



High  
Performance  
Computing

# Intelligent Extraction and Indexing of Atmospheric Research in HPC Environment



A L C H E M I

T R A N S F O R M I N G   D A T A

Powered by

**Citadel**<sup>TEK</sup>

## Overview

Valuable atmospheric data has been collected over decades and stored for future research; however, gaining access to this information has been a challenge due to the technology used to store it, and the lack of simple user initiated mechanisms to search and find the specific information needed. ALCHEMI was designed to solve these types of problems, overlaying the existing environment, providing users the ability to search and access content regardless of where its located or on what technology its stored.

## Challenge

Over the past 40+ years, atmospheric research has been collected and stored in a tape based digital archive managed by a High Performance Storage System (HPSS). Each of these files includes header information describing the research contained in the file. While these files have followed official formatting standards, those standards have changed over time such that terms and acronyms have transformed as well. Currently there is no simple mechanism to search across this repository using the header information or to create a common term that incorporates all the variances in the standard in order to ensure complete results.

## Solution

The ALCHEMI solution from CitadelTek transforms how content is accessed, shared and analyzed, unlocking the value of data and making it useful well beyond its original intent. At the core of ALCHEMI is its Insight Engine, which automatically discovers, extracts and enhances data.

After connecting to the HPSS interface, ALCHEMI scans the file store, systematically opening files, extracting and indexing the header information and storing that information along with all of the associated file attributes in its Elasticsearch index.

ALCHEMI enables a customizable taxonomy to be created though a simple external file that allows the customer to construct custom terms based on key words, phrases, numbers or any other unique content their files might contain.

Users are now able to search across the atmospheric data store, using basic keywords or the custom terms to identify desired results. ALCHEMI includes automated and manual tagging processes, which allows search results to be further classified – including the ability to create virtual containers called *Collections*, which enhances the community's ability to share and collaborate on the research.

## Results

The valuable atmospheric research, which was locked away due to technology limitations, has now been freed. ALCHEMI provides the portal into this valuable information, making it available on all platforms users desire including their mobile devices as well as creating a collaborative environment for data sharing, version management and user interactive communication.