

Library eArchiving with ZONTAL Space and the Allotrope Data Format

Dennis Della Corte^{1,2,}, Wolfgang Colsman³, Ben Welker⁴, Brian Rennick⁵*

¹Brigham Young University, Department of Physics and Astronomy

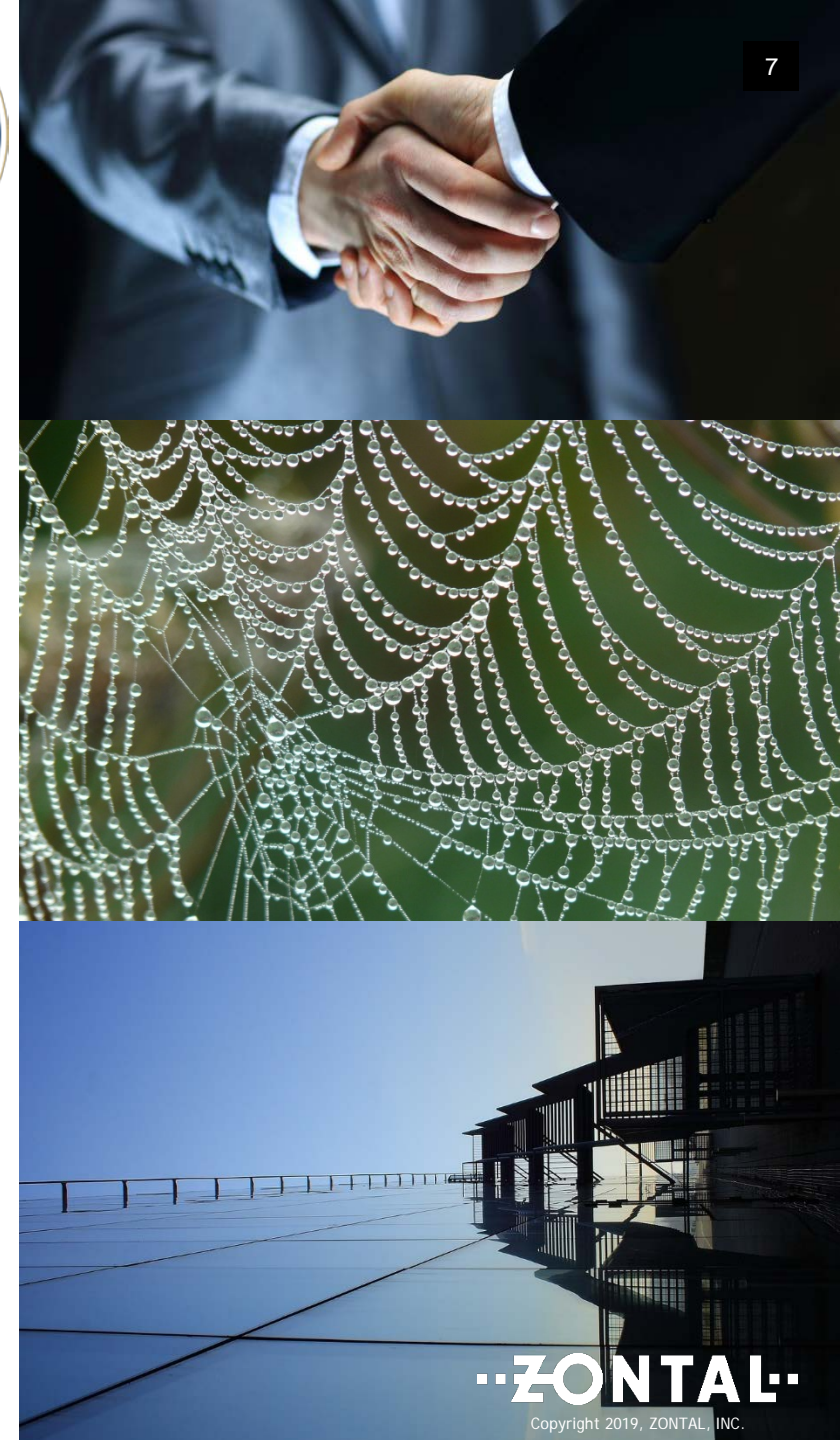
²ZONTAL, inc Vice President, Marketing & Sales

