

# Jurnal Internasional Kurasi Digital

Edisi 1, Volume 3 | 2008

## Menuju Teori Pelestarian Digital

Reagan Moore,

San Diego Supercomputer Center

Juni 2008

### Abstrak

Lingkungan pelestarian mengelola komunikasi dari masa lalu sambil berkomunikasi dengan masa depan. Informasi yang dihasilkan di masa lalu dikirim ke masa depan oleh lingkungan pelestarian saat ini. Bukti bahwa pelestarian lingkungan menjaga keaslian dan integritas saat melakukan komunikasi merupakan teori pelestarian digital. Kami memeriksa informasi representasi yang diperlukan tentang lingkungan pelestarian untuk teori pelestarian digital. Informasi representasi mencakup deskripsi kebijakan manajemen preservasi, proses preservasi, dan informasi negara yang diperlukan untuk memverifikasi perilaku kerja yang benar dari sistem. Kami mendemonstrasikan kisi data berbasis aturan yang dapat memverifikasi bahwa kebijakan sebelumnya menerapkan properti preservasi dengan benar,<sup>1</sup>

---

<sup>1</sup> Pekerjaan ini didukung oleh Administrasi Arsip dan Arsip Nasional di bawah perjanjian kerja sama NSF 0523307 melalui suplemen untuk SCI 0438741, "Cyberinfrastructure: From Vision to Reality" dan oleh National Science Foundation, ITR 0427196, "Constraint-based Knowledge Systems for Grids, Perpustakaan Digital, dan Arsip Persisten". Pandangan dan kesimpulan yang terkandung dalam dokumen ini adalah milik penulis dan tidak boleh diartikan sebagai mewakili kebijakan resmi, baik tersurat maupun tersirat, dari National Science Foundation, National Archives and Records Administration, atau pemerintah AS.

## pengantar

Konsep pelestarian dapat dicirikan sebagai komunikasi dengan masa depan. Kita tahu bahwa di masa depan akan digunakan teknologi baru yang lebih hemat biaya dan lebih canggih daripada teknologi saat ini. Komunikasi dengan masa depan kemudian berhubungan dengan pemindahan catatan ke pilihan teknologi baru. Lingkungan pelestarian perlu menggabungkan jenis sistem penyimpanan baru, protokol baru untuk mengakses data, format pengkodean data baru, dan standar baru untuk mengkarakterisasi asal. Jadi, tantangan utama yang dihadapi pelestarian adalah bagaimana menerapkan teknologi baru secara efektif, sambil melestarikan properti pengawetan seperti keaslian, integritas, dan lacak balak. Teknologi jaringan data memberikan kemampuan yang diperlukan untuk menggabungkan teknologi baru, tanpa mengorbankan properti lingkungan pelestarian,

[2006](#)). Kami akan mengeksplorasi konsep ini di bawah "Kemandirian Infrastruktur".

Kami juga dapat melihat pelestarian sebagai validasi komunikasi dari masa lalu. Seorang arsiparis membuat pernyataan tentang penerapan proses pengawetan sebelumnya dan kebijakan pengawetan pengawetannya. Klaim apa pun tentang status keaslian dan integritas saat ini bergantung pada deskripsi lengkap tentang tindakan sebelumnya. Ini adalah tantangan besar kedua untuk lingkungan pelestarian, kemampuan untuk mengkarakterisasi bagaimana proses pengawetan sebelumnya telah dikendalikan oleh kebijakan pengelolaan pelestarian. Data grid berorientasi aturan mendukung karakterisasi eksplisit dari kebijakan dan proses (Rajasekar, Wan, Moore & Schroeder, [2006](#)). Kami akan mengeksplorasi konsep ini di bawah "Sistem Data Berorientasi Aturan Terpadu".

Dengan adanya karakterisasi proses pengawetan dan kebijakan pengelolaan, maka properti lingkungan pengawetan dapat diverifikasi. Secara khusus, adalah mungkin untuk memverifikasi kepercayaan seperti yang didefinisikan oleh Audit & Sertifikasi Repositori Terpercaya RLG / NARA: Kriteria dan Daftar Periksa (TRAC) ([2007](#)). Ini adalah tantangan besar ketiga untuk pelestarian lingkungan, kemampuan untuk memverifikasi bahwa mereka bekerja dengan benar. Kami mengeksplorasi persyaratan ini di bawah "Kriteria Penilaian".

Teori pelestarian digital kemudian menjadi mungkin. Ini didasarkan pada definisi sekumpulan minimal proses pengawetan yang diperlukan untuk mengimplementasikan kebijakan pengelolaan, dan sekumpulan minimal metadata pengawetan (informasi status tetap) yang diperlukan untuk memvalidasi kriteria penilaian. Tujuannya adalah untuk menunjukkan bahwa lingkungan pelestarian mempertahankan properti yang diinginkan seiring perkembangan teknologi yang mendasarinya. Ketika kami memeriksa komponen-komponen yang diperlukan untuk membangun teori pelestarian digital, kami mencatat bahwa diperlukan informasi representasi tentang lingkungan pelestarian itu sendiri. Lingkungan pengawetan tidak hanya membutuhkan kemampuan untuk melindungi catatan dari perubahan yang terjadi pada infrastruktur pengawetan, tetapi juga kemampuan untuk menandai perubahan yang terjadi dalam kebijakan pengawetan.

