U.S. General Services Administration
Federal Acquisition Service
Technology Transformation Services
Office of Solutions
Data and Analytics Portfolio
Data.gov Program

Metadata Management Services Product-Preliminary Plan

Background:

Initial funding was provided during FY 2018 through Cross Agency Priority (CAP) goal funding.

The proposal for CAP Funding was described as:

- Project Name: Data Catalog Kit to Support Agencies Implementation of the Federal Data Strategy

- Background and Project Description: Federal agencies need access to consistent and easy metadata management, data catalog, and data hosting capabilities.

- Project Description: Creating a government-wide data catalog platform that supports individual catalog instances that can be branded, based on Data.gov's standardized open source codebase (CKAN). This is similar to the capabilities of sites.usa.gov, but is instead focused on data management.

CAP funding supported the initial award of a contract for the work on 9/28/2018, with one base period of six months, and four, six-month long option periods. CAP funding supported the base period only. All subsequent option periods have been funded by the existing Data.gov budget. Data.gov's existing, annually appropriated budget, which does not contain additional funding for this project, currently supports the work (referred to as "CKAN project") and the existing Data.gov system. The government staff working on this project is the existing Data.gov Program Management Office (PMO) of 3 full time employees.

Current Status:

The CKAN project contract is currently in option period 3, which will end in October 2020. There will be one remaining six month option period thereafter.

Work completed to date:

Base period (9/2018 to 5/2019, original base period extended due to government shutdown):

- Infrastructure research - prototype
- Multi-tenant infrastructure
- Architecture plan
- Design instance creation page
- Provisioning app design/UI and API
- Docker images: CKAN
- Create documentation
- FedRAMP research/security planning
- DCAT-US metadata extension integration
- Compliance automation research
- Prototype and demonstration of a multi-tenant CKAN platform

Option Period 1 (5/2019-11/2019)

- Data.gov themes and extensions
- Compliance automation for future ATOs/Open Controls
- FedRAMP template research
- CKAN Cloud Operator (CCO) refactoring
- CCO Unit tests and documentation
- CKAN Cloud Docker
- User audit and baselining
- Agency user interviews/journey-mapping
- Catalog harvester improvements
- CI/CD analysis
- CSW Harvester
- CKAN 2.8 theme
- Next Generation harvester work
- Airflow for harvester

Option Period 2 (11/2019-5/2020)

- CKAN Cloud Operator (CCO) improvements
- Legacy Harvester and extensions research
- Security work - multi-tenant platform MVP
- CCO on AWS
- CKAN on Cloud.gov research
- USWDS research
- Next Generation harvester
- Design of “top tasks” survey/conduct survey
- Research DCAT-US metadata standard
- Catalog upgrade work CKAN 2.8
- SAML2 work - user authentication
- Metadata form research plan
- Metadata form user research

Option Period 3 (5/2020 to 11/2020)

- Design improved metadata entry form
    - Incorporating human-centered design and research with dataset owners
    - Following current metadata schema
    - Improvements in navigation, pre-populated fields, logic, help text, visibility
    - Incorporating US Web Design Standards (USWDS)

- User research rounds on improved metadata form

- CKAN 2.8 work - data catalog
    - Modernization of catalog.data.gov
    - Addressing security and compliance issues
    - Reduction in required operations and maintenance tasks

- Improved harvesting - CKAN 2.8

- Remaining work:
    - Deploy improved metadata form
    - Research plan on additional user research with data managers
    - Conduct human-centered design and research with data managers
    - Determine product features from user research (workflow/dataset hosting)
    - Competitive analysis of different metadata management platforms

- Iterative research and development on government-wide data management tools and solutions
- Inventory.data.gov/catalog.data.gov on CKAN 2.8
- Migration of data.gov services to cloud.gov
- Enterprise data inventory reference implementation and playbook
- Design and implementation of additional features through experimental prototypes and human-centered design
- Investigation of agency use of APIs and standards for interoperability of federated data across government
- Migration plan for seamless transition of existing data to production system
- Development of production system
- Research of security/infrastructure requirements including agency customer requirements
- Infrastructure/Security analysis with respect to existing Data.gov
- Migrating existing agencies to metadata management system featuring improvements
- Onboard new agencies to metadata management system featuring improvements
- Staffing plan for ongoing operations and maintenance

