

# NARA, the OAIS-RM, and NOAA's Next-Generation Cloud Archive

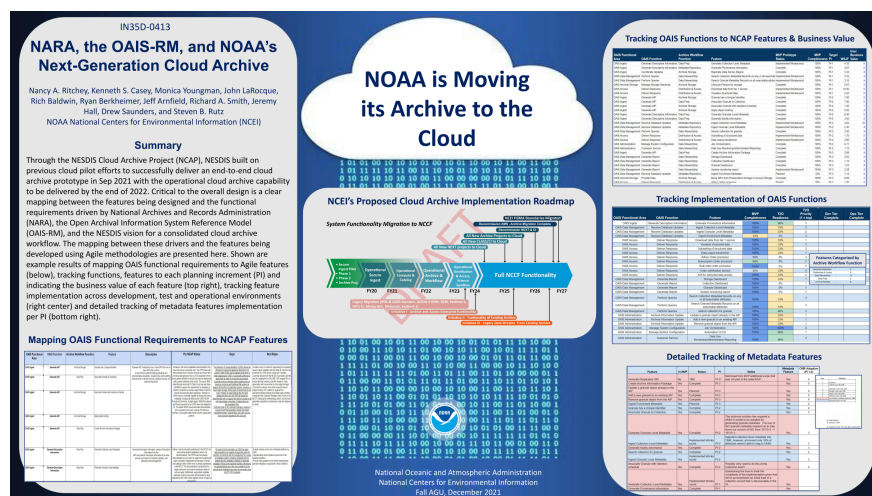
Nancy Ritchey<sup>1</sup>, Kenneth Casey<sup>1</sup>, Monica Youngman<sup>1</sup>, John LaRocque<sup>1</sup>, Rich Baldwin<sup>1</sup>, Ryan Berkheimer<sup>1</sup>, Jeff Arnfield<sup>1</sup>, Richard A. Smith<sup>1</sup>, Jeremy Hall<sup>1</sup>, Drew Saunders<sup>1</sup>, and Steven Rutz<sup>1</sup>

<sup>1</sup>NOAA/NCEI

November 24, 2022

## Abstract

Through the NESDIS Cloud Archive Project (NCAP), NESDIS is building on previous cloud pilot efforts and making significant progress on delivering an end-to-end cloud archive prototype in late 2021 and operational cloud archive capability in 2022. Critical to the overall design is a clear mapping between the features being designed and the functional requirements driven by National Archives and Records Administration (NARA), the Open Archival Information System Reference Model (OAIS-RM), and the NESDIS vision for a consolidated cloud archive workflow. The mapping between these drivers and the features being developed using Agile methodologies will be reviewed, and a technical summary of NESDIS' plans to implement data archiving using commercial cloud services will be provided.





# NARA, the OAIS-RM, and NOAA's Next-Generation Cloud Archive

Nancy A. Ritchey, Kenneth S. Casey, Monica Youngman, John LaRocque, Rich Baldwin, Ryan Berkheimer, Jeff Arnfield, Richard A. Smith, Jeremy Hall, Drew Saunders, and Steven B. Rutz  
NOAA National Centers for Environmental Information (NCEI)

