



Germany's Climate Archive, Rebuilt for the Exabyte Era

INDUSTRY

HPC CLIMATE MODELING

THE CHALLENGE

DKRZ operates one of the world's largest scientific data archives: 254 PB of climate research data growing by ~30 PB every year, across two sites, two library vendors, and five generations of LTO media. The legacy HSM couldn't keep up with the concurrency, the small-file workloads, or the operational overhead, and locked DKRZ into proprietary formats.

THE SOLUTION

DKRZ selected Versity ScoutAM to build a unified archive designed for a 1 EB, 500-million-file target namespace over a seven-year contract. ScoutAM gives the systems team 100% local administrative control and direct tape access, while its end-user interface gives researchers direct self-service data access with automatic metadata extraction, custom tagging, and instant search.

KEY RESULTS

01

Exabyte-scale archive designed for a 500-million-file namespace

02

Zero-data migration brought a 254 PB tape archive online in hours

03

Provided researchers with self-service search, automatic metadata extraction, and custom tagging

04

Non-proprietary, open-format architecture with 100% local admin autonomy

05

Automated cross-site replication between Hamburg and Garching for designated critical datasets



A NEW KIND OF ARCHIVE FOR A NEW KIND OF SCIENCE

The way researchers interact with archived climate data is changing. For most of DKRZ's history, the tape archive was a destination: a place to store finished work and retrieve it when needed. Increasingly, it's becoming a working resource.

Climate science is becoming a data-reuse discipline. Researchers need to revisit decades of simulation output to reproduce published results and support emerging AI-driven weather and climate workflows. At the same time, new models, higher-resolution simulations, and formats are changing the shape of the archive itself, producing data at higher volume, higher velocity, and far greater file counts than previous systems were designed to manage.

For DKRZ, the archive could no longer operate as a passive preservation tier. It needed to become an active scientific resource: searchable, policy-driven, open, and ready to support the next generation of climate research.

WHY THE ARCHIVE NEEDED TO CHANGE

DKRZ has hosted every generation of climate supercomputing in Germany, from a single-processor CDC 205 in the mid-1980s to today's Bull/Atos system. That continuity created one of DKRZ's greatest strengths (a deep, long-lived archive researchers return to for reproducibility, comparison, and new analysis) and, over time, a growing infrastructure challenge.

The tape archive alone now spans six aging Oracle StorageTek SL8500 libraries holding



ABOUT DEUTSCHES KLIMARECHENZENTRUM (DKRZ)

The German Climate Computing Center (DKRZ) is Germany's national supercomputing center for climate research. Founded in 1987 by Nobel laureate Klaus Hasselmann and headquartered in Hamburg, it operates as a non-profit and non-commercial limited company with four shareholders: the Max Planck Society, the City of Hamburg, the Alfred Wegener Institute, and Helmholtz-Zentrum Hereon.

DKRZ sits at the heart of Germany's climate research infrastructure, providing high-performance computing and data services to more than 5,000 climate and earth-system researchers. The vast majority of climate models, simulations, and data sets produced by the country's leading research institutions passes through its systems.

roughly 90–100 PB on five generations of LTO tape media; three Spectra Logic TFinity libraries with approximately 32,000 slots serving as the active production tier; a remote site in Garching for offsite replication; and a 130+ PB Lustre file system feeding 3,000 HPC compute nodes.

That heterogeneity had become a ceiling rather than a foundation. Peak days pushed **200 TB into the archive or 100 TB back out**, restore times were growing unpredictable, and new climate codes and formats were producing small files that a tape-based archive was never designed to handle. Data movement remained almost entirely user-driven, forcing researchers to reason about storage tiers instead of science.

Underneath the operational strain sat a deeper strategic concern. After more than a decade on HPSS followed by StrongLink, the systems team wanted a decisive break from proprietary tape formats (HPSS) and appliance-style deployments (StrongLink), with a documented exit strategy written into the next procurement from day one.

VERSITY SCOUTAM: AN OPEN BY DESIGN

DKRZ ran a formal public tender with a demanding list of technical requirements shaped by hard-won operational experience. Versity ScoutAM, delivered in partnership with NEC, cleared every mandatory requirement and gave DKRZ what previous systems couldn't:

- 100% local admin autonomy — no proprietary abstraction between the team and the system
- Direct tape access in an open format
- A flexible policy engine capable of modeling the full range of workflows Germany's climate community depends on
- Site licensing with no per-server or per-capacity surprises

The joint solution was engineered from day one to scale beyond 254 PB toward an eventual exabyte, with a 500-million-file namespace target to support the volume, velocity, and variety of modern climate workloads.