## Kemandirian Infrastruktur

Lingkungan pelestarian adalah middleware perangkat lunak yang melindungi catatan dari evolusi teknologi yang cepat. Lingkungan pengawetan menyediakan operasi standar pada ruang nama tetap yang dapat digunakan untuk mengimplementasikan proses pengawetan. Operasi standar diporting ke setiap versi baru dari infrastruktur. Jadi dari perspektif arsiparis, nama-nama yang digunakan untuk mengidentifikasi arsip, arsiparis dan sistem penyimpanan tidak boleh berubah, dan operasi yang dilakukan pada arsip harus tetap konsisten dari waktu ke waktu. Lingkungan pelestarian secara efektif melindungi catatan dari ketergantungan apa pun pada pilihan infrastruktur saat ini.

Data grid adalah middleware perangkat lunak yang memungkinkan pembentukan koleksi bersama dari data terdistribusi. Data mungkin berada di berbagai jenis sistem penyimpanan, yang terletak di berbagai institusi yang terhubung dengan Internet. Kisi data memelihara informasi yang diperlukan untuk mengidentifikasi di mana file sebenarnya disimpan, mengelola metadata deskriptif tentang setiap file, mengelola replika file, dan mengelola komunikasi melalui jaringan area luas. Ruang nama tetap dipertahankan yang mencakup nama yang digunakan untuk mengidentifikasi arsiparis, nama yang digunakan untuk mengidentifikasi catatan, nama yang digunakan untuk mengidentifikasi atribut deskriptif, dan nama yang digunakan untuk mengidentifikasi sumber daya penyimpanan. Data grid mengimplementasikan virtualisasi kepercayaan melalui otentikasi setiap pengguna dan manajemen kontrol akses eksplisit pada file, metadata, dan sistem penyimpanan. Jaringan data juga menyediakan serangkaian operasi standar untuk mengakses file jarak jauh. Akibatnya, kisi data mengelola semua properti catatan secara independen dari pilihan penyimpanan dan teknologi basis data. Kemampuan ini menjadi dasar konsep kemandirian infrastruktur.

Operasi standar yang disediakan oleh jaringan data untuk memanipulasi data telah dirancang untuk mendukung proses yang diperlukan untuk berbagi data, publikasi data, dan penyimpanan data. Operasi tersebut mencakup manipulasi level byte (buka, tutup, baca, tulis, seek, stat, dll.), Manipulasi level file (replikasi, buat versi, backup, agregat dalam container, parse metadata), dan manipulasi level koleksi (rekursif operasi pada direktori, akses massal, pergerakan data massal). Saat ini, jaringan data Storage Resource Broker (SRB) menyediakan lebih dari 80 operasi untuk mendukung akses dan manipulasi data jarak jauh<sup>2</sup>.

Kami mengamati bahwa protokol penyimpanan yang didukung oleh jenis sistem penyimpanan tertentu berbeda baik dalam semantik (maksud operasi) atau manipulasi (hasil berbeda yang dihasilkan) di seluruh produk vendor. Teknologi SRB mengimplementasikan operasi standar yang tidak bergantung pada sistem penyimpanan dengan membuat driver khusus sistem penyimpanan untuk memetakan dari operasi standar ke protokol sistem penyimpanan masing-masing vendor. Kisi data Storage Resource Broker adalah contoh sistem yang memberikan kemandirian infrastruktur untuk lingkungan pelestarian.

Jaringan data SRB digunakan untuk produksi di seluruh dunia, dalam cakupan yang bervariasi dari lingkungan lokal yang mengelola data pada disk lokal dan arsip pita, hingga jaringan data internasional yang memindahkan ratusan terabyte data antara Amerika Utara dan Eropa. Di San Diego Supercomputer Center, jaringan data SRB mengelola satu Petabyte data

<sup>2</sup>Kisi data Storage Resource Broker <http://www.sdsc.edu/srb/index.php/>

terdiri dari lebih dari 200 juta file. Kisi data SRB produksi mendukung perpustakaan digital, lingkungan pelestarian, aliran data sensor waktu nyata, dan platform analisis data.

Kemampuan jaringan data untuk menggabungkan teknologi baru yang transparan menjadi dasar pemilihannya sebagai infrastruktur di mana lingkungan pelestarian dapat dibangun. Pada saat teknologi penyimpanan baru tersedia, driver sumber daya penyimpanan jaringan data baru dapat diterapkan yang memetakan operasi standar ke protokol akses penyimpanan baru. Rekaman yang telah disimpan pada teknologi lama dapat dimigrasikan ke sistem penyimpanan baru tanpa mengubah nama catatan, tanpa kontrol akses pada catatan yang berubah, dan tanpa hubungan antara metadata deskriptif dan catatan yang dipecahkan. Kisi data secara transparan mengelola semua metadata administratif yang diperlukan untuk memastikan bahwa keaslian dan lacak balak dipertahankan sementara teknologi baru dibawa ke lingkungan pelestarian.