³ZONTAL, inc Chief Executive Officer

⁴Brigham Young University, Library, Associate University Librarian for
Library IT

⁵Brigham Young University, Library, Software Engineer

Currently under peer review



Presentation Outcomes

- Gain better understanding of Allotrope business benefits outside of pharma
- See differences an Allotrope based information management solution can make
- Get you excited about exploring how ZONTAL Space can benefit your business

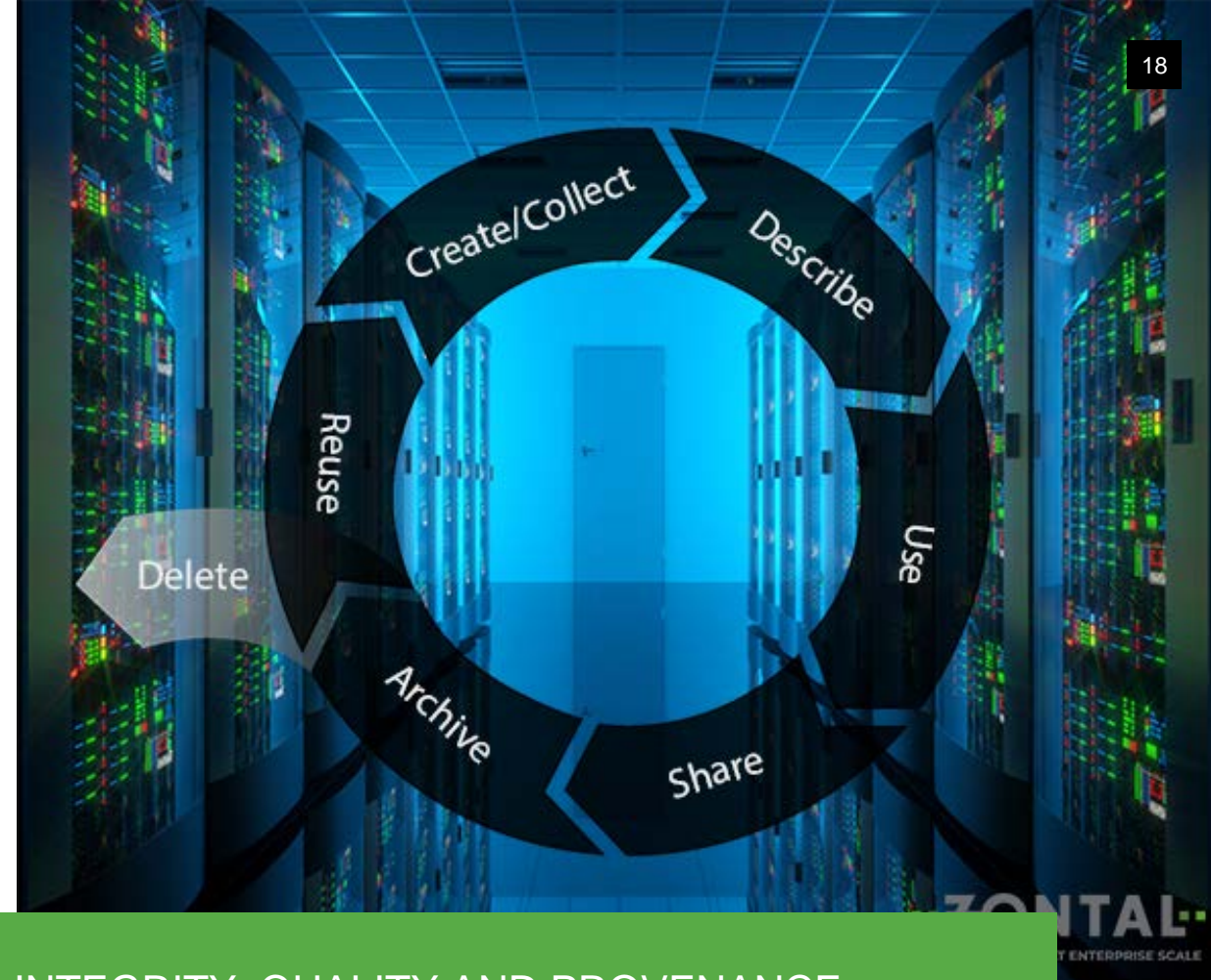
DATA INNOVATION AT ENTERPRISE SCALE

What is ZONTAL Space all about?