## Summary

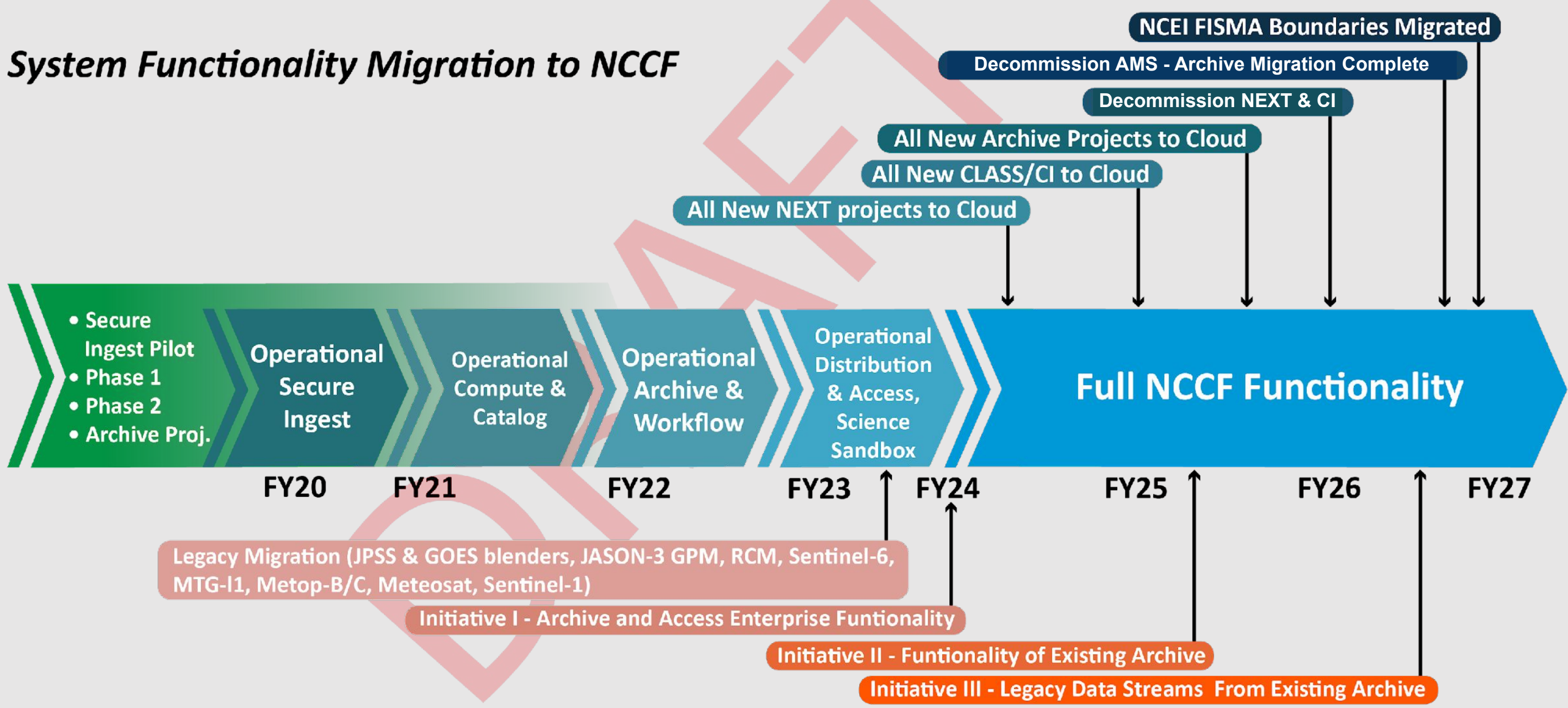
Through the NESDIS Cloud Archive Project (NCAP), NESDIS built on previous cloud pilot efforts to successfully deliver an end-to-end cloud archive prototype in Sep 2021 with the operational cloud archive capability to be delivered by the end of 2022. Critical to the overall design is a clear mapping between the features being designed and the functional requirements driven by National Archives and Records Administration (NARA), the Open Archival Information System Reference Model (OAIS-RM), and the NESDIS vision for a consolidated cloud archive workflow. The mapping between these drivers and the features being developed using Agile methodologies are presented here. Shown are example results of mapping OAIS functional requirements to Agile features (below), tracking functions, features to each planning increment (PI) and indicating the business value of each feature (top right), tracking feature implementation across development, test and operational environments (right center) and detailed tracking of metadata features implementation per PI (bottom right).