Kisi data juga mengelola integritas data. Checksum dipertahankan untuk setiap rekaman, bersama dengan beberapa replika. Perintah administratif dapat dikeluarkan untuk memverifikasi integritas (kebenaran checksum), memverifikasi keberadaan replika (sinkronisasi), dan memverifikasi bahwa untuk setiap arsip yang diarsipkan ada metadata deskriptif yang sesuai. Berdasarkan kemampuan ini, berbagai lingkungan pengawetan telah dibuat. Contoh penting termasuk NARA Transcontinental Persistent Archive Prototype (TPAP)<sup>3</sup>, arsip persisten Perpustakaan Digital Sains Nasional<sup>4</sup>, dan Repositori Pelestarian Digital Perpustakaan Digital California<sup>5</sup>. Mengingat keberhasilan ini, pertanyaannya adalah apa lagi yang diperlukan agar lingkungan pelestarian menjadi layak?

Kami mengamati bahwa tenaga kerja yang diperlukan untuk mengelola koleksi berukuran 100 terabyte yang memiliki puluhan juta file dapat menjadi berat. Tidaklah cukup untuk menyediakan fungsi administratif (proses pemeliharaan) yang dapat dijalankan untuk memulihkan kesalahan yang terdeteksi. Alih-alih, jaringan data harus mengotomatiskan sebanyak mungkin fungsi administratif, termasuk verifikasi integritas dan keaslian, agar sistem pengawetan yang besar dapat berjalan.

### Sistem Data Berorientasi Aturan Terintegrasi

Otomatisasi kebijakan manajemen dimungkinkan jika kebijakan dapat dinyatakan sebagai aturan yang diberlakukan pada setiap akses atau manipulasi data. Sistem Data Berorientasi Aturan terintegrasi iRODS menyediakan pemetaan dari kriteria penilaian, hingga kebijakan preservasi yang menegakkan kriteria penilaian, hingga proses preservasi yang memanipulasi catatan. Perangkat lunak middleware grid data digunakan untuk memetakan dari kebijakan manajemen yang diinginkan ke sistem penyimpanan dan database yang tersedia secara komersial. Berdasarkan pengalaman dengan data grid Storage Resource Broker (Baru, Moore, Rajasekar, & Wan, 1998), kami menyadari bahwa operasi yang didukung oleh sistem penyimpanan jarak jauh tidak secara langsung berhubungan dengan kemampuan pengawetan. Kami menggunakan operasi standar yang didukung oleh jaringan data SRB untuk menentukan layanan mikro yang dijalankan langsung di lokasi penyimpanan jarak jauh. Dengan demikian, setiap kemampuan pengawetan adalah agregasi dari beberapa layanan mikro, dan setiap layanan mikro adalah agregasi dari beberapa operasi di sistem penyimpanan jarak jauh.

<sup>3</sup>Prototipe Arsip Persisten Lintas Benua NARA <http://www.sdsc.edu/NARA>

<sup>4</sup>Perpustakaan Digital Sains Nasional <http://nsdl.org/>

<sup>5</sup>Gudang Penyimpanan Digital Perpustakaan Digital California

<http://www.cdlib.org/inside/projects/preservation/dpr/>

<i>Kelestarian Lingkungan Hidup</i>	<i>Konservasi Properti</i>	<i>Kontrol Mekanisme</i>	<i>Kemampuan</i>
<b>Pengelolaan Fungsi</b>	Penilaian Kriteria	Kelestarian Kebijakan	Kelestarian Proses
<b>Jaringan data - Virtualisasi manajemen</b>			
<b>Manajemen data Infrastruktur</b>	Status Persisten	Aturan	Layanan mikro
<b>Data grid - Data dan virtualisasi kepercayaan</b>			
<b>Fisik Infrastruktur</b>	Database	Mesin Aturan	Sistem penyimpanan

Tabel 1. Karakterisasi kebijakan pengelolaan pelestarian sebagai aturan.

Seperti yang ditunjukkan pada Tabel 1, kami menyatakan proses pengawetan sebagai satu set layanan mikro. Kami mengekspresikan kebijakan manajemen sebagai kumpulan aturan yang mengontrol pelaksanaan setiap layanan mikro. Kami mengevaluasi kriteria penilaian melalui kueri pada informasi status persisten yang dihasilkan oleh penerapan aturan. Aturan disimpan di mesin aturan yang diinstal di setiap lokasi penyimpanan jarak jauh, dan informasi status yang terus-menerus disimpan dalam katalog metadata pusat.

Perhatikan bahwa meskipun layanan mikro tetap tidak berubah dari waktu ke waktu, kami masih memerlukan serangkaian operasi standar yang diterapkan pada sistem penyimpanan jarak jauh yang juga tetap tidak berubah. Dua tingkat virtualisasi data yang ditunjukkan pada Gambar 1 memungkinkan untuk membangun lingkungan pengawetan yang menjalankan layanan mikro standar (untuk mengekspresikan proses pengawetan) yang terdiri dari operasi standar (untuk berinteraksi dengan sistem penyimpanan). Dengan demikian, proses pengawetan yang dinyatakan sebagai kumpulan layanan mikro standar akan bekerja di sistem penyimpanan lama dan baru melalui operasi standar.

Layanan mikro iRODS dapat dianggap sebagai mendefinisikan sekumpulan minimal fungsi pengawetan yang perlu dijalankan tepat waktu. Demikian pula, aturan iRODS dapat dianggap sebagai mendefinisikan sekumpulan minimal kebijakan manajemen yang diperlukan untuk menegakkan kepercayaan. Dengan deskripsi minimal dan lengkap tentang komponen lingkungan pelestarian, seseorang kemudian dapat membuat sistem berbasis aturan yang konsisten secara internal. Operasi yang diperlukan menghasilkan informasi status persisten yang diperlukan untuk memvalidasi pernyataan tentang kepercayaan.

