ZONTAL
DATA INNOVATION AT ENTERPRISE SCALE

Library eArchiving with ZONTAL Space and the Allotrope Data Format

***Dennis Della Corte*[1,2,*], *Wolfgang Colsman*[3], *Ben Welker*[4], *Brian Rennick*[5]**

[1]Brigham Young University, Department of Physics and Astronomy
[2]ZONTAL, inc Vice President, Marketing & Sales
[3]ZONTAL, inc Chief Executive Officer
[4]Brigham Young University, Library, Associate University Librarian for Library IT
[5]Brigham Young University, Library, Software Engineer

Currently under peer review

ZONTAL

Presentation Outcomes

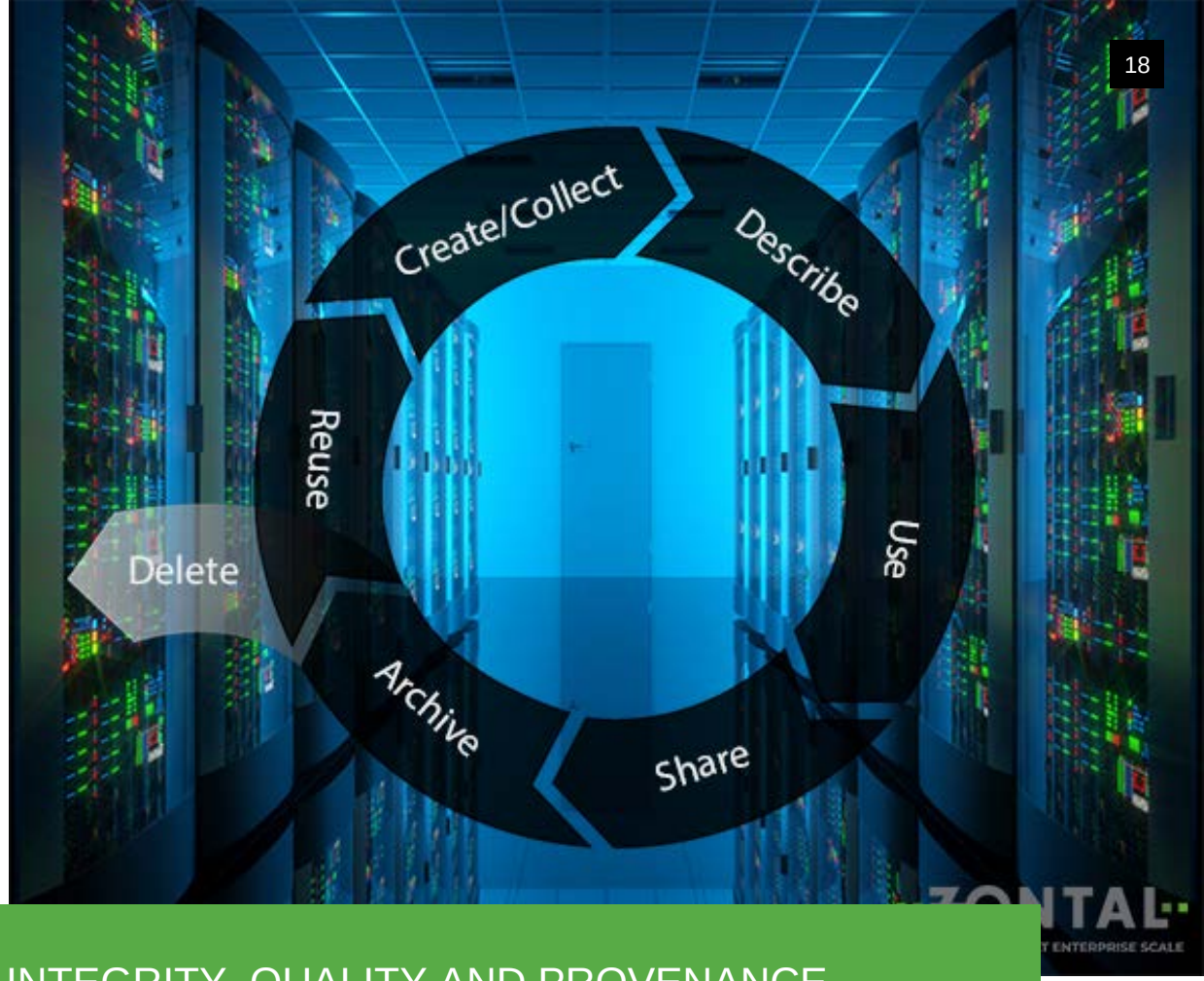
- Gain better understanding of Allotrope business benefits outside of pharma
- See differences an Allotrope based information management solution can make
- Get you excited about exploring how ZONTAL Space can benefit your business