Option Period 4 (11/2020-5/2021)

Planned work for CKAN product:

- Deployment of production system
- Ongoing human-centered design and research as needed
- Security analysis on ATO
- Agency outreach and communication on current services and reimbursable plan
- Market research - agency demand, pricing, security, SLAs,
- Estimate of infrastructure, professional services, other costs for agency services

<u>Data.gov Program Work FY 2021:</u>

- Complete migration of existing Data.gov to cloud.gov
- SSP completion, assessment, obtain GSA ATO at FISMA moderate for existing system
- Determine requirements for contract support for recompete
- Determine requirements re ATO on agency services vs existing Data.gov/catalog (GSA ATO/FedRAMP/FISMA levels)

- Research procurement options with flexible capacity for growth in agency customers
- Research infrastructure options for scaling for growth in agency customers/usage (cloud.gov, FCS, existing TTS vehicles for cloud infrastructure)
- Execute procurements for technical support and infrastructure
- Determine government staff and contract needs for supporting agency services on reimbursable basis
- Agency outreach/market research on cost levels and reimbursable structure
- Develop program materials (terms, pricing, interagency agreements)
- GSA internal reviews and approvals on reimbursable plan
- Development and modernization of government data standards

Data.gov Program Work Post May 2021 on Metadata Management Services Product:

- Begin new contracts on technical support to CKAN product
- Procure infrastructure (cloud.gov or internal GSA options [IAAs], existing GSA vehicles for cloud infrastructure)
- ATO process as determined by agency customer requirements
- Formalize agency product program and reimbursable plan
- Obtain initial staffing and resources for reimbursable product
- Finalize pricing
- Create and execute interagency agreements
- Continue human-centered design and research into filling gaps in federated data management

Preliminary Estimate of Resource Requirements

Assumptions:

- Agency product with weekday, business hours SLA
- Metadata management tool with additional services, without large scale dataset hosting or other, yet unidentified capabilities that would require significantly larger support staff

Staffing (Annual):

1 FTE program management (can be combination of partial government FTE and contractor program management)	$260,000

1 FTE System Architect	$270,000
1.5 FTE Web Developer	$375,000
Infrastructure (Annual)	$125,000
Miscellaneous (Subscription, OTS software, annual)	$10,000
Total Annual Cost:	$1,040,000

Percentage of recoverability of cost will depend on the number of agencies agreeing to use the service on a reimbursable basis, and the results of pricing research on what pricing level would be competitive to other options available to agencies.

Potential Benefits:

- Could be determinative in whether agencies have datasets in Data.gov and meet their OPEN Government Data Act requirement
- Increase in number of agencies and datasets being made available, for benefit of the public, students, businesses, researchers
- Increase in the number of evidence-based policy decisions leveraging internal agency datasets
- Increase in data sharing among federal agencies
- Efficiencies in agencies using common solution rather than duplicative costs in infrastructure or program management across multiple agencies
- Increase in interoperability between government systems
- Cost savings in agencies avoiding more expensive options
- Reduction of risk of data mishandling, avoiding liability in integrity, availability, and confidentiality.
- Reputational gains for Solutions/TTS/FAS/GSA in providing a new shared service
- Increased innovation by multiple agencies using a common, open source technology
- Potential to solve problems for agencies beyond open data

Potential Risks:

- Agency user research does not identify significant interest in a shared services product
- User research does not identify market need beyond basic metadata entry currently provided through inventory.data.gov
- Lack of agency budget in any additional expenditures for open data purposes

- Agency requirements on ATO (is FedRAMP and FISMA moderate required) are prohibitive of a shared service on metadata management
- Infrastructure obstacles (cloud.gov not possible, or current TTS vehicles in obtaining infrastructure are not available)
- Lack of procurement options to support a reimbursable product with flexible demand