Selain kemampuan untuk mencirikan lingkungan pelestarian, persyaratan implementasi kedua adalah otomatisasi pelaksanaan kebijakan manajemen. Kami mengamati bahwa koleksi meningkat dalam ukuran, jumlah tenaga kerja yang dibutuhkan untuk pulih dari masalah yang terjadi di lingkungan terdistribusi menjadi memberatkan (Moore, Wan, & Rajasekar, 2005). Sistem iRODS mengotomatiskan penerapan

aturan sebagai operasi atomik, dijalankan pada setiap permintaan proses preservasi, atau sebagai operasi yang ditangguhkan, dijalankan saat sumber daya tersedia, atau sebagai operasi berkala. Dukungan untuk operasi yang ditangguhkan diperlukan untuk menangani situasi di mana sistem penyimpanan jarak jauh tidak dapat diakses untuk sementara. Operasi berkala diperlukan untuk mengelola validasi integritas dan keaslian. Sebuah aturan dapat ditulis untuk memvalidasi checksum, menyinkronkan replika, dan mengoreksi salinan yang telah rusak. Demikian pula, aturan dapat ditulis yang membandingkan metadata asal dengan metadata yang diperlukan untuk rangkaian rekaman, dan kemudian mengidentifikasi informasi keaslian yang hilang, atau mengekstrak informasi yang diperlukan dari perjanjian pengiriman.

Masalah terkait saat menerapkan lingkungan preservasi yang dapat diskalakan adalah pengelolaan sejumlah besar catatan. Ini diatasi dengan membuat operasi yang bertindak pada kumpulan catatan. Dalam praktiknya, rangkaian operasi SRB yang dilakukan di setiap sistem penyimpanan atau database bergantung pada tingkat agregasi yang diberlakukan pada setiap ruang nama logis yang digunakan untuk mengidentifikasi catatan, pengarsip, dan sistem penyimpanan:

- Operasi pada ruang nama logis pengguna {pengguna tunggal, grup pengguna, federasi grid data}
- Operasi pada ruang nama logis sumber daya penyimpanan {sistem penyimpanan tunggal, sistem gabungan dengan cache data, cluster}
- Operasi pada ruang nama logis file {file tunggal, kontainer yang menggabungkan beberapa file, direktori hierarki file}
- Operasi pada metadata yang ditentukan pengguna {atribut tunggal, tabel hierarki, koleksi}

Ini beberapa tingkat agregasi diperlukan untuk skalabilitas pelestarian lingkungan. Serangkaian tingkat agregasi serupa sedang diterapkan dalam sistem iRODS melalui pendanaan NSF untuk pengembangan sistem manajemen data berbasis aturan umum.

Kriteria desain akhir untuk lingkungan pelestarian adalah dukungan untuk evolusi lingkungan pelestarian itu sendiri. Ini menyiratkan bahwa kebijakan dan proses manajemen preservasi dapat berkembang untuk menangani jenis catatan baru, jenis metadata asal baru, dan persyaratan hukum baru. Untuk memungkinkan aturan dan layanan mikro berkembang, kami mendukung tiga ruang nama logis tambahan di lingkungan iRODS yang mengidentifikasi layanan mikro, aturan, dan informasi status persisten. Kami juga menambahkan tingkat agregasi yang serupa dalam setiap ruang nama:

- Operasi pada ruang nama logis layanan mikro {atomic, deferred, periodic}
- Operasi pada aturan ruang nama logis {layanan mikro tunggal, kumpulan layanan mikro, hierarki aturan rekursif}
- Operasi pada ruang nama logis informasi status persisten {atribut tunggal, jejak audit pada peristiwa, penguraian jejak audit untuk kepatuhan dengan kriteria penilaian}

Ini tiga ruang nama memungkinkan virtualisasi manajemen. Seseorang dapat menambahkan yang baru kebijakan manajemen, kemampuan penyimpanan baru, dan informasi status persisten baru tanpa memengaruhi kemampuan untuk menjalankan kebijakan dan proses manajemen sebelumnya. Dengan tiga ruang nama logis tambahan ini, lingkungan pelestarian juga dapat berkembang. Juga, kemampuan untuk mengkarakterisasi kebijakan manajemen sebagai aturan menyiratkan bahwa kebijakan manajemen dapat diotomatiskan, meminimalkan kebutuhan tenaga kerja yang diperlukan untuk koleksi besar.

Dengan menetapkan aturan yang mengotomatiskan validasi kriteria penilaian, lingkungan pelestarian dapat ditentukan yang memvalidasi kepercayaannya sendiri. Ini juga menjadi mungkin untuk memigrasi catatan, aturan, layanan mikro, dan informasi status persisten ke lingkungan pelestarian independen lainnya. Kebijakan pengelolaan yang diterapkan pada lingkungan pelestarian pertama dapat terus diterapkan pada lingkungan pelestarian kedua. Ini berarti bahwa pernyataan tentang keaslian dan integritas dapat terus divalidasi saat rekaman dipindahkan di antara penerapan lingkungan pelestarian yang berbeda.

