n p r

IO: NPR's Enterprise Asset Management System

Storage Abstraction - Workflow Orchestration   - Long Term Preservation

David Julian Gray: Sr. Enterprise Architect;   Cory Aiken: Director Software Engineering
Joe Kampschmidt: Sr. Software Architect;   Surya Mudiya: Software Development Manager
Shahryar Arvin: Sr. Software Architect;   Julie Rogers: Curator/Historian

Library of Congress
Designing Storage Architectures for Digital Collections
March 15, 2022

Overview

The problem

- *Where's my stuff?*
- *Digital media evanescent*
- *Archival format is "constant migration"*

Solution - Data/storage abstraction

- *Build vs buy?*

Technical Overview of IO - NPR: "How I built this"

npr

NPR IO: *Problems to Resolve*

- *Ensure File Integrity*
  - Find it *now* - Find it at *any arbitrary later date*
- *Streamline and lubricate workflows* :
  - Integrate disparate systems
  - Eliminate siloed storage and content management
  - Incentivize compliance
- *Contain costs*
  - Provide usage and capacity analytics
  - Move content without breaking relationships

npr

What's the Bright Idea?
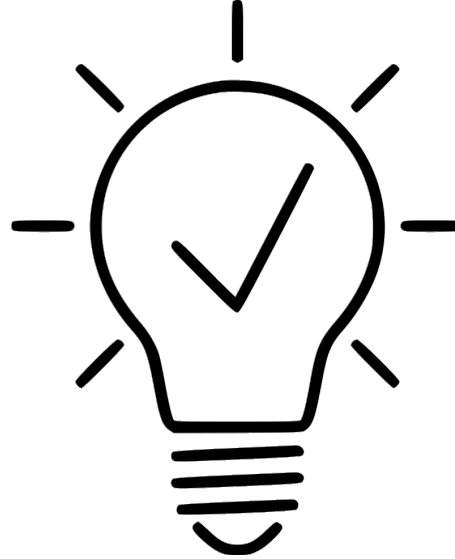
Data abstraction!

"Storage Broker"

- **Intermediate access layer pointing to perennial GUID and link**
- **Media can move among and between storage as performance and cost containment require without breaking content management and production tools.**

Federated search engine

This Cloud

On Prem Here

On Prem There

That Cloud

Storage "Broker"

