

Ana2

---

# Ana2:

Systematic, centralized archiving and automation of laboratory data using Allotrope technologies

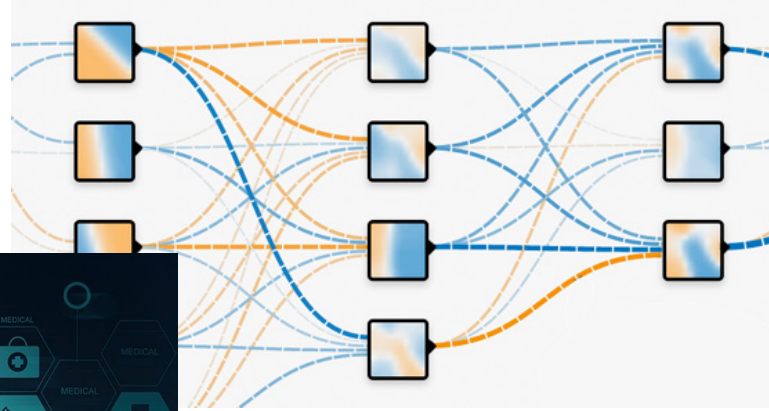
Corey E. Bakalarski, PhD

Scientist

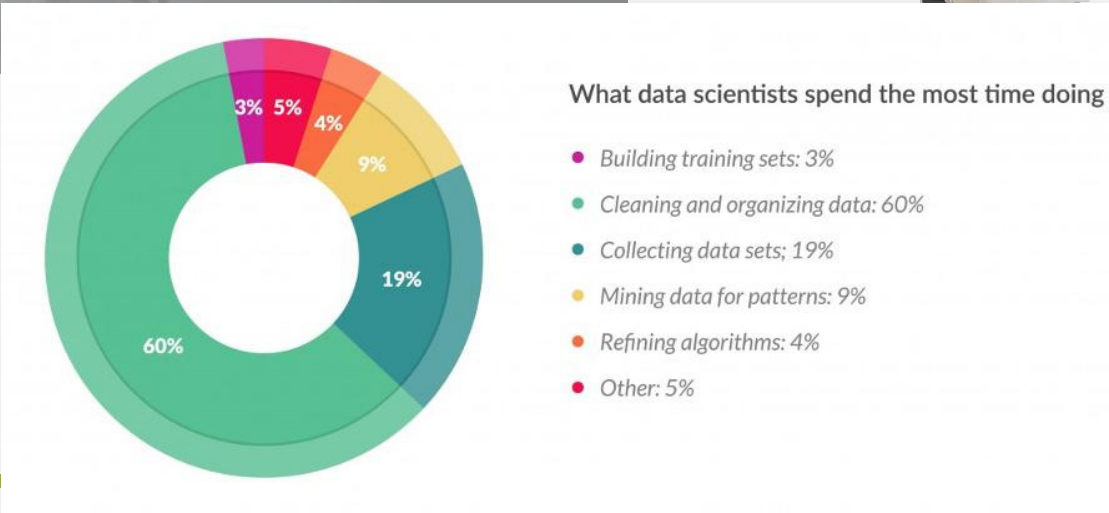
Depts. Of Microchemistry, Proteomics & Lipidomics  
and Bioinformatics & Computational Biology

Genentech Research and Early Development (gRED)

# Digital Transformation in the Life Sciences



# Digital Transformation in the Life Sciences



Gil Press, [www.forbes.com](http://www.forbes.com), March 23, 2016; based on CloudfLower survey of data scientists

# A Diversity of Data and Analysis Tools: The Research Perspective

Many  
Workflows  
× Many  
Instruments  
× Many  
Formats  
× Many  
Analysis Types  
× Many  
Result Types

A Challenging  
Landscape to  
Connect  
Information



But where do we start?