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 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.

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, …)

ZONTAL SPACE

ADF is superior to BagIt format

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>
```

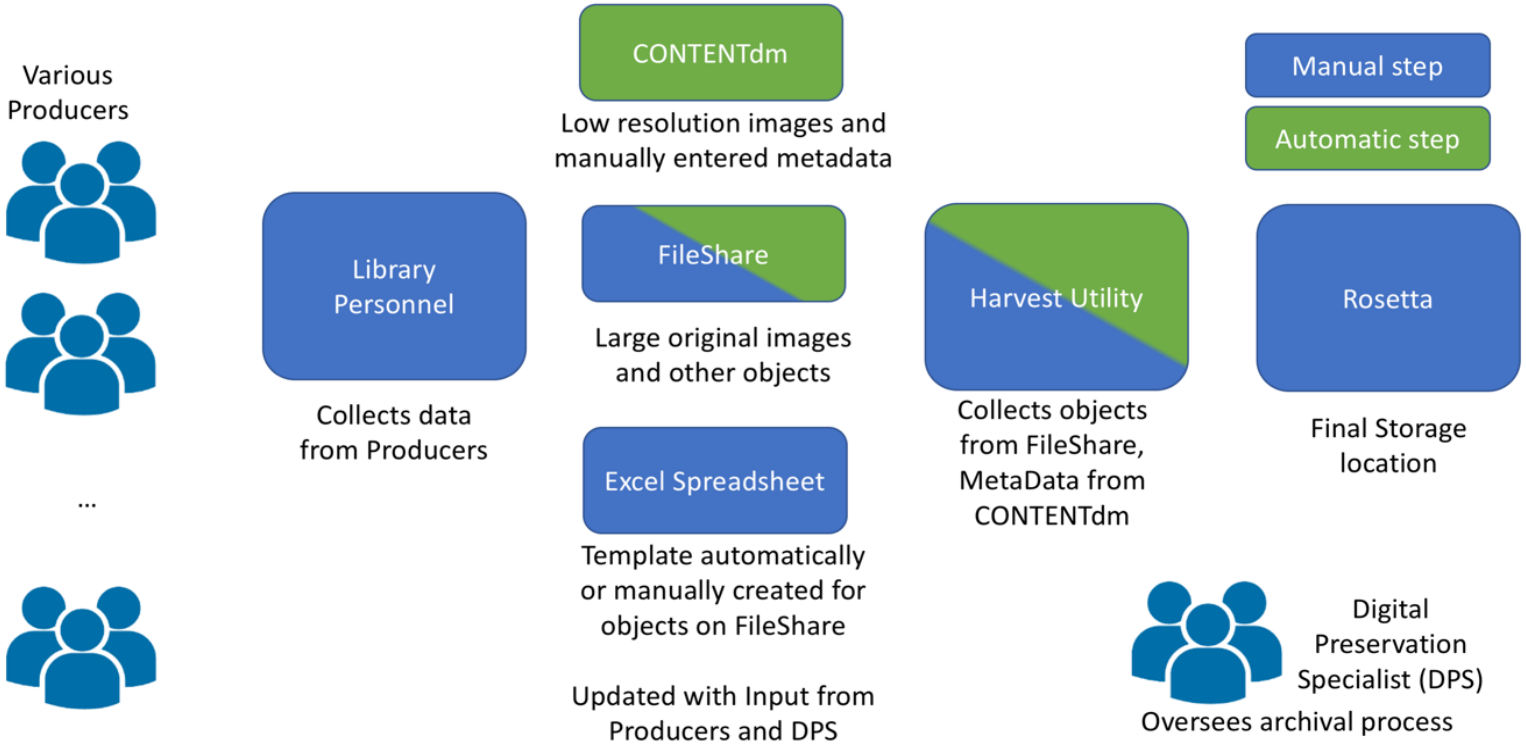
ZONTAL Space automatically maps XML files containing metadata to the correct ADF and stores them as RDF triples.