2018 Blue Sky Solution

NewsFlex API

SEAMUS API

Artemis API

Content Service

Federated Content Browser & Ingest formatter

Edit Chosen Asset

Processing Engine

etc
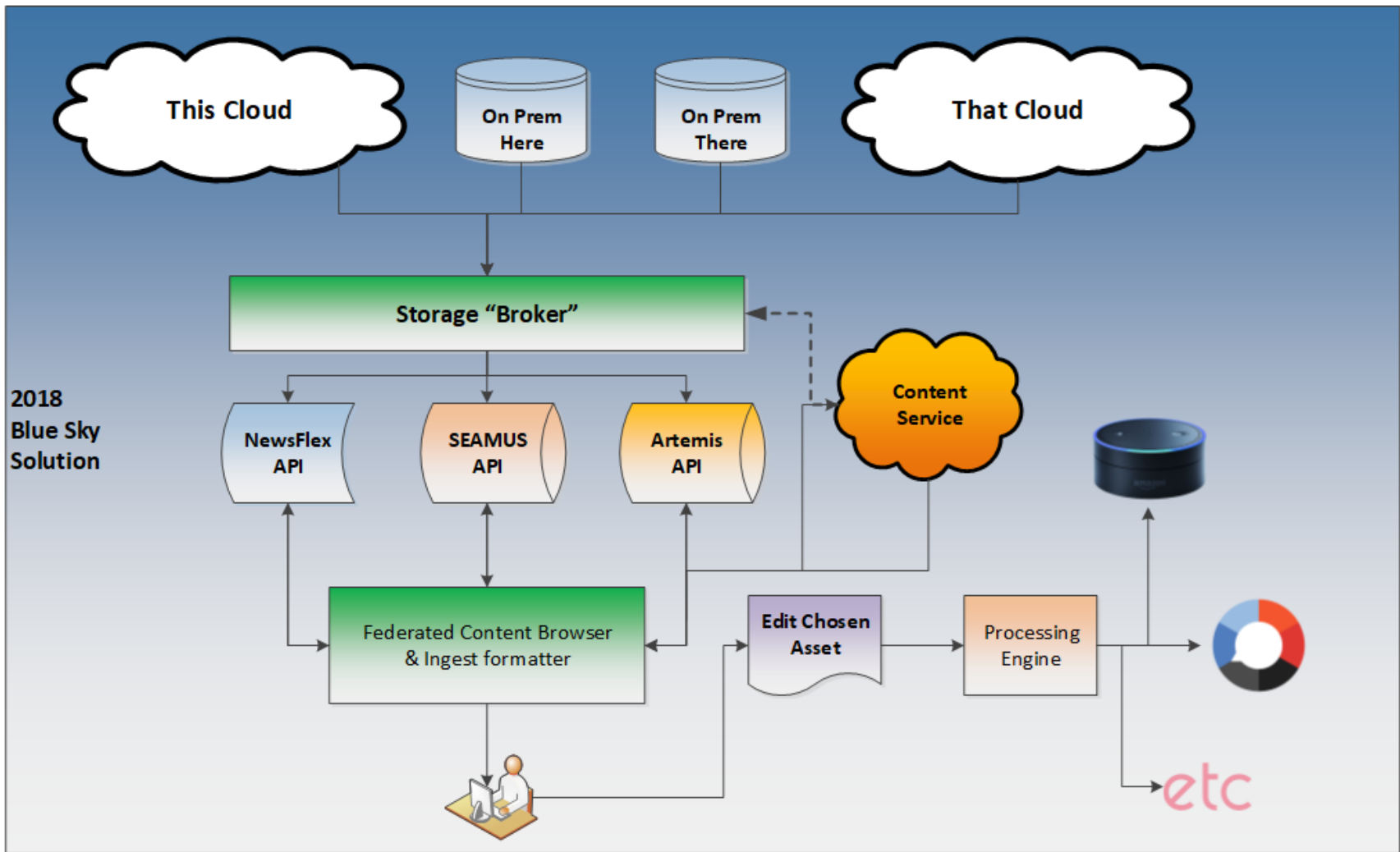
NPR IO

- Enterprise Media Asset Management

- Content Repository

- Workflow Orchestrator

- Asset Repurposing Engine

- Preservation Asset Arbitrator

- Federated Search Platform

NPR IO - Key concepts and components

- **Storage "Buckets"**
  - Mapped to logical units
  - Can be "on-prem" or in cloud
  - abstracted from client usage
- **WENDU**
  - Workflow Orchestrator
  - Business Rules Engine
  - Enterprise Service Bus
- **Federated Search Platform**

BUCKETS!

Ingest Buckets

- DC Ingest
- LA Ingest
- NY Ingest
- Video Ingest
- Live Recordings
- Perishable

Working Buckets

- Video Editing
- Audio Editing
- General Editing (Other)
- Local Editing

Archive Buckets (Read-Only)

Tier 1
- Perennials
- Ready for Presentation
- External MAM (e.g. JW)

Tier 2
- Object Storage
- Preserve

Tier 3
- Glacier (S3)
- Tape?
- ...

Published Buckets

Tier 1
- Recently Published

Tier 2
- Published this Year
- Published Archive

Tier 3

Labels show basic concepts - not all buckets have yet been implemented or used at this time

IO Object Model

Storage
- Storage Type
- Storage Tier

Key
- Top level classes
- Attribute classes
- Enum types

* attributes that can be treated as a tag

Bucket
- id
- title
- type
- category*
- dataAccess
- region*
- owner*
- tier*
- maxFiles
- maxSize
- allowedTypes*

Tag
- key
- value
- type

<<enumeration>> Category
- Ingest
- Working
- Archive
- Published

Region
- id
- name
- TBD

Policy
- applyForAll
- assetFilters
- frequency
- owner (Principle)
- Replication
- Fixity
- etc.

Schedule
- TBD

Asset
- id
- filename
- filePath
- role
- title
- type
- subType
- state
- checksum
- tags
- technicalData
- externalData
- versions
- relatives
- bucket
- timeInBucket
- perennial*

AssetRelation
- id
- relationType
- relatedAsset

<<enumeration>> AssetType
- Audio
- Video
- Image
- Document
- Other

<<enumeration>> AssetRole
- master
- derivative
- clone
- linked

Asset's role controls which properties are required vs optional or not valid.