INFORMATION LIFECYCLE MANAGEMENT

Build Trust to Maximize Data Reuse

- Data ownership secures data value over time
- Data governance manages documentation standards, ensures interoperability and simplifies collaboration
- Data quality management maximizes data value and compliance
- Data provenance builds trust and encourages reuse
- Configurable workflows allow implementing business specific requirements
- Automated retention schedules minimize efforts of archive management



DATA INTEGRITY, QUALITY AND PROVENANCE

Assessing data for its trustworthiness becomes important in data reuse with the growth in data creation because of the lack of standards for ensuring data quality and potential harm from using poor-quality data.

DATA INNOVATION AT ENTERPRISE SCALE

BUILD ON OPEN STANDARDS

“Drive” Innovation and Sustainable Development

- Allotrope Data Format, GxP compliant Data Container (ADF)
- ISO 14721: Reference Model for Open Archival Information Systems (OAIS)
- ISO 16363: Trustworthy Repositories Audit and Certification (TRAC)
- EMA “Best Archiving Practice” (BAP)
- Public Ontologies (allotrope, prov, foaf, dc terms, qudt, ...)
- S3 compatible storage (Amazon S3, Dell ECS, Hitachi HCP, ...)

DATA INTEROPERABILITY

The seamless, secure and controlled exchange of data is a serious challenge in maximizing the potential of the many digital tools and resources.

The lack of interoperability means that a complete record often will not be available, compromising compliance and inhibiting innovation.

ADF is superior to BagIt format

5

Functionality	Required by	BagIt	ADF
Content Data Object	OAIS	Stored in native format only	Stored in native format or converted to Data Cube
Representation Information	OAIS	Optional: Metadata as key/value pair	Fully open standards: RDF for metadata, HDF5 for Data Cube
Reference information	OAIS	Optional: Metadata as key/value pair	Optional: Semantic metadata according to RDF.
Provenance Information	OAIS	Not included	Audit trails in Data Provenance as RDF metadata
Fixity information	OAIS	Checksums stored in manifests	Checksums are dynamically created and verified
Hierarchical file packaging	OAIS	“Payload”	Data Package Layer
Semantic MetaData	FAIR	Not included	Data Description Layer
Interoperability between user groups	FAIR	Limited by key/value metadata pairs	Semantic connection of user groups enables flexible metadata enrichment
Vendor agnostic scientific format	FAIR	Not included	Data Cubes
Migration: Conversion between BagIt and ADF	FAIR	BagIt -> ADF possible	ADF -> BagIt possible, but semantic information will get lost

Ontology Mapping

- Dublin Core ontology is established standard for library application
- Current IT systems do not enforce or check correct usage of ontology