*Our raw instrument data is the starting point for all of our biological knowledge...*

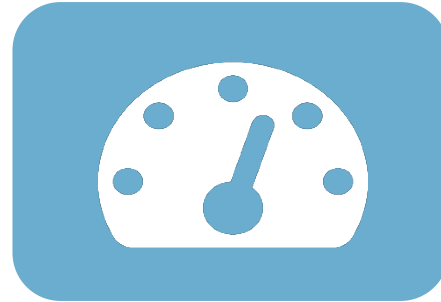
*...and the foundation upon which everything is built.*

# Key Ana2 Requirements



## Scalability

1000s of instruments



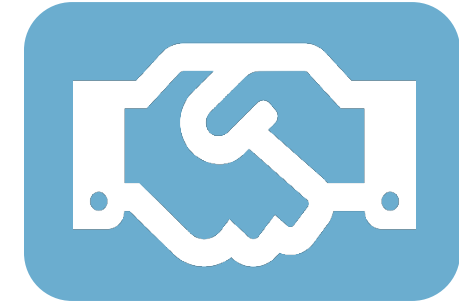
## Speed

Delivery in minutes



## Interface

Global, web-based



## Interoperability

Clearinghouse for data



## Flexibility

Broad support for any instrument



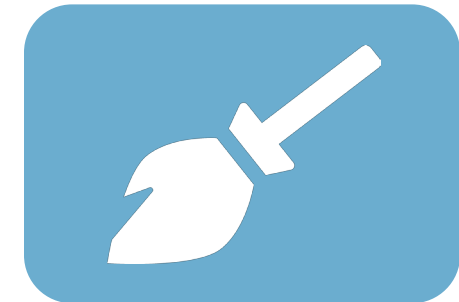
## Evolution

Support scientific advances, maintain old



