SNL CSP Library Archive Project FY19 Report

Kenneth M. Armijo, Luis Garcia-Maldonado, Alice Parsons, Sandra Pacheco, Annette Chavez, Micaela Galindo, Dimitri Madden, Charles Andraka and Paul Gauche.

October 2019

1. CSP PROJECT DEFINITION

1.1. Vision, Purpose, and Justification

Sandia National Laboratories (SNL) National Solar Thermal Test Facility (NSTTF) and Tech Library have been collaborating over the course of the FY19 period to establish and maintain the first and only digital collection in the world of Concentrating Solar Power (CSP) related historical documents, dating back to the CSP program inception here at Sandia in the 1970's thru to the present. The unclassified, unrestricted (UUR) collection, comprised of internally generated Sandia documents as well as a significant number of external reports will be searchable via both the Sandia website and OSTI, DOE's document repository. DOE is currently championing efforts to get the collection launched, where international partners, which include Australia and Germany, plan to forward related documents to be included in the CSP archive. Advancing this transformative project will make the CSP collection accessible to the Sandia and global communities.

1.2. Project Scope

The scope of this project only considers activities that pertain to archival of historical documents and not for any current or future documents. Additionally, this work does not include archival of solar documents that pertain to photovoltaics (PV) and other renewable energy resources that are not at all related to CSP. However, activities within the scope of this effort pertain to the planning and implementation of a searchable, digital repository for historical CSP documents within Sandia, and accessible (by permission) by external entities including the U.S. DOE. The activities within the scope of work performed in this effort include:

- Scanning and ingestion of CSP documents into the Library catalog
- Implementing a search function and document access protocols
- Coordinating collection access via OSTI

Documents include:

- SAND reports (~800)
- External documents held by 8823 (~1800)
- SAN reports (~150)
- UCI documents (~30)
- Documents held by DOE (~3000)
- Documents held by international partners (unknown)

2. FY19 PROGRESS & FY20 WORK SCOPE

During FY19, the SNL team completed the following tasks where the team requests from DOE a cost-extension to complete the remaining work items by September 2020 (End of FY20).

1. All UUR CSP SAND reports where uploaded into the SNL SRN Library catalog.
   a. All full text reports have been sent to OSTI (about 700).
2. The Library catalog record is continuing to be updated for each record with its respective metadata inputs, such as OSTI_ID, a CSP Series tag, and a link to the full text report.
3. The group has built a local archival record of over 3,000 documents where over FY19, another 500 documents were scanned for archiving.

a. After the documents were scanned, they were turned into an OCR pdf. The documents were inputted into a database. The database contains all the relationships of the documents.

4. 700 SAND reports have been sent to OSTI from Sandia, where another 80 SAND reports are now cataloged on the SCN, which need to have their records downshifted to the SRN and processed into the catalog/sent to OSTI.

    a. While the team was hoping to have all the SAND reports processed by end of FY19, the process appears to be taking longer due to classification and volume processing challenges. Since this activity is labor intensive, we propose the following changes to the overall project plan:

        i. Change the due date of SAND report archiving, with metadata upload and search interfaces to FY20 and processing of external documents with a completion date of Sept. 2020.

For FY20 the following items need to be completed to close-out the project:

1. SNL is in midst of processing approximately 100-200 remaining SAND reports, where work includes re-classification of OUO SANDs (~30), moving catalog records from the SCN to the SRN Library Catalog (~80), troubleshooting OSTI records that aren't shown as "complete" (~135), and manually cataloging and processing reports that were not originally in the Library Catalog.

2. A list of external reports was received from 8823 and was evaluated against our Library Catalog for inclusion. The Library has just a handful of the ~1800 titles on the list. The digital files will be sent to the Library and metadata added to the Library catalog.

3. SNL is working with OSTI to provide metadata tagging for all reports as a special collection.

4. An estimated number of approximately 15,000 documents (some are up to 400 pages), and ~1000 Architecture Design (some up to 20ft in length), still need to be scanned and entered into the CSP archival database. The target goal date to complete all the scanning and the input to the database is estimated to be May 2020. We also anticipate completing the processing of existing external documents (~1900) by May 2020, however it will take until the end of FY20 to finish all others that are not currently at SNL.