## Mapping OAIS Functional Requirements to NCAP Features

OAIS Functional Area	OAIS Function	Archive Workflow Function	Feature	Description	Pre NCAP Status	Gaps	Next Steps
OAIS Ingest	Generate AIP	Archival Storage	Granule has a Unique Identifier	Generate AIP transforms one or more SIPs into one or more AIPs that conform to the OAIS data formatting standards and documentation standards. Included here are the NARA requirements to declare records, organize records, and execute disposition.	Immature, with some capabilities demonstrated in the cloud and some in progress only. The CP2P team has demonstrated how to declare records and ensure that incoming data granules have a UUID assigned to them and a parent identifier to ensure they have a relationship with a parent collection level record. The use of OAIS data lifecycle tools and SI Object Lock has also been identified to support the requirement for assigning a retention schedule (as well as legal holds) to specific records and executing their disposition. Analysis of OSIM shows it is already capable of storing and using metadata concepts identified by the OAIS. NOAA Procedural Directives, multiple legacy NESDIS systems, and those proposed by the CP2P team (see Appendix B). The earlier NCAP cloud pilot also demonstrated how to present to the user a virtual AIP directory interface, to simulate traditional file-system organization patterns.	The presence of unique identifiers (UUIDs) should be declared on ingest and applied at that time if not present. While UUIDs can be applied at the time of initial cataloging by OSIM, it is generally preferred to apply them as early in the process as possible. Currently, most on-premise archival systems rely on directory structures and traditional file systems to support the function of organizing records. However, in cloud object storage, these types of hierarchies do not exist. Instead, an organizing file like the AIPS catalog with the use of OAIS-RC file should be implemented with, to support the active concept of archival package consisting of related data and documentation files.	Create a story or stories for augmenting the operation before ingest capability with the ability to check for the presence of a UUID and to create one if needed. Learning lessons from the NCAP cloud pilot, pilot the use of a catalog and/or OAIS-RC "AIP wrapper" file to ensure that the contents of an AIP remain in close association with one another in cloud object storage. Connect the various more archive-specific pieces (object locking, AIP creation) to augment the well-demonstrated existing cloud data ingest workflow. Implement the Collection Manager suite of tools in the NCAP following the methodology used to move production algorithms (containerize code and implement in cloud).
OAIS Ingest	Generate AIP	Data Prep	Associate Granule to Collection				
OAIS Ingest	Generate AIP	Archival Storage	Associate Granule with retention schedule				
OAIS Ingest	Generate AIP	Archival Storage	Apply object locking				
OAIS Ingest	Generate AIP	Data Prep	Create Archive Information Package				
OAIS Ingest	Generate Descriptive Information	Data Prep	Generate Collection Level Metadata	Generate Descriptive Information: extract Descriptive information from the AIPs and collect Descriptive Information from other sources to provide to Coordinate Updates, and ultimately Data Management.	Mature, main functionality operational in the NCAP, but newly demonstrated capabilities need to be operationalized. The CP2P team has already demonstrated how to extract and ingest the hierarchical object metadata, independent of data type or format, and catalog it within OSIM which is already operational in the NCAP. This demonstration included both the single submission and routine submission styles of archive ingest. Additionally, augmented metadata extraction routines have been demonstrated to work seamlessly with OSIM, either applied at time of ingest or retroactively.	No need in this functionality, as OSIM has been demonstrated to be capable of supporting both the workflow corrections and the management of a large amount of metadata content, including complete and hierarchical dataset-specific metadata in addition to standard metadata. Pieces of the backend workflow still need to be operationalized and a few minor additions to the granule-level metadata have been recommended (see Jira JPL31232 for details).	Enable indexing of the minor metadata additions by OSIM. Operationalize demonstrated components of the backend workflow. Provide clear guidance to the teams implementing granule metadata to incorporate in their workflows.
OAIS Ingest	Generate Descriptive Information	Data Prep	Generate Granular Level Metadata				