AssetVersion
- id
- sequence
- asset

<<enumeration>> AssetState
- invalid
- pendingIngest
- ingesting
- locked
- checkedOut
- archived
- deleted
- purged

<<enumeration>> AssetRelationEnum
- descendant
- derivative
- linked
- working
- clone

9

WENDU

- Workflow Orchestration

- Business Rules Engine

- Enterprise Service Bus

WENDU
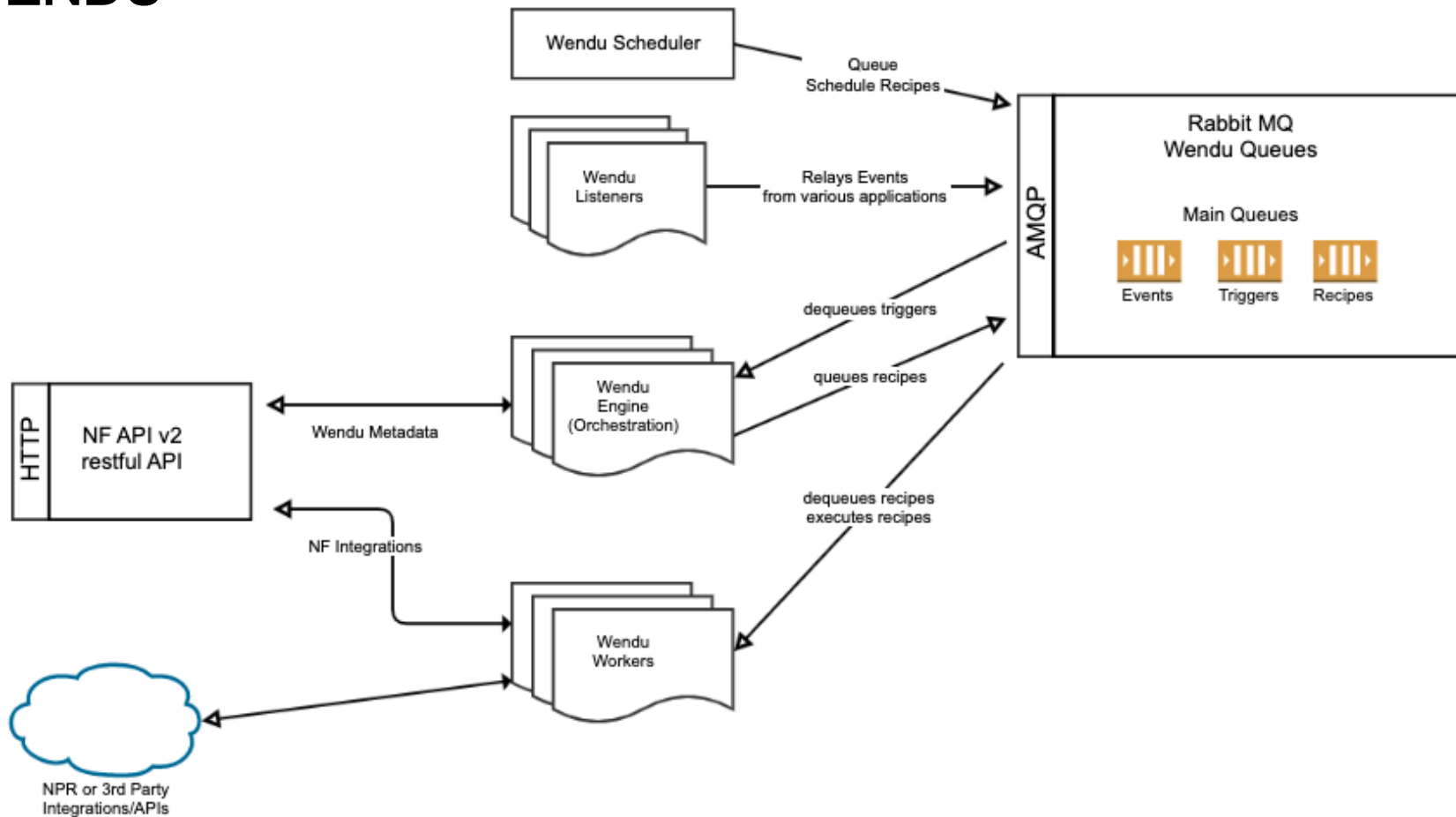
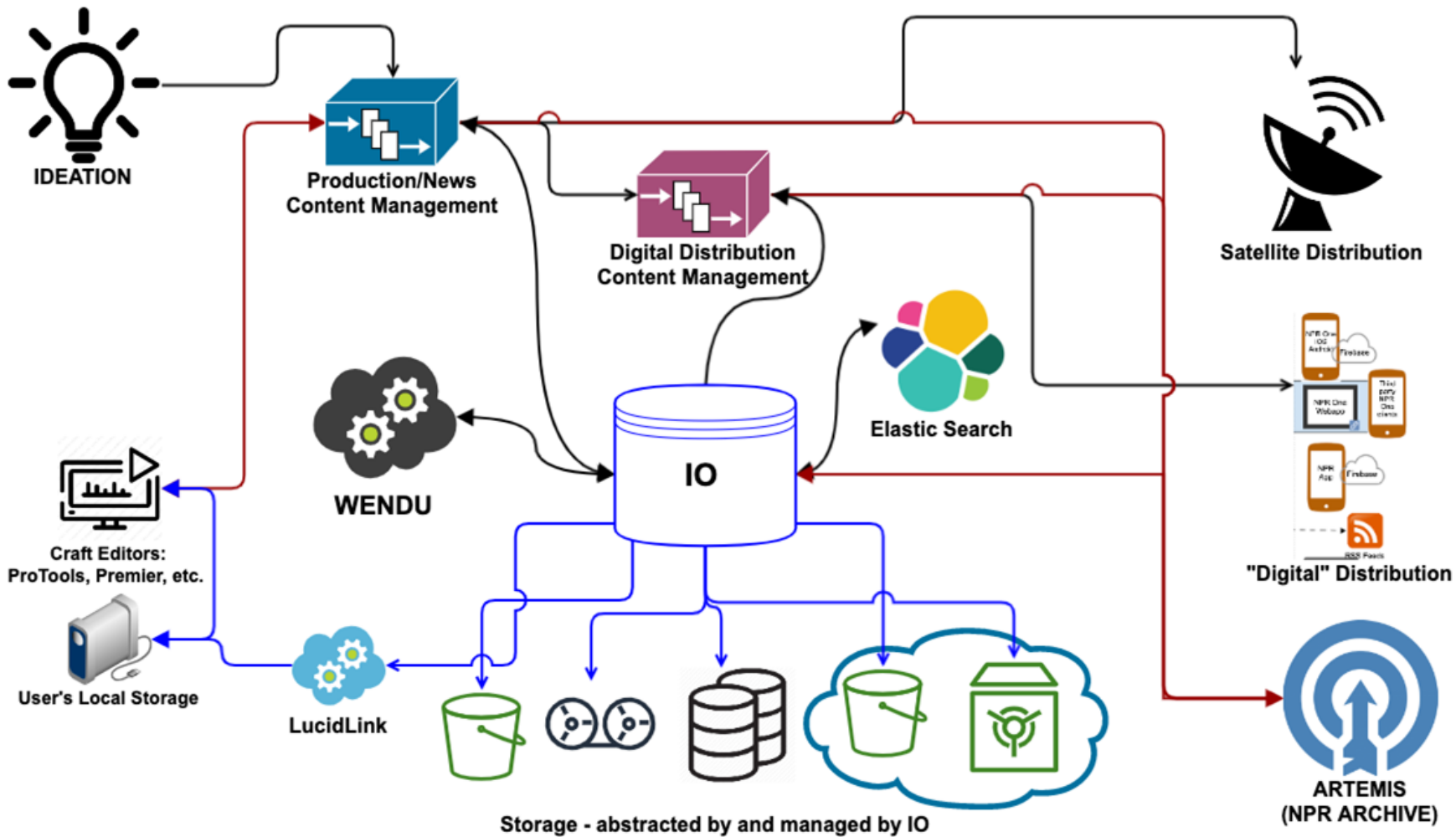# Long-Term Preservation

- NPR's Research, Archives & Data Strategy team (RAD)

- Historical Storage

- Digitization and Reformatting

IO

- **Frictionless workflows**

- **Enterprise lifecycle management**

- **Long-term media preservation**

IDEATION

Production/News Content Management

Digital Distribution Content Management

Satellite Distribution

Elastic Search

WENDU

IO

Craft Editors: ProTools, Premier, etc.

User's Local Storage

LucidLink

"Digital" Distribution

ARTEMIS (NPR ARCHIVE)

Storage - abstracted by and managed by IO

14

QUESTIONS?