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```

▼ <xml>
  <creato/>
  <contri/>
  <title>[John David Smith?] letter to John Smith</title>
  <altern/>
  ▼ <descri>
    Letter from "Jod", possibly John David Smith, to his father, John Smith. Regarding health of the family and thanks for presents wh
    October 7, 1884.
  </descri>
  <date>1884 October 7</date>
  <datea>1884</datea>
  <editio>Electronic version</editio>
  <publis/>
  <physic>1 folder</physic>
  <scale/>
  <extent>1 folder</extent>
  <medium/>
  <dimens/>
  <arrang/>
  <degree/>
  ► <contra>...</contra>
  <names/>
  <topics/>
  <geogra/>
  <genre>Letters</genre>
  <occupa/>
  <functi/>
  <subjec/>
  <langua>English; eng; en</langua>

```

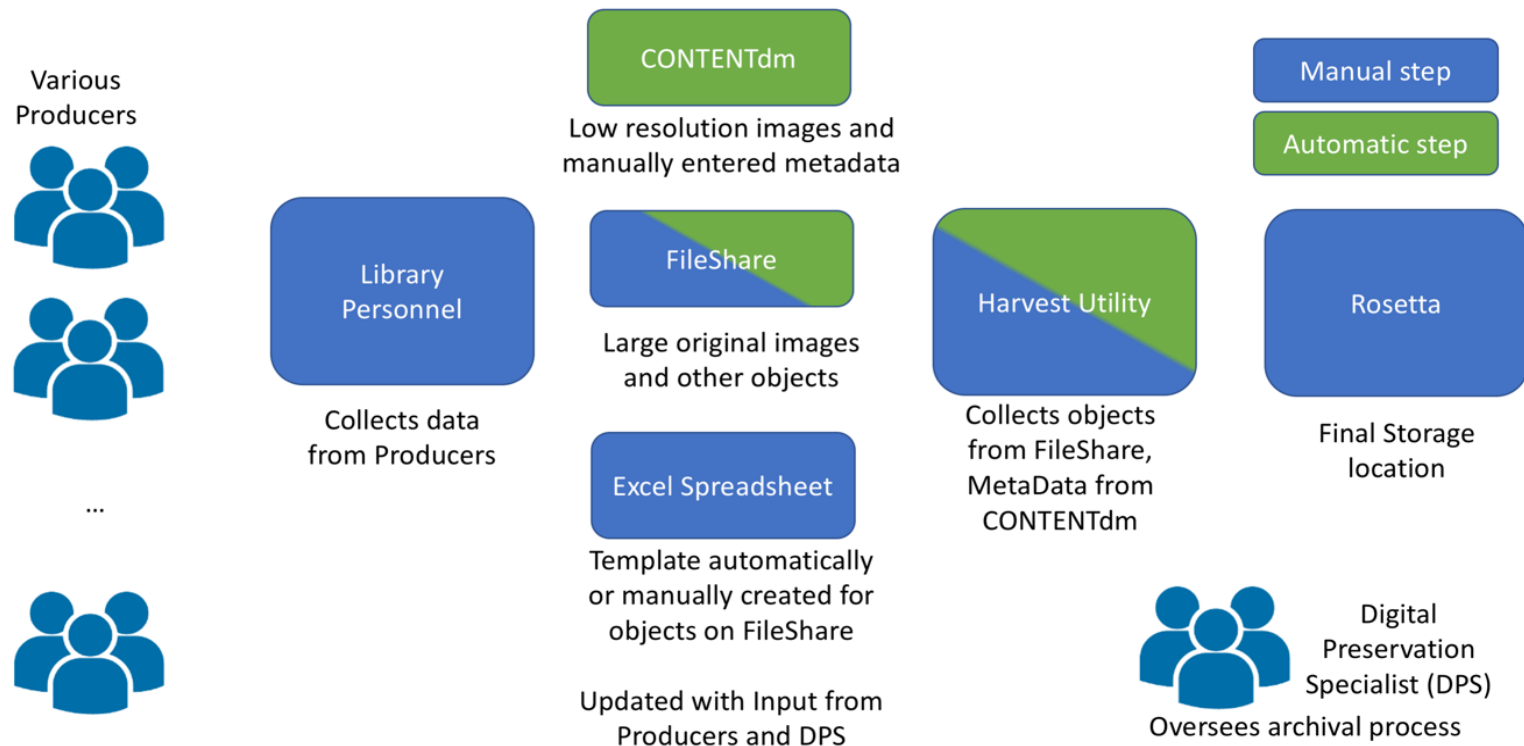
ZONTAL Space automatically maps XML files containing metadata to the correct ADF and stores them as RDF triples.

Dublin Core Terms

dc:title
 dcterms:created
 dc:date
 dc:coverage
 dcterms:extent
 dcterms:rightsHolder
 dc:type
 dc:language
 dc:relation
 dcterms:bibliographicCitation
 dc:identifier
 dc:rights
 dcterms:license
 dcterms:rightsHolder
 dcterms:accessRights
 dc:publisher
 dc:format
 dc:description
 dcterms:available
 dcterms:isPartOf

BYU Pre-PoC Workflow

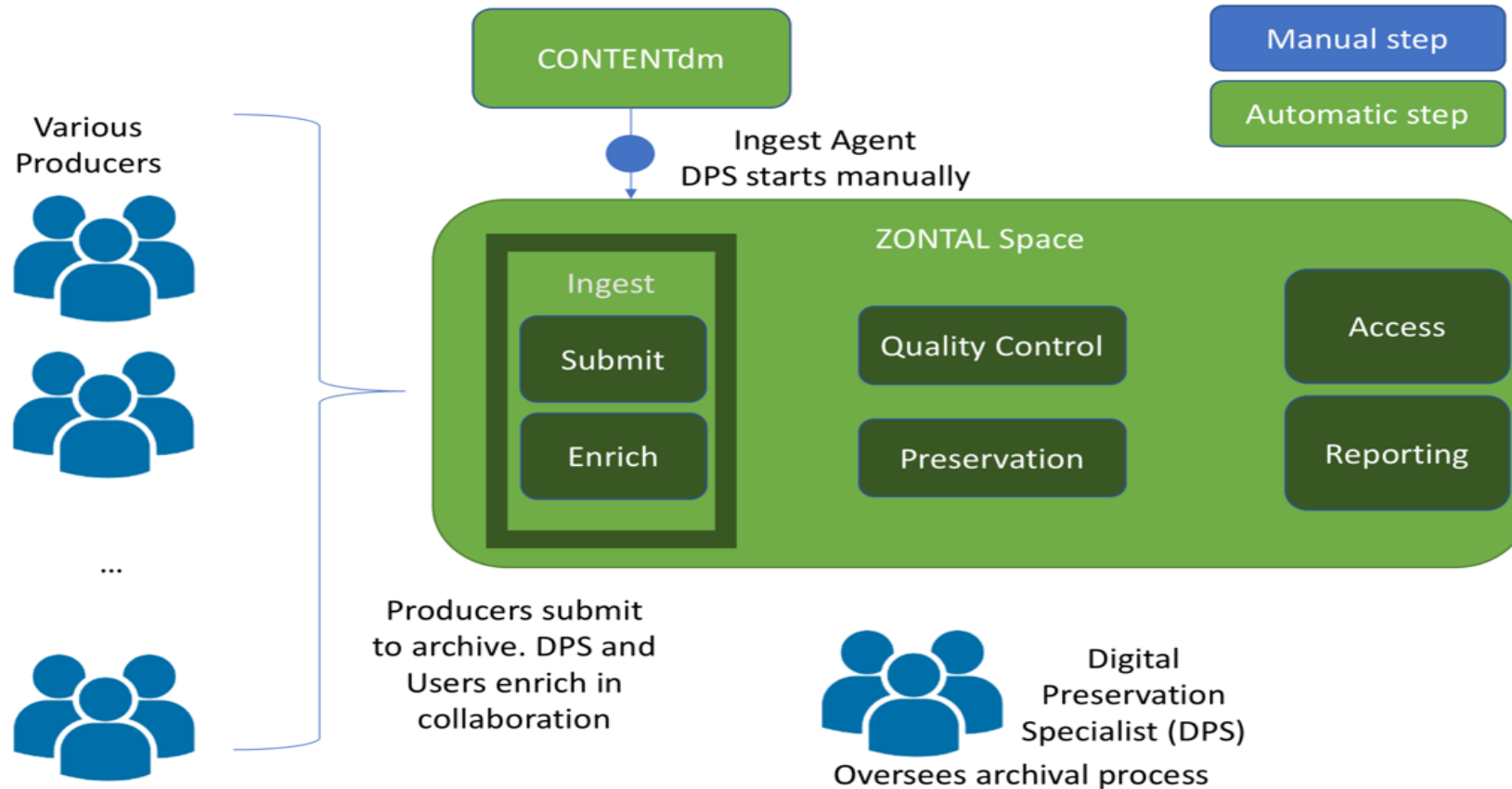
Many manual steps and multiple resources are required



BYU Post-PoC Workflow

8

Nearly automatic data preservation with minimal resource overhead





DATA INNOVATION AT ENTERPRISE SCALE

Why does this work?

Standardized APIs

Secure and Automate Access to Metadata and Content

- Web service APIs integrate into the existing application landscape
- Agents integrate laboratory instruments to improve data integrity
