Agreements Engine (SCSB) Pilot Final Report

Date: September 2024

Lead Authors

Marian Leon, BTAA
Kelly Sattler, Michigan State University Libraries
Maurice York, BTAA

Reviewers

Agreements Engine (SCSB) Pilot team (see below)

Report Review & Approval History

As of Sept. 10, 2024

- BIG Collection Steering Committee [9/16/24]
- BTAA Library Deans Executive Committee [10/2/24]
- Report published [12/9/24]

Overview

The Agreements Engine pilot was initiated in May 2021. At that time, the BIG Collection was still in its conceptual stage. Because of this, the pilot team's initial efforts were focused on understanding the components of SCSB and how it was being deployed by ReCAP, followed by exploring its potential application by Big Ten libraries within the design architecture of the Systems of Trust for Print Collections. Throughout 2021, the team gained insights into the tool and began to identify relevant use cases. In 2022, a preliminary model for the Shared Print Program was drafted, and in January 2023 the Visiting Program Officer for Shared Collections joined the pilot team. With the framework for the shared print program articulated, the Agreements Engine (SCSB) pilot team was able to concentrate its efforts on testing the SCSB software to determine its suitability for supporting the objectives of the BTAA shared print program.

The pilot team was tasked with investigating a specific, tightly focused question:
Does SCSB, the middleware developed by HTC and used by ReCAP, retain the information that Big Ten libraries need to verify that we are meeting the agreements for items designated to the Shared Collection?

Project Team Members

Administration:

- Maurice York - Project Sponsor; BTAA Director of Library Initiatives
- Marian Leon - Initiator; BTAA
- Susanne Garrison - Master Scheduler; BTAA
- Kelly Sattler - Project Manager for the ReCAP SCSB pilot; MSU
- Mary Laskowski* - Resource & Connector between BTAA projects
- Krisellen Maloney* - Functional Lead; Rutgers/BTAA
- Karla Strieb - VPO for Shared Collections

Fulfillment:

- Marie Waltz* - Group Lead; CRL
- Aaron Tomak - MSU
- Kim Pierce - MSU
- Mary Laskowski* - IL
- Rachel Watters - WI

Governance & Sustainability:

- Ian Bogus - ReCAP
- Maurice York - BTAA
- Nabeela Jaffer* - U of Mich

System and Technology:

- Barak Zahavy - ReCAP
- Bruce Barton* - WI
- Lee Konrad - WI
- Rachel Watters - WI
- Krisellen Maloney* - Rutgers / BTAA

Metadata and Data Analysis:

- Amy Wood* - CRL
- Lisa Lorenzo* - MSU
- Steve Meyer - UW-Madison
- Nate Florin*- CRL
- Dao Rong Gong - MSU
- Margaret Kelly - U of Mich
- Michael North - Northwestern
- Bob Trautvetter - Northwestern
- Alice Tippit - Northwestern
- Edgar Garcia - Northwestern

* rolled off the pilot project before completion

Project Charge & Deliverables

The team was charged with overseeing a test installation of the ReCAP SCSB middleware to assess its potential for supporting the BTAA Shared Print Program.

The basic questions this pilot was asked to address were:

- How does the ReCAP SCSB code, as written, match up with BTAA functional requirements?
- Can BTAA install and run the ReCAP middleware in a BTAA university AWS environment using our own metadata systems requirements and integrations?
- Will the middleware scale to BTAA needs?
- Is there a feasible path toward partnership, sustainability, and shared governance with ReCAP?

Outcomes and Deliverables:

The key outcomes as articulated in the charge:
Cycle 1:

- A report on required functional and design requirements for BIG Collection shared collection middleware and comparison to ReCAP existing code and technical roadmap

Cycle 2:

- A functioning cloud-hosted installation of the software, populated with a sample set of records
    - For the pilot, we will use an AWS environment hosted at one of the participating universities.

    - Understanding and articulating the hosting environment needs for bringing this to scale.
    - Additional hosting models will be explored and a final recommendation will be provided.
- A report on technical tests of system functions for:
    - the agreements engine and its capabilities
    - metadata record ingest
    - indexing
    - fulfillment functions
    - etc.
- Test integrations with key BTAA ILS and discovery systems (such as Alma, Voyager, Folio, etc.)

Cycle 3:

- An assessment of the resourcing and effort that would be required for BTAA hosting and development of ReCAP software
- A report on the feasibility and scalability of implementing the SCSB software in the BTAA environment
- Reports from sub-groups formed to assess specific functions of the software (e.g. policy configuration and functions, fulfillment, etc)
- A report on potential partnership and governance models with ReCAP

Vision & Principles

The shared collection is currently just over 500,000 print volumes; the vision is for this to grow to tens of millions. From this starting point, we do not know the pace at which the shared pool will grow and scale. However, we expect the process of converting from local to shared will take many years and incorporate new content as it is collected by members. From a systems or infrastructure services standpoint, the vision is that every volume committed will be immediately incorporated into shared stewardship and that a set of operational agreements governing committed shared items will be immediately applicable and operable for that item. We refer to the primary piece of middleware that we envision as the backbone for building the Shared Collection as the "Agreements Engine."