## Kriteria penilaian

Komunitas pelestarian sedang mengembangkan kriteria penilaian untuk memvalidasi kepercayaan dari repositori digital. Sebuah sistem yang mampu memvalidasi kriteria penilaian dapat dianggap dapat dipercaya, dan dengan demikian akan menjadi lingkungan yang wajar untuk penyimpanan data untuk jangka panjang. Serangkaian kriteria penilaian awal telah diusulkan oleh Research Libraries Group dan National Archives and Records Administration (RLG & NARA, 2007). Analisis kriteria penilaian dan pemetaan kriteria kebijakan manajemen telah diterbitkan (Moore & Smith,

2006; Smith & Moore, 2006). Harapannya adalah seseorang dapat menentukan kebijakan manajemen yang memastikan kepercayaan, menetapkan aturan yang menerapkan kebijakan, menentukan kemampuan yang mengimplementasikan fungsi preservasi yang diperlukan, dan mendefinisikan metadata preservasi yang menangkap informasi tentang penerapan fungsi preservasi. Seseorang kemudian dapat menanyakan metadata pelestarian untuk menilai apakah kriteria penilaian telah dipenuhi. Pendekatan berdasarkan kriteria penilaian mencoba untuk menentukan semua kebijakan pengelolaan yang diperlukan untuk membuktikan bahwa lingkungan pelestarian akan berhasil melestarikan catatan.

Kriteria penilaian didasarkan pada prinsip pelestarian tradisional:

- Keaslian: pernyataan tentang asal-usul catatan
- Respect des fonds: pernyataan tentang organisasi asli catatan
- Chain of custody: pernyataan tentang kepemilikan catatan
- Integritas: pernyataan tentang pengelolaan catatan

Masing-masing prinsip pengawetan ini mendefinisikan properti yang harus dijaga oleh lingkungan pengawetan. Minimal, lingkungan pengawetan memerlukan nama yang tetap untuk mengidentifikasi catatan, pengarsip, dan tempat penyimpanan. Pernyataan tentang manajemen catatan kemudian dapat didasarkan pada atribut yang terkait dengan ruang nama yang tetap. Contohnya termasuk ruang nama yang digunakan oleh lingkungan pengawetan untuk melacak asal (metadata deskriptif), integritas (aturan, proses pengawetan, informasi status sistem), lacak balak (nama arsiparis, nama sumber daya penyimpanan), dan respek des fonds (catatan nama dan metadata deskriptif). Ruang nama ini harus tetap tidak berubah dari waktu ke waktu, memastikan bahwa operasi yang dilakukan oleh arsiparis sebelumnya dapat diinterpretasikan dengan benar. Kriteria penilaian juga mensyaratkan agar fungsi yang dijalankan oleh proses pengawetan tetap konsisten dari waktu ke waktu. Ini berarti lingkungan pengawetan harus diukur dalam hal operasi pengawetan aktual yang dilakukan dan kebijakan manajemen yang mengontrol pelaksanaan proses pengawetan. Dengan demikian, kami memiliki dua konteks terpisah yang membutuhkan deskripsi; properti catatan, dan kebijakan serta proses lingkungan pelestarian.

Sistem data yang berorientasi pada aturan menentukan sekumpulan minimal proses pelestarian dan kebijakan manajemen yang menjadi dasar lingkungan pelestarian. Sistem berorientasi aturan juga melacak perubahan pada lingkungan pelestarian. Perlu kita ketahui bagaimana proses pengawetan yang kita terapkan saat ini terkait dengan proses pengawetan yang diterapkan di masa lalu, agar dapat menegaskan integritas dan keaslian.

Karakterisasi seperangkat minimal kebijakan pengelolaan pelestarian sulit dilakukan. Kita perlu memetakan dari kriteria penilaian, ke kebijakan manajemen yang menegaskan asersi pelestarian, hingga kemampuan pelestarian yang menerapkan proses pengawetan. Prototipe sistem manajemen data iRODS dirancang untuk

menerapkan lingkungan pelestarian lengkap yang terbukti. Upaya pembangunan saat ini meliputi:

- Penerapan kapabilitas Arsip Elektronik sebagai aturan iRODS, layanan mikro, dan informasi negara (Arsip Nasional dan Administrasi Arsip [NARA], 2003 ). Tantangan utama adalah desain tingkat agregasi yang benar dari operasi penyimpanan jarak jauh ke dalam layanan mikro. Layanan mikro harus cukup sederhana sehingga semua kemampuan pelestarian dapat diimplementasikan dari layanan mikro standar. Tetapi jika mereka berada pada level yang terlalu rendah (operasi berorientasi byte) mereka menjadi sulit untuk diterapkan. Penilaian saat ini telah menetapkan 174 aturan dan sejumlah kecil layanan mikro yang perlu dilaksanakan. Lingkungan iRODS saat ini menyediakan 73 layanan mikro. Anggapannya adalah bahwa jika tingkat agregasi yang benar dari operasi standar yang dilakukan pada sistem penyimpanan jarak jauh ditentukan untuk setiap layanan mikro, sejumlah minimal layanan mikro akan dibuat, yang mampu mengekspresikan setiap kemampuan pemeliharaan yang diperlukan.
- Penerapan kriteria penilaian RLG / NARA untuk repositori digital terpercaya. Kami telah menetapkan 105 kebijakan manajemen yang diperlukan untuk mengekspresikan kriteria penilaian dalam hal aturan iRODS dan layanan mikro.
- Perbandingan pelestarian pendekatan berbasis metadata dengan pendekatan berbasis aturan. Seseorang dapat menentukan metadata pelestarian yang diperlukan, dan kemudian mengidentifikasi seperangkat aturan pelestarian yang diperlukan untuk mengelola metadata. Kami telah mengambil pendekatan sebaliknya. Kami menentukan kebijakan manajemen preservasi yang diperlukan, dan kemudian mengidentifikasi metadata preservasi yang dihasilkan dari penerapan kebijakan tersebut. Pendekatan ini memberikan seperangkat metadata pengawetan yang lebih komprehensif tidak hanya tentang catatan, tetapi juga lingkungan pengawetan.