- Allotrope Data Format (ADF) files transfer raw data, contextual metadata, audit trails and checksums in a vendor neutral format



ARTIFICIAL INTELLIGENCE AND ROBOTIC PROCESS AUTOMATION

80 percent of the work data scientists perform today, according to this [*Forbes Magazine*](#) report, can be automated. This puts great pressure on companies to FAIRify their data today, so that automated systems may be of the highest possible quality and performance.

PoC: One manual ingest command bulk imported hundreds of records and automatically enriched them with metadata

The screenshot displays the Zontal Space interface. On the left, a sidebar contains navigation icons. The main area shows a table of records with columns: Lifecycle Status, Preferred Label, Created On, Information Pack..., Imported from Host, and Imported from URL. The table lists 29 results, with 1 selected. The selected record is 'p15999coll31-1075', which is 'Archived' and has a 'sample' label. To the right of the table, a detailed view for the selected record is shown, including Dublin Core, Reference Properties, File Properties, and Provenance information.

Lifecycle Status	Preferred Label	Created On	Information Pack...	Imported from Host	Imported from URL
Submitted	sample	2019/Aug/29 08:4...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-1075	2019/Sep/23 15:3...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-1062	2019/Sep/23 15:5...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-1041	2019/Sep/23 15:3...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-1016	2019/Sep/23 15:3...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-721	2019/Sep/23 15:5...	BYU Dublin Core	unx10080	/opt/www/webdav...