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

ZONTAL SPACE

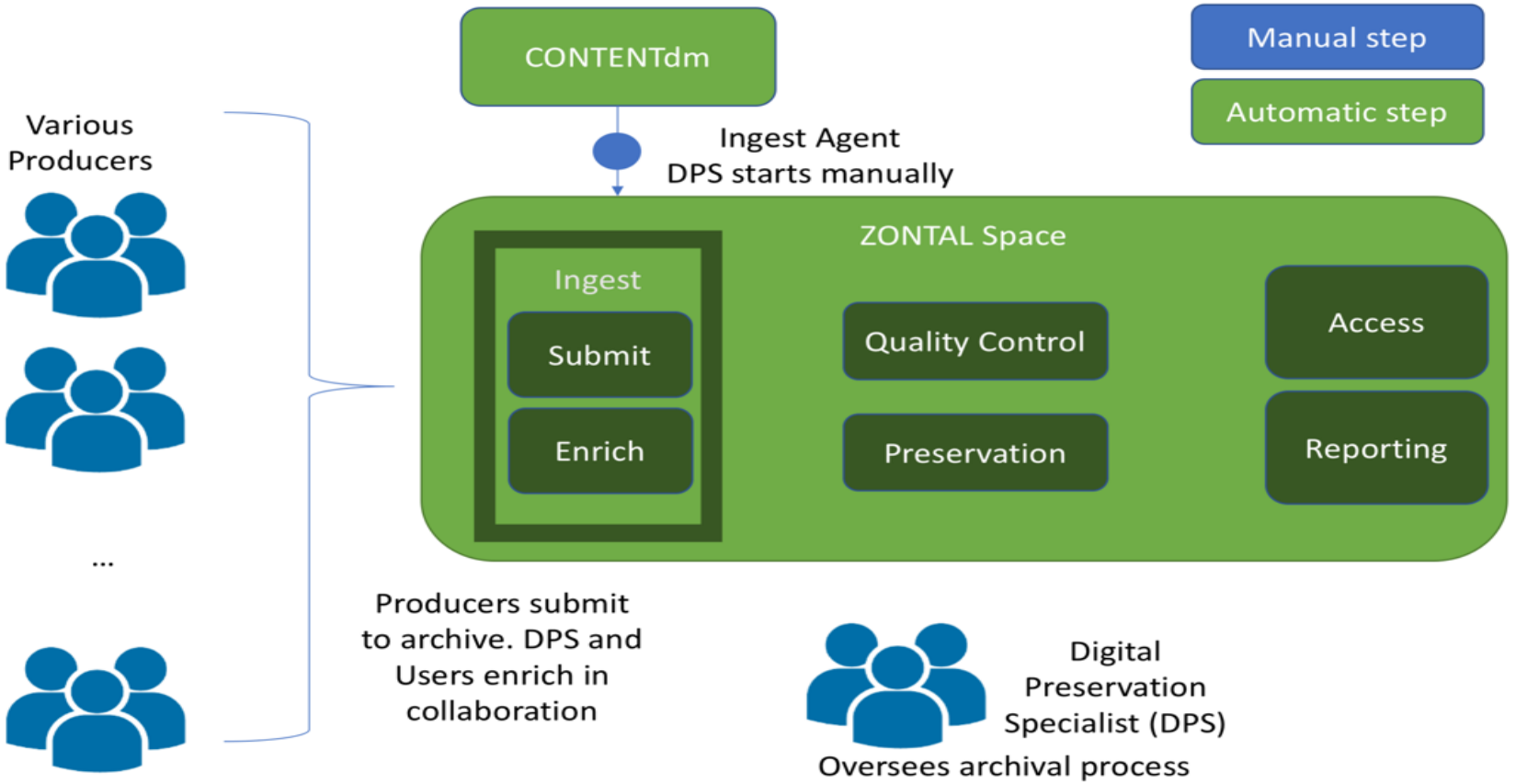
BYU Pre-PoC Workflow

Many manual steps and multiple resources are required



BYU Post-PoC Workflow

Nearly automatic data preservation with minimal resource overhead



DATA INNOVATION AT ENTERPRISE SCALE

Why does this work?

Standardized APIs

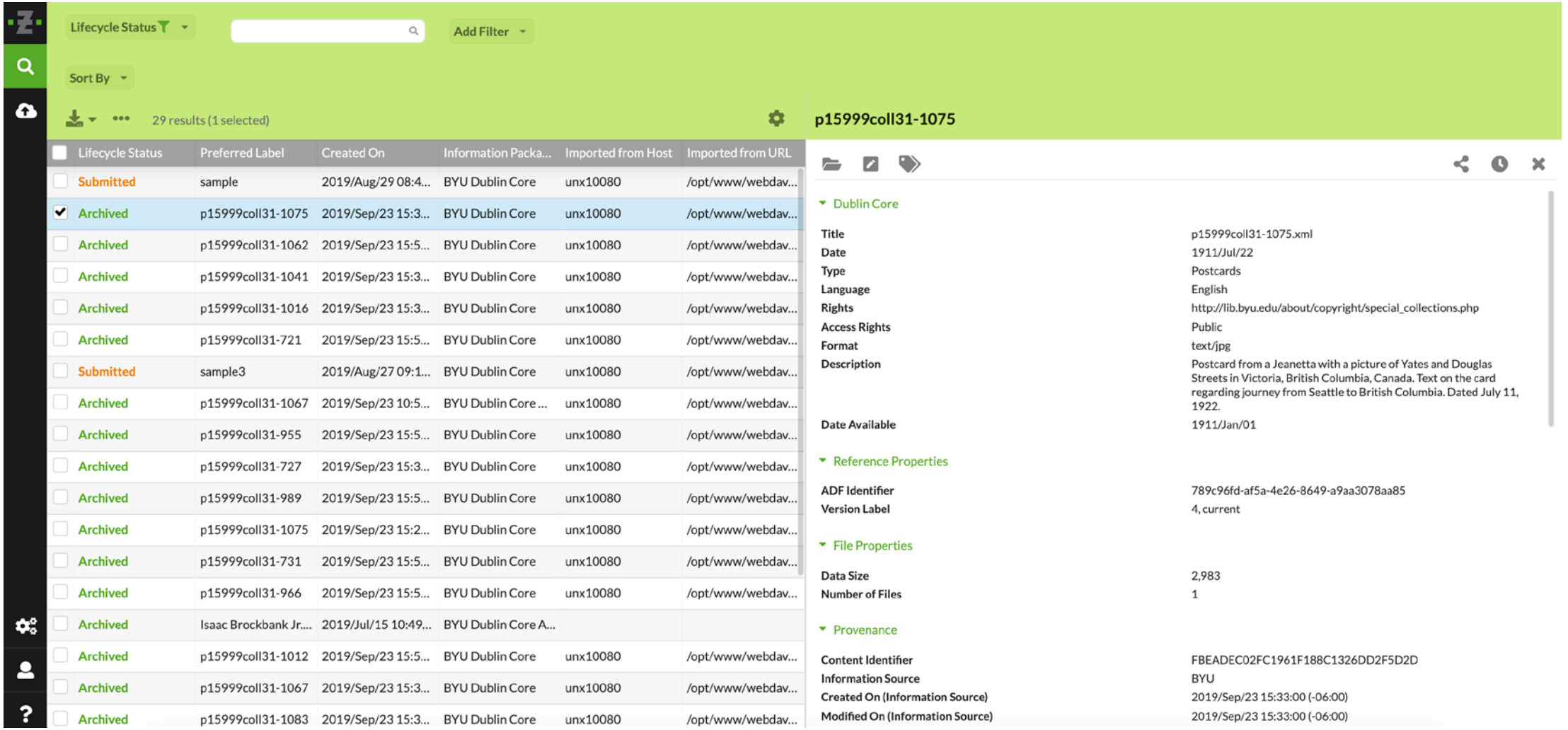