Sistem data berorientasi aturan iRODS tersedia sebagai perangkat lunak sumber terbuka dari wiki <sup>6</sup>. Rilis produksi pertama perangkat lunak dibuat pada tanggal 23 Januari 2008.

## Teori Pelestarian

Pelestarian dapat dianggap sebagai komunikasi dengan masa depan. Informasi yang dipahami saat ini dikirim ke sistem yang tidak diketahui di masa depan di mana ia akan diinterpretasikan dan ditampilkan. Sistem masa depan mungkin tidak hanya memiliki perangkat keras dan perangkat lunak yang berbeda, tetapi juga standar yang berbeda untuk informasi pengkodean. Secara efektif, kita perlu mengirimkan ke masa depan tidak hanya informasi (catatan), tetapi juga deskripsi lingkungan yang digunakan untuk mengelola dan membaca catatan tersebut.

Sistem pengawetan di masa depan dapat dikaitkan dari waktu ke waktu dengan sistem pengawetan asli. Untuk mempertahankan kemampuan menafsirkan dan menampilkan rekaman, lingkungan pelestarian harus mencirikan evolusinya sendiri, dan dampak evolusi lingkungan pelestarian terhadap manajemen rekaman. Teori pelestarian membuat pernyataan tentang kemampuan untuk menjaga konteks informasi, pengaturan, dan pengelolaan catatan, serta konteks informasi.

---

<sup>6</sup>iRODS <http://irods.sdsc.edu/>



(kebijakan manajemen dan prosedur pelestarian) dari lingkungan pelestarian.

Contoh pengelolaan informasi kontekstual didefinisikan dalam standar OAIS (Consultative Committee for Space Data Systems [CCSDS], 2002). Ini berfokus pada kemampuan untuk mengakses dan menafsirkan catatan melalui pembuatan informasi representasi. Informasi representasi mendefinisikan struktur yang ada dalam record dan label semantiknya. Komunitas yang ditunjuk didefinisikan yang mempertahankan kemampuan untuk menafsirkan label semantik. Namun, standar OAIS tidak memberikan informasi representasi tentang lingkungan pelestarian itu sendiri.

Konsep kemandirian infrastruktur mengukur pernyataan tentang pelestarian lingkungan, termasuk tidak hanya ruang nama, tetapi juga proses pelestarian, dan kebijakan pelestarian. Dengan mendemonstrasikan bahwa lingkungan pengawetan mengontrol konteks informasi yang diperlukan untuk melestarikan kemampuan untuk menerapkan prosedur pengawetan, kita dapat membuat teori pengawetan di mana konten informasi dari catatan dan konteks informasi dari lingkungan pengawetan dikomunikasikan ke masa depan.

Karena tidak ada sistem yang dapat sepenuhnya mendeskripsikan dirinya sendiri, teori pelestarian perlu menentukan kumpulan asumsi minimal yang menjadi dasar lingkungan pelestarian, dan kemudian menunjukkan bagaimana asumsi ini dilestarikan seiring dengan berkembangnya lingkungan pelestarian. Untuk melakukan ini, kita memerlukan beberapa konsep lagi untuk teori pelestarian. Kita membutuhkan unit fundamental dari "fungsi" yang menjadi dasar proses pelestarian. Kandidat untuk unit fungsi pengawetan adalah kumpulan minimal kebijakan pengelolaan dan rangkaian layanan mikro minimal yang diperlukan untuk mengekspresikan dan menerapkan kemampuan pengawetan yang diperlukan. Kami ingin mengkarakterisasi dampak penerapan proses preservasi sebagai perubahan informasi negara, dan fungsi transformasi yang diterapkan pada rekaman. Untuk proses pengawetan yang layak, kami memerlukan transformasi yang dapat dibalik: kemampuan untuk mengubah kembali ke catatan asli. Ini melibatkan karakterisasi catatan sebagai berikut:

- Setiap record adalah urutan bit
- Isi informasi dijelaskan dengan mendefinisikan struktur yang ada dalam urutan bit, dan kemudian menamai strukturnya. Nama struktur mewakili istilah semantik yang digunakan untuk mendefinisikan arti catatan. Isi pengetahuan didefinisikan sebagai hubungan antara dan pada struktur. Contohnya termasuk:
  - Hubungan logis. Istilah semantik dapat dipetakan ke dalam ontologi, dan penalaran dilakukan pada atribut yang disimpulkan (kisi semantik).
  - Hubungan temporal. Struktur tersebut mungkin mewakili cap waktu yang dapat digunakan untuk memaksakan kausalitas.
  - Hubungan spasial. Struktur dapat dipetakan ke sistem koordinat yang dapat dipetakan secara bergiliran menjadi geometri dan ditampilkan dalam sistem GIS.
  - Hubungan prosedural. Struktur tersebut dapat mewakili hasil dari suatu proses dalam alur kerja seperti penerapan proses pelestarian.
  - Hubungan fungsional. Struktur tersebut mungkin memerlukan penerapan algoritme transformasi untuk membuat data turunan produk.