# NOAA is Moving its Archive to the Cloud

## NCEI's Proposed Cloud Archive Implementation Roadmap



## Tracking OAIS Functions to NCAP Features & Business Value

OAIS Functional Area	OAIS Function	Archive Workflow Function	Feature	MVP Prototype Status	MVP Completeness	Target PI	User Business WSJF Value
OAIS Ingest	Generate Descriptive Information	Data Prep	Generate Collection Level Metadata	Implemented-Workaround	100%	PI-1	4.33
OAIS Ingest	Generate Descriptive Information	Metadata Repository	Generate Provenance information	Complete	100%	PI-1	3.67
OAIS Ingest	Coordinate Updates	Archival Storage	Replicate Data Across Region	Complete	100%	PI-1	3.33
OAIS Data Management	Perform Queries	Data Stewardship	Search Collection Metadata Records on any or all searchab	Implemented-Workaround	100%	PI-1	3.33
OAIS Data Management	Perform Queries	Data Stewardship	Search Granule Metadata Records on all searchable attrib	Implemented-Workaround	100%	PI-1	3.33
OAIS Archival Storage	Manage Storage Hierarchy	Archival Storage	Lifecycle Policies for storage	Complete	100%	PI-1	2.67
OAIS Access	Deliver Response	Distribution & Access	Download data from tier 1 service	Implemented-Workaround	100%	PI-1	15.00
OAIS Access	Deliver Response	Distribution & Access	Visualize structured data	Implemented-Workaround	100%	PI-1	2.20
OAIS Ingest	Generate AIP	Archival Storage	Granule has a Unique Identifier	Complete	100%	PI-2	7.00
OAIS Ingest	Generate AIP	Data Prep	Associate Granule to Collection	Complete	100%	PI-2	7.00
OAIS Ingest	Generate AIP	Archival Storage	Associate Granule with retention schedule	Complete	100%	PI-2	1.50
OAIS Ingest	Generate AIP	Archival Storage	Apply object locking	Complete	100%	PI-2	5.50
OAIS Ingest	Generate Descriptive Information	Data Prep	Generate Granular Level Metadata	Complete	100%	PI-2	6.50
OAIS Ingest	Generate Descriptive Information	Data Prep	Generate fixdity information	Complete	100%	PI-2	3.50
OAIS Data Management	Receive Database Updates	Metadata Repository	Ingest Collection Level Metadata	Implemented-Workaround	100%	PI-2	4.00
OAIS Data Management	Receive Database Updates	Metadata Repository	Ingest Granular Level Metadata	Implemented-Workaround	100%	PI-2	2.40
OAIS Data Management	Perform Queries	Data Stewardship	Search collection for granule	Complete	100%	PI-2	3.00
OAIS Access	Deliver Response	Distribution & Access	Subsetting of structured data	Implemented-Workaround	100%	PI-2	1.75
OAIS Access	Deliver Response	Distribution & Access	Data output transformer	Implemented-Workaround	100%	PI-2	2.60
OAIS Administration	Manage System Configuration	Data Stewardship	Job Orchestration	Complete	100%	PI-2	0.71
OAIS Administration	Customer Service	Data Stewardship	Data Use Monitoring/Administration/Reporting	Complete	100%	PI-2	1.15
OAIS Ingest	Generate AIP	Data Prep	Create Archive Information Package	Complete	100%	PI-3	2.88
OAIS Data Management	Generate Report	Data Stewardship	Storage Dashboard	Complete	100%	PI-3	2.50
OAIS Data Management	Generate Report	Data Stewardship	Collection Dashboard	Complete	100%	PI-3	1.13
OAIS Data Management	Generate Report	Data Stewardship	Granule Dashboard	Complete	100%	PI-3	1.23
OAIS Data Management	Generate Report	Data Stewardship	System monitoring report	Implemented-Workaround	100%	PI-3	2.38
OAIS Data Management	Generate Database Updates	Metadata Repository	Enrichment Metadata	Planned	33%	PI-3	1.13
OAIS Archival Storage	Provide Data	Archival Storage	Move AIPs from Preservation Storage to Access Storage	Complete	100%	PI-3	0.60
OAIS Access	Deliver Response	Distribution & Access	Archive Order response	Started	00%	PI-3	1.00