Submitted	sample3	2019/Aug/27 09:1...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-1067	2019/Sep/23 10:5...	BYU Dublin Core ...	unx10080	/opt/www/webdav...
Archived	p15999coll31-955	2019/Sep/23 15:5...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-727	2019/Sep/23 15:3...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-989	2019/Sep/23 15:5...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-1075	2019/Sep/23 15:2...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-731	2019/Sep/23 15:5...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-966	2019/Sep/23 15:5...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	Isaac Brockbank Jr....	2019/Jul/15 10:49...	BYU Dublin Core A...		
Archived	p15999coll31-1012	2019/Sep/23 15:5...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-1067	2019/Sep/23 15:3...	BYU Dublin Core	unx10080	/opt/www/webdav...
Archived	p15999coll31-1083	2019/Sep/23 15:3...	BYU Dublin Core	unx10080	/opt/www/webdav...

p15999coll31-1075

Dublin Core

Title: p15999coll31-1075.xml
 Date: 1911/Jul/22
 Type: Postcards
 Language: English
 Rights: http://lib.byu.edu/about/copyright/special_collections.php
 Access Rights: Public
 Format: text/jpg
 Description: Postcard from a Jeanetta with a picture of Yates and Douglas Streets in Victoria, British Columbia, Canada. Text on the card regarding journey from Seattle to British Columbia. Dated July 11, 1922.
 Date Available: 1911/Jan/01

Reference Properties

ADF Identifier: 789c96fd-af5a-4e26-8649-a9aa3078aa85
 Version Label: 4, current

File Properties

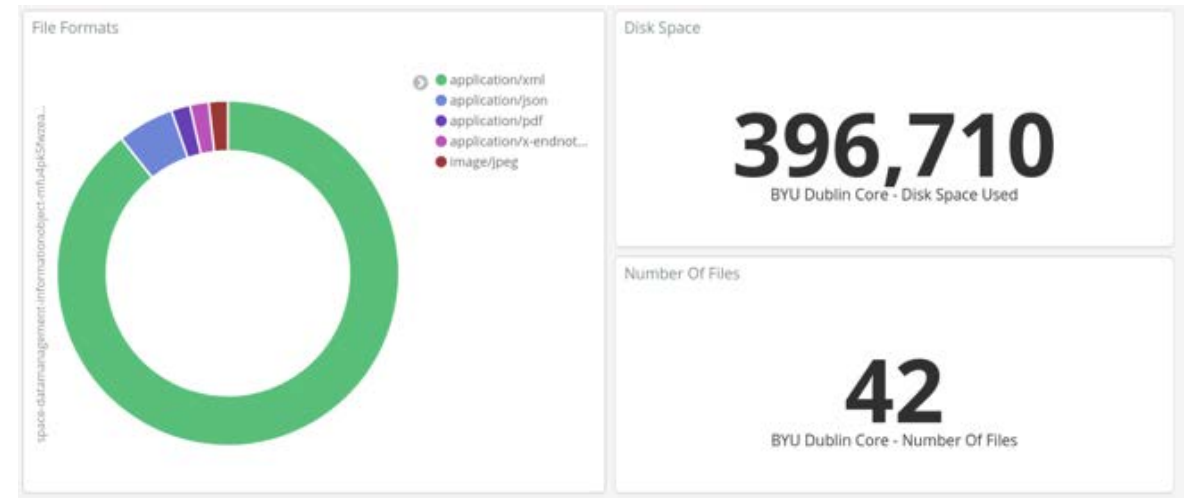