Ini karakterisasi catatan mendukung konsep objek persisten

(Moore, 2003 ). Dimungkinkan untuk membuat objek persisten yang dapat ditampilkan di masa depan menggunakan teknologi yang belum diketahui, meskipun struktur dan hubungan internal objek tersebut didasarkan pada teknologi saat ini. Untuk mengaktifkan tampilan objek persisten, kita memerlukan satu konsep lagi, yaitu manipulasi record di masa depan dapat diekspresikan dalam bentuk manipulasi struktur dan hubungan yang telah dijelaskan untuk setiap record. Kami mendasarkan kemampuan untuk menampilkan rekaman melalui tingkat tipuan berikut:

- Mencirikan struktur dan hubungan standar yang ada dalam catatan
- Mencirikan operasi standar yang dapat diterapkan pada setiap jenis hubungan untuk setiap jenis struktur.
- Mencirikan manipulasi yang dilakukan oleh aplikasi tampilan dalam hal operasi standar pada hubungan standar. Akibatnya, aplikasi tampilan tidak memanipulasi catatan. Alih-alih menjalankan operasi standar pada hubungan standar pada struktur standar. Seseorang dapat memetakan dari tindakan aplikasi tampilan ke operasi standar. Dengan adanya pemetaan ini, aplikasi tampilan apa pun dapat memanipulasi rekaman apa pun, atau setidaknya struktur dalam rekaman yang hubungan yang diperlukan telah ditentukan. Perangkat lunak yang menerapkan operasi standar dapat diubah menjadi layanan mikro yang dikelola oleh lingkungan pelestarian.

Proses pengawetan kemudian terdiri dari manipulasi struktur dalam catatan, atau penugasan properti ke kumpulan catatan, atau pembentukan hubungan antar catatan. Jika kami memiliki serangkaian proses pengawetan yang dapat dibalik yang ditentukan, kami dapat menegaskan bahwa lingkungan pengawetan di masa mendatang dapat mengubah semua catatan kembali ke bentuk aslinya. Lingkungan pelestarian masa depan dapat dengan benar menafsirkan konteks informasi pelestarian dari masa lalu dan menerapkan kebijakan pelestarian yang sama. Konsep "objek persisten" secara efektif menggabungkan konsep migrasi (transformasi format rekaman ke standar baru), dan emulasi (memindahkan kemampuan aplikasi kreasi asli ke teknologi penyimpanan baru).

Teori pengawetan perlu menunjukkan konsistensi internal antara kriteria penilaian, aturan yang mengontrol pelaksanaan proses pengawetan, dan informasi status persisten yang dihasilkan oleh penerapan proses pengawetan. Sistem ini konsisten secara internal jika semua atribut preservasi yang diperlukan untuk mengukur prinsip preservasi (keaslian, integritas, lacak balak, respek des fonds) dapat dihasilkan atau divalidasi melalui penerapan proses manajemen; dan jika informasi status persisten yang dihasilkan oleh penerapan kebijakan manajemen dipertahankan sebagai atribut preservasi yang dapat ditanyai untuk memverifikasi properti preservasi. Kami tidak dapat memiliki situasi di mana atribut pengawetan yang diperlukan untuk menilai prinsip pengawetan tidak dapat dikontrol atau diverifikasi oleh salah satu aturan pengawetan. Kita juga tidak dapat memiliki situasi di mana informasi status persisten yang dihasilkan oleh penerapan aturan tidak disertakan dalam representasi konteks lingkungan pelestarian yang dimigrasi ke masa depan.

Uji asam lingkungan pengawetan adalah apakah ia menjelaskan seluruh konteks informasi pengawetan dengan cukup baik sehingga rekaman dapat dipindahkan ke lingkungan pengawetan independen tanpa kehilangan keaslian atau integritas. Ini membutuhkan migrasi tidak hanya catatan, tetapi juga karakterisasi

konteks pelestarian lingkungan. Lingkungan pengawetan baru harus menerapkan kebijakan pengelolaan yang sama, proses pengawetan yang sama, menggunakan ruang nama logis yang sama, dan mengelola informasi status persisten yang sama. Jika semua komponen konteks ini dapat diekspresikan dan dimigrasi ke lingkungan pengawetan baru, maka konteks pengawetan dijelaskan dengan benar.