3. IMPLEMENTATION PLAN

In FY20, pending the approval of the no-cost extension by DOE, the SNL team plans to proceed with the following CSP Library Archival implementation plan:

CSP TEAM MATRIX

ROLE	PERSON
Product Manager	Kenneth Armijo
CSP Collection Mgmt. Team	Luis Garcia Maldonado, Dimitri Madden, Micaela Galindo, Sandra Pacheco, Alice Parsons, Annette Chavez
Stakeholders	Paul Gauche, Manager, Concentrating Solar Technologies Don Guy, Manager, Technical Library
Champion	Matt Bauer, DOE/EERE/SETO (Solar Energies Tech. Office)
IT support	TBD

3.1 Schedule, Milestones, and Deliverables

1. **MILESTONE #1: Full-text for all UUR SAND documents in the CSP collection are ingested, catalogued, and sent to OSTI. (12/31/19)**

   a. All UUR CSP SAND documents (~700) ingested, catalogued in Library database, and sent to OSTI. (11/30/19)

   b. Catalog records on SCN for UNC CSP SANDS downshifted to SRN Horizon, updated, and full text sent to OSTI (~80). (11/30/19)

   c. OUO SAND documents (~30) rereviewed by Classification and added to CSP Collection. (12/31/19)

2. **MILESTONE #2: All external documents, including those held by 8823, external partners, and DOE, are ingested and catalogued. (9/30/20)**

   a. External documents held by 8823 ingested and catalogued in Library database (~1900). (5/1/20)

   b. DOE documents (~3000) ingested and catalogued in Library database. (9/30/20)

   c. International partners identified, and documents ingested and catalogued in Library database. (9/30/20)

3. **MILESTONE #3: CSP Collection is browsable by the public at both Sandia and, and procedures are in place for handling document requests. (9/30/20)**

   a. Public search interface on Sandia's website is functional and accessible. (5/1/20)

   b. Public search ability through OSTI is functional and accessible. (9/30/20)

   c. Workflow for report requests is defined and documented. (9/30/20)

   d. Any IT processes that will remain in place to make public access functional are documented. (9/30/20)

3.2 Risk Management

Risk identification will occur on an ongoing basis through planned meetings and team/stakeholder feedback. The team has identified five risks (below), where we are

implementing a risk-reduction strategy to manage these challenges within the constraints of time, scope, and cost.

1. If the Library database is switched from Horizon to Koha, there may be a delay in getting documents ingested, catalogued, and meeting Milestone #1.

    - An effort to move the Library catalogue to Koha is actively underway and should be completed by the end of CY19. The team is mitigating this strategy by continuing to maintain and grow the current SNL Horizon library catalog, which will run concurrently with the Koha application until Koha can mirror the necessary functionality provided by Horizon.

2. Due to issues related to changing information sensitivities, the Library does not distribute external reports to external requestors. The Library does not send external reports to OSTI, which may create a significant difference between Sandia's CSP collection and OSTI's.

    - SNL will mitigate this risk by defining how the collection will be accessed and consider solutions that will make the content of both repositories as consistent as possible.

3. If DOE does not fund the project in FY20, it is unclear how much support Org. 8823 can provide to:

    - Ingest documents from our DOE and international partners, which would jeopardize meeting Milestone #2; and

    - Support any IT expertise needed to do batch processing and create a user-friendly front-facing search/request capability, impacting Milestone #3.

        o   Considerable progress on this collection should be made by 9/30/19 which will hopefully inspire DOE to continue funding the project through 2020.

4. The project is labor intensive and may be hard to absorb given other funded work priorities.

    - The SNL team is hiring two student interns to help with the work load to complete these tasks in an aggressive manner.

5. It is unclear how external, non-SAND reports will be added to the internal SNL archival database, which may reduce its overall utility, particularly to external entities and SNL-collaborators.

    - The SNL team is having on-going discussions with management to determine if solutions, such as an SNL SAND report wrapper, with re-approval process could be used to process these documents. The External Reports, which includes those currently at SNL, and that being added by U.S. and international partners, can be added to our collection and treated as UUR unless otherwise specified, even though they are not marked Unlimited Release. If any contain proprietary or OUO info, they would most certainly be marked as such. All Sandians would have immediate access via the SRN to all UUR reports, SAND reports and externals. The general public would not have direct access to externals but would have to gain access through request made to Sandia. However, they would have access to the rest of the documents sent to OSTI, which include all of the SAND reports.

3.3 Completion Criteria

The CSP Collection Project will be complete when:

- All documents have been successfully acquired, stored, catalogued, and made accessible (to the extent they can be accessible) via the Sandia and OSTI websites.
- Documentation is created of any IT-related processes that require maintenance.
- Workflows are in place for handling CSP document requests.
- All internal SAND, and external documents have been included within the SNL CSP Library archival database.
- An internal and external accessibility process has been put in place to access the documents.