## Access

Balance requirements



## Cleanup

Maintain instrumentation

# Ana2 Automatically Archives and Delivers All Lab Data

Push vs. Pull:  
Performance



Lightweight  
Monitoring Agent

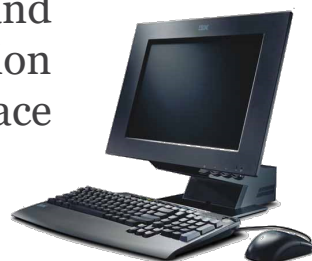


Ana  
submission  
servers



Scale to 100s of  
Instruments

Ana web-based  
user and  
administration  
interface

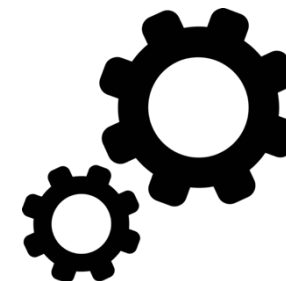


Whole Dataset  
Delivery

Ana  
Data Lake



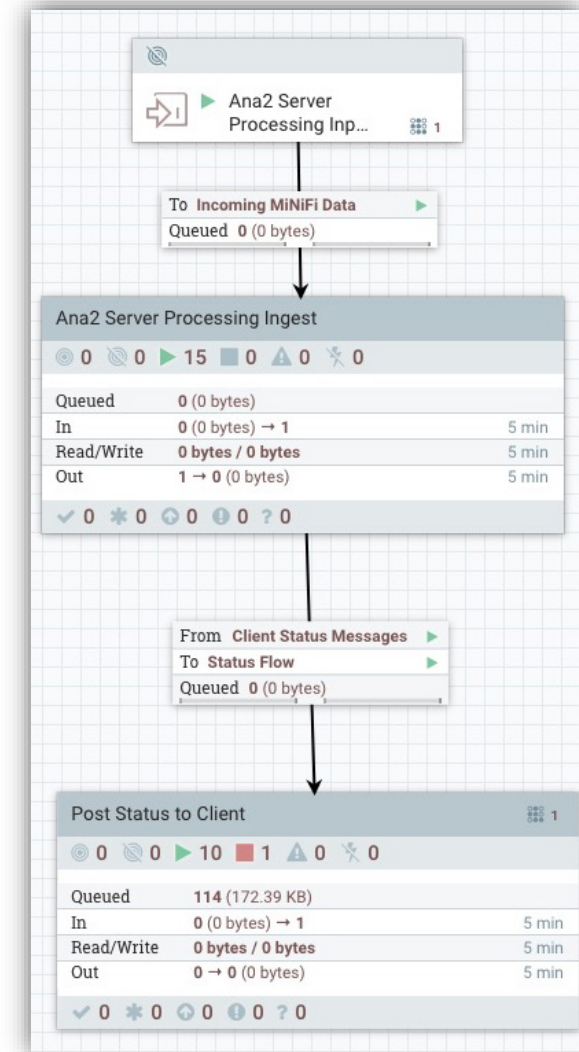
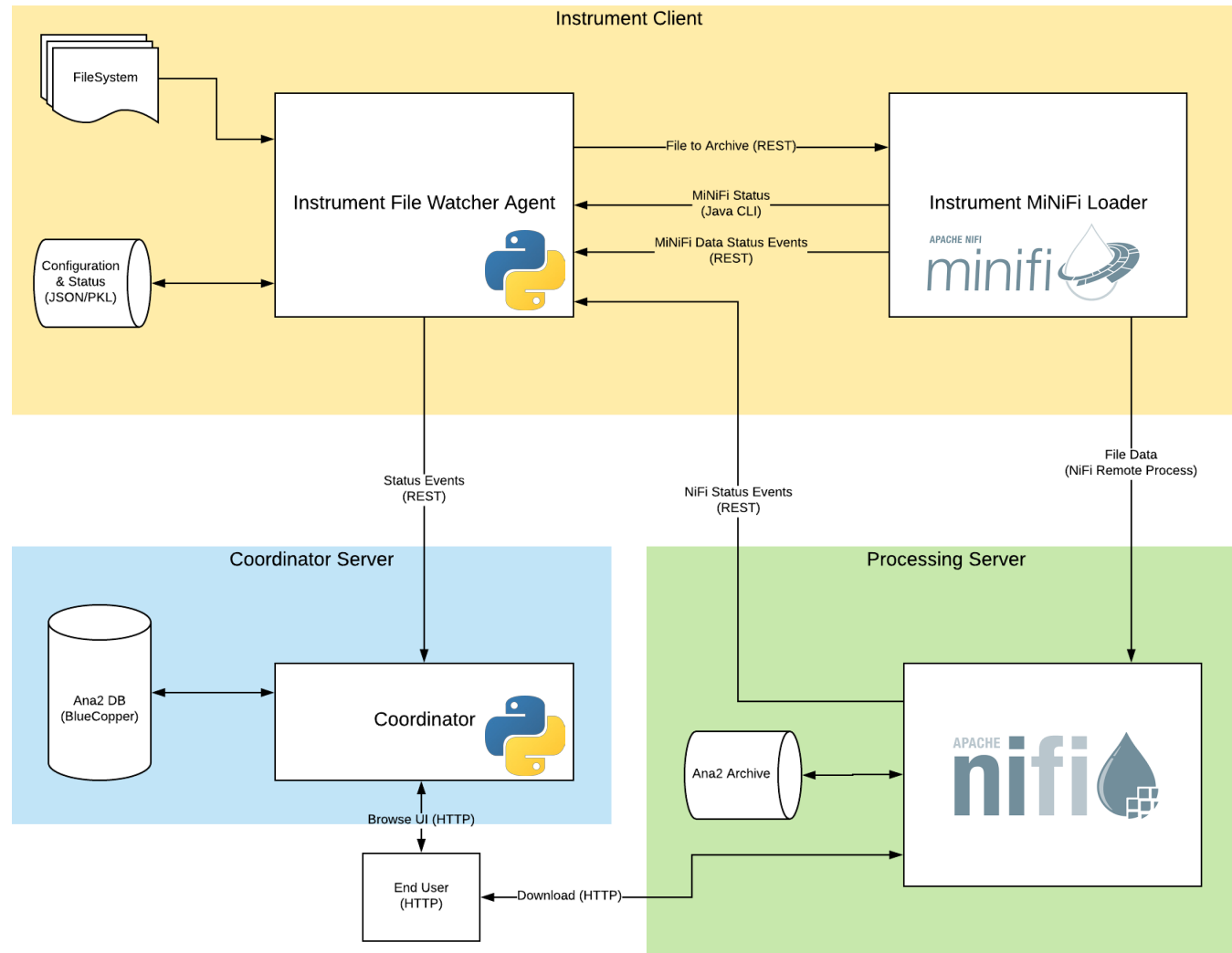
Automate Data  
Processing



