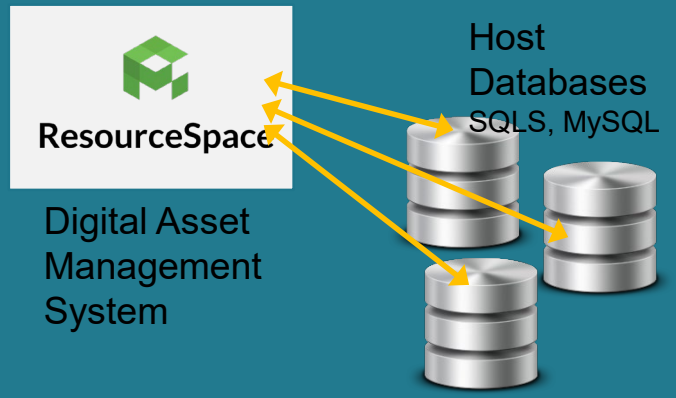
NM BUREAU OF GEOLOGY AND MINERAL RESOURCES

# Breaking Down Data Silos – Integrating Disparate Data Systems

**Problem**: Integrate multiple separate data systems together in a meaningful, robust, scalable manner.

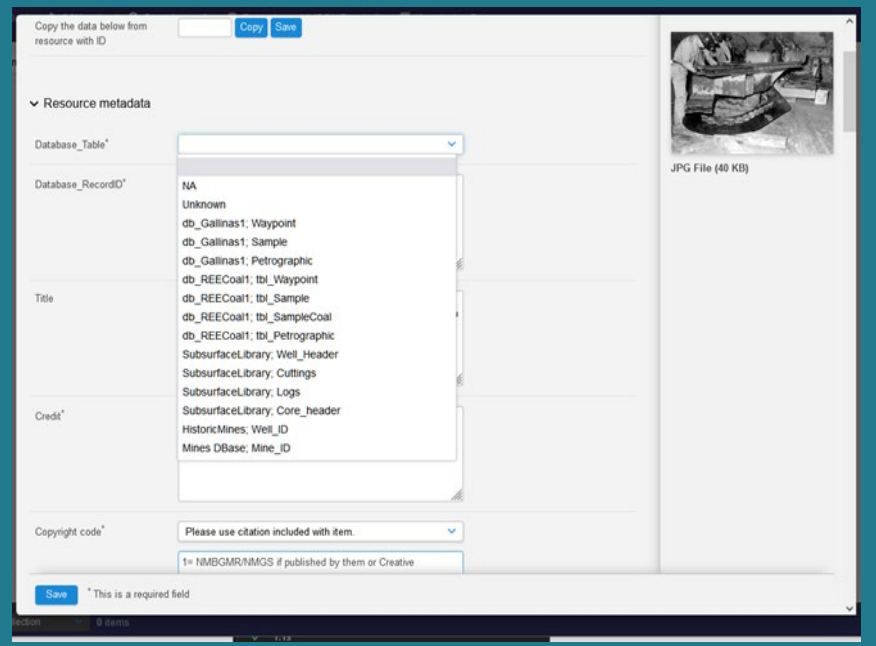



**Constraints**:

- Home grown "host" databases - multiple schemas and platforms that don't always mesh.
  - Too many existing apps and front ends to start from scratch.
  - No data standardization so merging databases is convoluted.
- Vendor systems
  eg., *Resource Space* - open source Digital Asset Management System (DAMS) used for Bureau's publicly available photograph and document archive (photoarchive.nmt.edu).
  - Have proprietary database structures and applications that make integration difficult.
  - Want to avoid changing database schema since future upgrades can break things and become a maintenance headache.
  - Use existing user-definable metadata fields – configurable through web-based UI; no DB changes needed.

**Our integration plan**:

- Host Database → DAMS: In each host DB we create a cross-walk or reference table with entries for various entities (waypoint, well, mine, sample, lab data, etc.) and their corresponding DAMS resource(s) [ allows 0 to many relationships ].
- DAMS → Host Database: Utilize user-defined metadata for fields indicating host database, table and record ID(s).
- Queries to the host DB can reference resources in the DAMS and likewise, data in the DAMS "knows" which database and table has the related host data.
- Integration designed specifically for connecting *Resource Space* (MySQL) to our Host DBs (SQLS, MySQL, Access) but will work and scale for other systems and databases.




Mark Leo-Russell
DBA, NMBGMR
Mark.Leo-Russell@nmt.edu

*A Research Division of New Mexico Institute of Mining and Technology*

New Mexico Tech

ResourceSpace

# Breaking Down Data Silos – Integrating Disparate Data Systems

**Lightning Talk: AAGS/USGS Data Preservation Workshop, Butte Montana**

August 31 2022

Mark Leo-Russell

New Mexico Bureau of Geology and Mineral Resources

Good morning. I'm Mark Leo-Russell, Database Administrator with the New Mexico Bureau of Geology and Mineral Resources at NM Tech.

My colleague Amy Trivitt spoke on Tuesday on how we are using Resource Space, an open source Digital Asset Management System, to manage and provide our historical photographs and other documents to Bureau staff and the general public [photoarchive.nmt.edu]. We are also beginning to use this system to store and manage photos and other materials for projects in progress including our DPP contracts, one of which involves core photos as several other presenters have spoken of this week.

An issue we realized last year after we went live was how do we "connect" assets in this system with the Bureau's many other host databases, mostly SQL Server and MySQL.

We wanted to avoid making changes to the Resource Space database since that potentially could be a maintenance headache when new versions of RS become available. Likewise, we have constraints on some of our in-house databases in that the data owners would strongly prefer that we not modify existing tables over concerns that changes would break applications, data loading processes, and reporting.

As we got more experience using Resource Space we began to see how the system's custom metadata functionality could be used to defined fields to connect resources to other DBs. We added a field with a drop-down list of our databases and tables as shown in the screen shot. That took care of the Resource Space to external DBs connection, but what about the other way?

To provide a link between records in our other databases, without modifying existing tables, I create a "cross-walk" table that contains records IDs such as Waypoints, Sample IDs, etc. to the corresponding resource or resources. The cross-walk table allows a 'zero-to-many' relationship between host DB entities to resources. For example, we can have a field site record (Waypoint) linked to multiple photographs via the reference table. Likewise for sample or core records—they can each have multiple photographs or other documents in the Photo Archive system.

This architecture allows queries against the Resource Space database to know where the corresponding host records reside (specific database and table) and likewise, to be able to know what digital assets belong to waypoints, samples, and analysis records.

My contact info is on the slide if you have any questions. Or see me later during our breaks.

Thanks for listening.