We are exploring whether and to what extent SCSB as built provides a starting point for our conceptual specifications for the Agreements Engine to:

1. identify the gaps from current state to what we envision, and gain an assessment of resources (cost and time) that would be necessary to close those gaps;
2. or, if SCSB could be used for our immediate needs while the features in our desired future state are built out
3. or, whether repurposing SCSB is infeasible and what we are looking at would be a fresh build of a new system (with the as-built SCSB software either serving as a bridge or not as the case may be).

The purpose of the Agreements Engine is to ensure that the shared collection, along with the agreements governing stewardship and access, is accessible and functional for BTAA library members. The Agreements Engine serves as a unified System of Record for these agreements and offers a suite of services that operate on the designated items, such as matching, management rules, and indexing.

Methodology

Initially, the pilot team divided into subgroups, each focusing on specific system requirements related to fulfillment, governance and sustainability, metadata and data analysis, and systems and technology. Each subgroup set their own goals and met regularly between the monthly full-team meetings. Early in the process the team received a demonstration of the SCSB software from the ReCAP consortium to understand how it functions within the ReCAP environment.

As each team delved deeper into their areas, questions were formed.

- There was a noted overlap and interest in the findings from the UBorrow system refresh pilot.
- Possible testing scenarios were created.
- A survey was developed to determine what institutions would be able to provide in terms of data and how onerous the extraction process would be.
- Concerns were also brought forth on how SCSB appeared to deal with records and duplication.

At this point, two things happened:

- A White Paper on the Agreements Engine for the BTAA Shared Collection was produced and distributed to the pilot team. The paper clarified the learnings of the pilot to date.
- The pilot team decided to hire HTC Consulting, the developers of SCSB, to set up a test instance for the BTAA instead of having a Big Ten university host.

HTC created a SCSB instance for BTAA in their AWS space and the full pilot team was given access for testing. The Gold Rush team worked with the Metadata subgroup to select a broad range of data that was mostly successfully loaded into the system. The data came from three Alma catalogs and one Folio catalog. Meanwhile, testing requirements were clarified, and testing was successfully completed once the data was loaded.

Learnings and Recommendations

Learnings

1. The pilot team determined that the Big Ten libraries would not use SCSB's circulation features, refocusing SCSB's purpose for BTAA to serve primarily as an index of records. This index would include a field(s) to track compliance with agreements related to the Shared Collection.
2. If a decision is made to use SCSB for the agreements engine, the field "CGD Status" would need defining and refining to match the agreements as determined by the Shared Print Agreements & Practices Working Group.
3. SCSB did not fully pass the testing for a minimum viable product as it does not do a FRBR-like roll up. This would require development work by HTC, but is not necessary for near-term use.
4. Higher level requirements desired for our future state, such as notifications, are also missing from SCSB as noted in the Functional test list tab - "High level requirements". We have noted which features exist now, which would be easy to implement, and which would be harder/more time consuming to add. We have also identified which features ReCAP would like to see built into SCSB.

5. At this point, BTAA members would need to log in directly to SCSB to run reports. Eventually, the connecting software (Ethl and Flora on the Systems of Trust diagram) would need to be designed, written and implemented in order for the discovery of these committed items to be found.
6. When uploading new records to SCSB, if even a single record in the file fails, the entire file upload will fail. This raises concerns about the efficiency and reliability of the process for production-level operations. Note - this holds true when adding new records. Once a record is in the system, updating that record can be accomplished with a traditional MARC record update.
7. The pilot team agrees that there are components of SCSB as written that can serve the immediate need of providing a shared repository for holding agreements related to items designated to the Shared Collection.
8. Based on our analysis, development will be required regardless of whether BTAA selects SCSB or another solution. While SCSB is a well-established system designed to meet certain functional requirements, such as facilitating shared repository circulation among consortial members, these are related but not entirely aligned with our requirements for an agreements engine.

Recommendations/ Future Considerations

1. We recommend that a new team is formed to refine the anticipated requirements and desired features for an agreements engine, including discovery. This work must be informed by the Shared Print Agreements (currently in draft pending review by the Deans in 2025) and the architecture in the Systems of Trust.
2. We recommend ongoing environmental monitoring for products as they are developed and made available for a more robust solution.
3. We recommend loading the data into an agreements engine from a single, centralized source, such as a data lake, rather than collecting data individually from each institution and then converting it into a proprietary format of the future commitments engine. This approach would reduce the burden on the institutions and require the development of only one conversion program rather than creating separate programs for each ILS system in use.

Next Steps

1. Consolidate the documentation from this pilot in BTAA spaces to be available to future pilot teams.
2. Create a new team to continue the work of identifying the needs and desires of an agreements engine and have that team scan the environment to see if there are already built systems that could possibly meet these needs.
3. If the decision is to leverage SCSB as an interim solution, we will need to commit resources to enable it to function. (i.e., AWS server space, and #2 above under Learnings)
4. Create a group to identify the functionality needed for a central data source that can function as a single source of metadata for analysis tools and a commitments engine.
5. Thank the members of the SCSB pilot team for their contributions and disband the team.

Appendix

I. Links

A. ReCAP SCSB Test Install Charge
B. DRAFT White Paper on the Agreements Engine
C. Functional Testing Spreadsheet
D. Shared Print Agreements (currently in draft pending review by the Deans in 2025)