Harapannya, kita bisa mengembangkan teori pelestarian. Komponennya adalah:

**Karakterisasi:**

- Definisi ruang nama tetap
- Definisi operasi standar yang dilakukan pada ruang nama yang ada
- Karakterisasi perubahan informasi status persisten (terkait dengan setiap ruang nama tetap) yang terjadi untuk setiap operasi standar
- Karakterisasi transformasi yang dibuat ke rekaman pada setiap operasi standar

**Kelengkapan:**

- Demonstrasi bahwa rangkaian proses pengawetan sudah selesai, memungkinkan layanan yang dekomposisi setiap proses pengawetan menjadi set mikro-menjalankan operasi standar
- Demonstrasi bahwa kebijakan manajemen preservasi sudah lengkap, Demonstrasi bahwa memungkinkan kontrol semua proses pelestarian, validasi
- informasi status persisten sudah lengkap, memungkinkan semua kriteria penilaian

**Tuntutan:**

- Jika operasi dapat dibalik, maka lingkungan pelestarian di masa depan dapat mendukung akses, dan membuat ulang rekaman dalam bentuk aslinya, menjaga keaslian dan integritas, menampilkan catatan
- Konsekuensinya adalah bahwa sistem seperti itu akan memungkinkan catatan dimigrasikan dengan antara implementasi independen lingkungan pelestarian, tetap menjaga keaslian dan integritas

Lingkungan berbasis aturan iRODS adalah langkah pertama menuju pembuatan repositori pelestarian digital yang dapat dipercaya. Definisi aturan standar, layanan mikro standar, operasi standar, dan informasi status persisten standar yang disediakan oleh iRODS dapat mengarah pada teori pelestarian digital.

Akhirnya, kami mengamati bahwa perpustakaan digital (Moore, Rajasekar, & Wan, 2005) dan koleksi bersama dapat diimplementasikan dari kemampuan yang sama yang digunakan untuk mendukung lingkungan pelestarian, asalkan kebijakan pengelolaannya diterapkan. Kisi data iRODS adalah infrastruktur manajemen data umum yang dapat disetel untuk mendukung penyimpanan data, publikasi data, berbagi data, atau analisis data melalui spesifikasi kebijakan manajemen data yang sesuai.

### Ucapan Terima Kasih

Pekerjaan ini didukung oleh Administrasi Arsip dan Arsip Nasional di bawah perjanjian kerja sama NSF 0523307 melalui suplemen untuk SCI 0438741, "Cyberinfrastructure: From Vision to Reality" dan oleh National Science Foundation, ITR 0427196, "Constraint-based Knowledge Systems for Grids, Perpustakaan Digital, dan Arsip Persisten".

### Referensi

- Baru, C., Moore, R., Rajasekar, A., & Wan, M. (1998). Sumber daya penyimpanan SDSC makelar. Dalam Prosiding *Konferensi CASCON'98, Toronto, Kanada*.
- Komite Konsultatif untuk Sistem Data Luar Angkasa (CCSDS). (2002). Model referensi untuk sistem informasi arsip terbuka (OAIS). Diakses 10 Juni 2008, dari <http://public.ccsds.org/publications/RefModel.aspx>
- Moore, R. (2003). Proyek San Diego: Objek yang gigih. *Archivi & Komputer, Automazione e Beni Culturali*. l'Archivio Storico Comunale di San Miniato: Pisa, Italia.
- Moore, R. (2006). Membangun lingkungan pelestarian dengan teknologi jaringan data. *Pengarsip Amerika* 69, 139-158.
- Moore, R., Rajasekar, A., & Wan, M. (2005). Kisi data, perpustakaan digital, dan arsip persisten: Pendekatan terintegrasi untuk menerbitkan, berbagi, dan mengarsipkan data. Edisi Khusus dari Prosiding *IEEE on Grid Computing, vol. 93, tidak. 3*, hlm.578-588. Masyarakat Komputer IEEE: Piscataway, New Jersey.
- Moore, R., & Smith, M. (2006). Penilaian repositori digital terpercaya RLG Persyaratan. Di *Lokakarya Konferensi Bersama tentang Perpustakaan Digital tentang Kurasi Digital & Repositori Terpercaya: Mencari Kesuksesan*. Chapel Hill, Carolina Utara.
- Moore, R., Wan, M., & Rajasekar, A. (2005). Pialang sumber daya penyimpanan: Generik infrastruktur perangkat lunak untuk mengelola data yang didistribusikan secara global. Dalam Prosiding *Konferensi IEEE tentang Data yang Didistribusikan Secara Global*, hlm. 65-69. Masyarakat Komputer IEEE: Piscataway, New Jersey.
- Arsip Nasional dan Administrasi Arsip. (2003). Arsip catatan elektronik daftar kemampuan. Diakses 10 Juni 2008, dari <http://www.archives.gov/era/pdf/requirements-amend0001.pdf>

- Rajasekar, A., Wan, M., Moore, R., & Schroeder, W. (2006). Prototipe berbasis aturan sistem manajemen data terdistribusi. Di *Lokakarya Komputasi Terdistribusi Berkinerja Tinggi tentang Manajemen Data Terdistribusi Generasi Berikutnya*. Paris, Prancis.
- Grup Perpustakaan Riset & Arsip Nasional dan Administrasi Arsip. (2007). Audit & sertifikasi repositori terpercaya: kriteria dan daftar periksa. Diakses 10 Juni 2008, dari <http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/trac.pdf>
- Smith, M., & Moore, R. (2006). Kebijakan arsip digital dan digital terpercaya repositori. Dalam Prosiding *Konferensi Kurasi Digital Internasional ke-2: Kurasi Data Digital dalam Praktek*. Glasgow, Inggris.