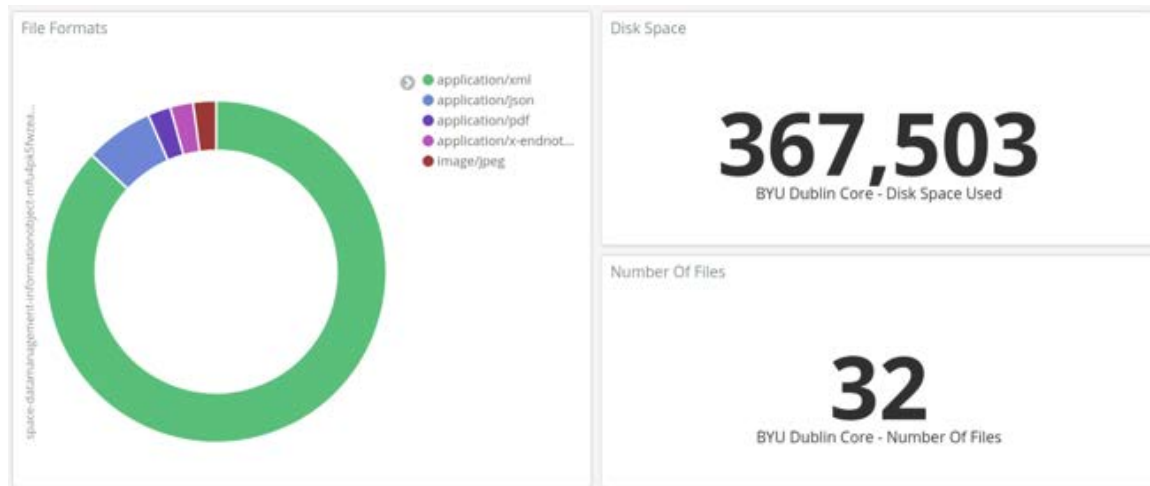
Data Size: 2,983
 Number of Files: 1

Provenance

Content Identifier: FBEADEC02FC1961F188C1326DD2F5D2D
 Information Source: BYU
 Created On (Information Source): 2019/Sep/23 15:33:00 (-06:00)
 Modified On (Information Source): 2019/Sep/23 15:33:00 (-06:00)

Little features with big impact

Dashboards monitor disk space usage per collection – admin can now monitor without infrastructure support



Connecting Data in Archives to Data Analytics

Capture

Reference & Masterdata Management

Web Clients

File Shares

CDS

LIMS

ELN

DMS

SDMS

Ingest



Manage Information Lifecycle

The screenshot shows the Zontal Space web interface. At the top, there's a search bar with 'HPL' entered. Below it, a table lists analytes with columns for Lifecycle Status, Analyte name, and Imported from. A search dropdown is open, showing a list of analytes: Aspirin, Cholecalciferol, Glucose, Glyphosate, Ibuprofen, and Norfloxacin. The table below shows results for 'Pistoia Meth...' and 'HPLC Method'.

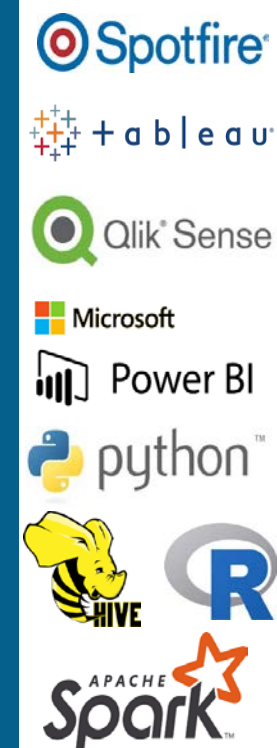
Lifecycle Status	Analyte name	Imported from
Archived	Pistoia Meth...	HPLC Method
Submitted	Pistoia Meth...	HPLC Method
Archived	Pistoia Meth...	HPLC Method
Submitted	Pistoia Meth...	HPLC Method
Submitted	Pistoia Meth...	HPLC Method
Archived	Pistoia Meth...	HPLC Method

At the bottom, there are logos for various cloud providers and storage solutions: Amazon S3, Azure, GCP, NAS, MINIO, and HITACHI.

Access



Reuse



What can you do with data in ZONTAL Space?

Next steps for library use case:

Enhance existing recommender systems to leverage descriptive information.