Secure and Automate Access to Metadata and Content

- 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

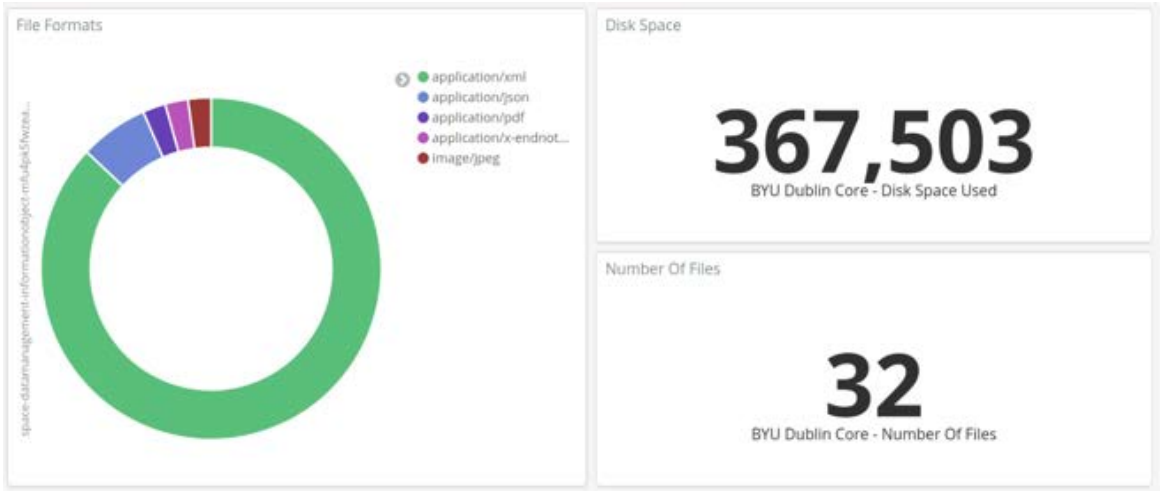
80 percent of the work data scientists perform today, according to this *Forbes Magazine* report, can be automated. This puts great pressure on companies to FAIRify their data today, so that automated systems may be of the highest possible quality and performance.

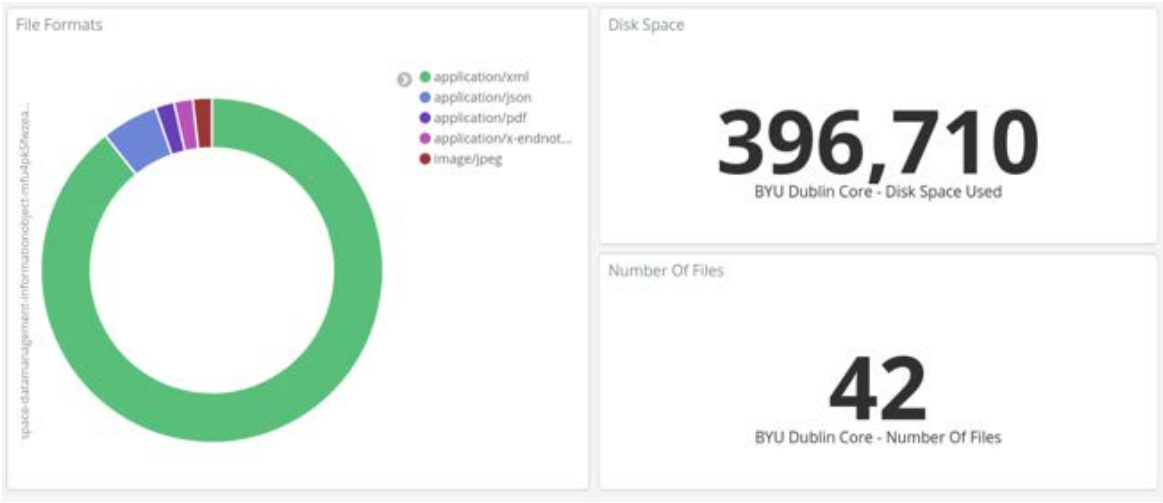