“Nobody wants to think about how they'll leave a project when they're just starting it. But you really have to think about it from the beginning, and it needs to be written down — even if you are absolutely happy with the project. **The openness of the Versity system is what makes that possible.**”

Carsten Schmitt

Head of Systems Department at DKRZ



254 PB ZERO DATA MIGRATION

One of the most important design decisions in the DKRZ deployment was avoiding a full tape rewrite.

Rewriting 254 PB across generations would have been a multi-year, high-risk undertaking with significant operational risk. **Instead, ScoutAM ingests the existing metadata and treats the legacy archive as a read-only tier.** New writes land in the open Versity format on current-generation media. Legacy data remains fully accessible in place, transparent to users.

"Users should not really be forced to think about where should I put my data in the archives. They should just get it."

Carsten Schmitt

Head of Systems Department at DKRZ

"The process itself was surprisingly fast. I was worried at first — I thought the conversion would take a long, long time," says Carsten Schmitt, Head of Systems Department at DKRZ. "But with support from NEC and Versity, we get the JSON output and it just runs a few hours. The import we tested so far also takes just a few hours."

The migration also surfaced value the team hadn't anticipated: during the metadata import, DKRZ was able to **identify and repair inconsistencies** invisible in day-to-day operation, cleaning up the archive's logical state as part of the move.

SELF-SERVICE ACCESS FOR RESEARCHERS

Historically, locating archived data often required administrator involvement for large data sets or custom metadata queries against the catalog. That model doesn't scale into the AI era, and it isn't how researchers want to spend their time.

Versity changes that. **Scientists can now search, locate, and retrieve their data directly, with automatic metadata extraction and custom tagging built into the platform.** A researcher can find exactly what they need — whether it's the output of a specific simulation run, a slice of a curated climate dataset, or every file matching a set of custom tags — and pull it back on demand.

"We want to help users not to think about the locality of data, so that they just can do their work and not be specialists in tape, S3, or HPC storage to just access the data," Schmitt notes.

For DKRZ, the difference is efficiency: the same retrieval and query operations run far faster than they did under the previous system.



FROM PASSIVE ARCHIVE TO ACTIVE DATA PLATFORM

ScoutAM was engineered from the ground up for archives that need to grow into the exabyte range without architectural discontinuities. **Its robust policy engine handles large job volumes with predictable resource allocation, and it opens the door to job prioritization and node specialization as DKRZ's automation matures.** Combined with intelligent staging across the flash cache and parallel tape operations, DKRZ can now deliver the predictable, reliable retrieval times its researchers have been asking for.

The design also addresses a growing headache in climate research: small files. Modern climate codes and data formats — Zarr chief among them — increasingly favor huge numbers of small files, a pattern that is notoriously inefficient on tape. Containerization was a hard requirement in DKRZ's procurement, and **ScoutAM's policy engine addresses it directly by automatically packing small files into larger objects on tape.**

Resilience is architectural, not bolted on. ScoutAM **asynchronously replicates** metadata, cache data, and archival data between DKRZ's Hamburg and Garching sites, providing the center with a up-to-date mirror for very important data sets at the secondary location. Critically, the metadata required to reconstruct the archive **resides on the tapes** themselves, in open format. and can be dumped and read without specialized tools. Versity also provides DKRZ a **free, read-only license in perpetuity**. DKRZ can recover its data, or read its own tapes with entirely different software, **with no dependency on proprietary tooling**.

READY FOR THE NEXT CHAPTER OF CLIMATE SCIENCE

DKRZ's mission has always been to reliably unlock the potential of accelerating technological progress for climate research.

With ScoutAM at the center of its archive architecture, DKRZ is positioned to support that mission at a far greater scale. The new archive preserves access to four decades of climate data while creating a foundation for AI-driven analysis, higher-resolution modeling, automated data management, and long-term scientific reuse. Germany's national climate archive, four decades of irreplaceable scientific record, is now on a foundation designed to serve the next four.

"DKRZ's mission is to be a partner to climate science, and our vision is to reliably unlock the potential of accelerating technological progress for climate research. Delivering on that vision requires a data infrastructure we can trust for decades — which is why we chose Versity."

Prof. Dr. Thomas Ludwig
Managing Director at DKRZ