## Tracking Implementation of OAIS Functions

OAIS Functional Area	OAIS Function	Feature	MVP Completeness	T2O Readiness	T2O Priority (1 = top)	Dev Tier Complete	Ops Tier Complete
OAIS Ingest	Generate Descriptive Information	Generate Provenance information	100%	66%	3		
OAIS Data Management	Receive Database Updates	Ingest Collection Level Metadata	100%	33%	2		
OAIS Data Management	Receive Database Updates	Ingest Granular Level Metadata	100%	33%	2		
OAIS Data Management	Receive Database Updates	Ingest Enrichment Metadata	33%	0%	3		
OAIS Access	Deliver Response	Download data from tier 1 service	100%	33%	1		
OAIS Access	Deliver Response	Visualize structured data	100%	33%	3		
OAIS Access	Deliver Response	Subsetting of structured data	100%	33%	3		
OAIS Access	Deliver Response	Data output transformer	100%	33%	1		
OAIS Access	Deliver Response	Adhoc Order processor	90%	0%	3		
OAIS Access	Deliver Response	Subscription Order processor	66%	0%	2		
OAIS Access	Deliver Response	Bulk order order processor	100%	0%	3		
OAIS Access	Deliver Response	Order certification service	90%	33%	3		
OAIS Access	Deliver Response	A/A for restricted data access	100%	33%	2		
OAIS Data Management	Generate Report	Storage Dashboard	100%	33%	1		
OAIS Data Management	Generate Report	Collection Dashboard	100%	0%	1		
OAIS Data Management	Generate Report	Granule Dashboard	100%	0%	1		
OAIS Data Management	Generate Report	System monitoring report	100%	0%	1		
OAIS Data Management	Perform Queries	Search Collection Metadata Records on any or all searchable attributes	100%	33%	3		
OAIS Data Management	Perform Queries	Search granule metadata records on all searchable attributes	100%	33%	3		
OAIS Administration	Archival Information Update	Update a granule object already in the AIP	100%	33%	1		
OAIS Administration	Archival Information Update	Add a new granule to an existing AIP	100%	33%	1		
OAIS Administration	Archival Information Update	Remove granule object from the AIP	100%	33%	1		
OAIS Administration	Manage System Configuration	Job Orchestration	100%	100%	3		
OAIS Administration	Manage System Configuration	Automation CI/CD	100%	66%	1		
OAIS Administration	Customer Service	Data Use Monitoring/Administration/Reporting	100%	66%	3		

Features Categorized by Archive Workflow Function

Metadata Repository	4
Distribution & Access	9
Data Stewardship	13
Data Prep	5
Archival Storage	6

## Detailed Tracking of Metadata Features

Feature	In MVP	Status	PI	Notes	Metadata Feature	CMR Adoption (PI 1-3)
Generate Replicated URI	No	TBD	PI-5*	Removed from MVP additional scope that was not part of the initial MVP.	Yes	0
Create Archive Information Package	Yes	Complete	PI-3		Yes	0
Update a granule object already in the AIP	Yes	Planned	PI-3		Yes	0
Add a new granule to an existing AIP	Yes	Complete	PI-3		Yes	0
Remove granule object from the AIP	Yes	Complete	PI-3		Yes	4
Ingest Enrichment Metadata	Yes	Planned	PI-3		Yes	2
Granule has a Unique Identifier	Yes	Complete	PI-2		Yes	4
Associate Granule to Collection	Yes	Complete	PI-2		Yes	3
Generate Granular Level Metadata	Yes	Complete	PI-2	The technical solution has required a UMM-G model to be adopted for generating granule metadata. The use of ISO granule metadata required us to step down our version of ISO from 19115-2 --> 19115-1.	Yes	3
Ingest Collection Level Metadata	Yes	Implemented-Workaround	PI-2	Ingested collection level metadata into CMR, however, at present only 16% of attributes weren't able to map to UMM.	Yes	3
Generate fixdity information	Yes	Complete	PI-2		Yes	3
Search collection for granule	Yes	Complete	PI-2		Yes	4
Ingest Granular Level Metadata	Yes	Implemented-Workaround	PI-2		Yes	3
Associate Granule with retention schedule	Yes	Complete	PI-2	Possibly only need to do this at the Collection level?	Yes	0
Generate Collection Level Metadata	Yes	Implemented-Workaround	PI-1	Questioning the how to track the complexity of the implementation given that we've accomplished an initial load of a collection record that is discoverable in the UI.	Yes	
Generate Provenance information	Yes	Complete	PI-1		Yes	

Key

Key	Definition
0	Unknown
1	Cannot work with CMR
2	Internally works with CMR
3	Partially works with CMR
4	Identify works with CMR
5	Partially works with CMR
6	Not actually a metadata feature in CMR
7	Metadata feature of CMR

20 Total Features  
16 Features in MVP  
16 Features in MVP That Have Been Implemented