PoC: One manual ingest command bulk imported hundreds of records and automatically enriched them with metadata



Little features with big impact

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





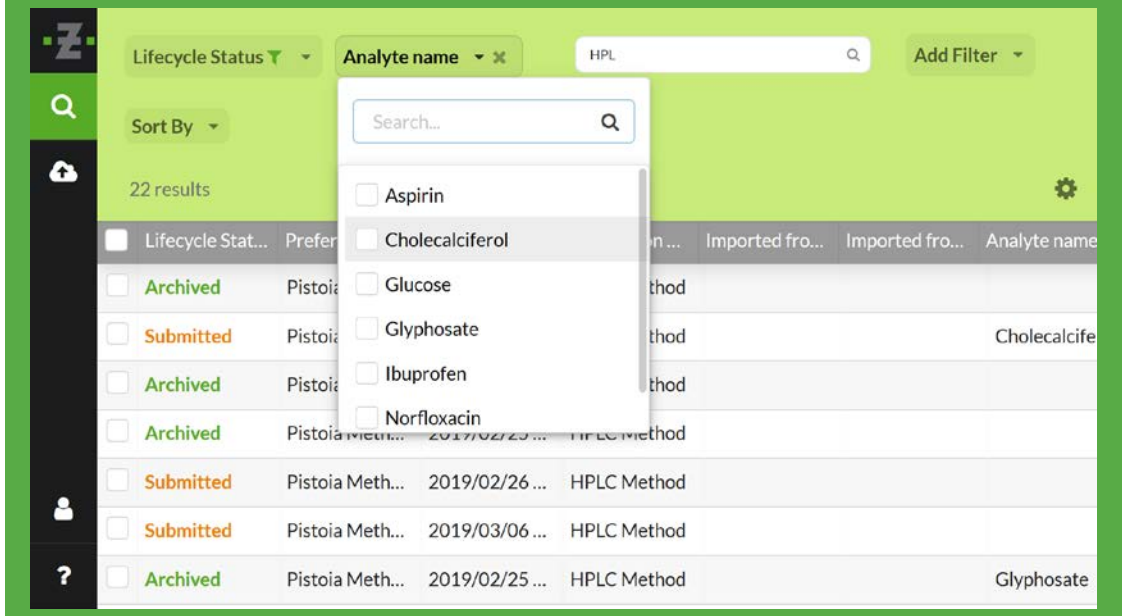
Connecting Data in Archives to Data Analytics

Capture

- Reference & Masterdata Management
- Web Clients
- File Shares
- CDS
- LIMS
- ELN
- DMS
- SDMS

Ingest

Manage Information Lifecycle



amazon web services S3 | AZURE | GCP | NAS | MINIO | DELL EMC ECS | HITACHI Inspire the Next

Access

REST API

OPENAPI INITIATIVE

SQL

SPARQL

{REST API}

Reuse

Spotfire

tableau

Qlik Sense

Microsoft Power BI

python

HIVE

R

APACHE Spark

What can you do with data in ZONTAL Space?

Next steps for library use case:

Enhance existing recommender systems to leverage descriptive information.