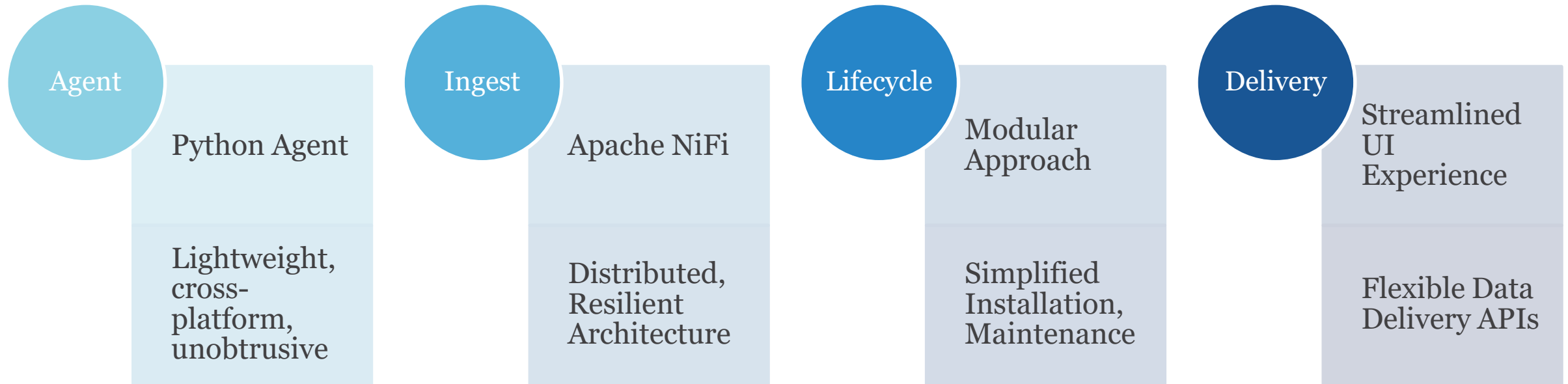
Analysis Tools

RESTful  
& Message APIs

# Decentralized Architecture Eliminates Bottlenecks



# Ana2: Designed to be Lightweight and Scalable



# Ana2: Analytical Data at Your Fingertips – Anywhere, Anytime

Ana2 Home Clients Help Admin panel Bluecopper COREYEB

Q blank DOWNLOAD SELECTED Columns Filters

Status: All Client: All Last status change date: - Client group: All

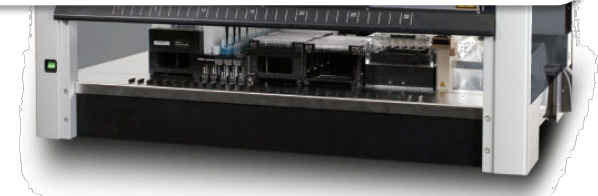
Status	Client	Name	Size	Last Modified	Last Status Change	First Seen	UUID
CLEANED	L-Ana2 Sciex VM2	blank2.wiff	2 GB	03/06/2019 11:24:34	03/06/2019 11:25:07	03/06/2019 11:24:33	c447aedc-0fa2...
VERSION_IN_PROGRESS	L-Ana2 Sciex VM2	blank1.wiff	3 GB	03/01/2019 4:31:47	03/01/2019 4:33:36	03/01/2019 4:31:46	947092a5-c16...
CLEANED	L-ANA Orb 1a	060812_blank...	51 MB	02/13/2019 12:15:49	02/13/2019 12:16:34	02/13/2019 12:15:48	e3417fe4-3f91...
ARCHIVED	L-ANA Orb 1a	060812_blank...	51 MB	02/12/2019 5:25:37	02/12/2019 5:26:25	02/12/2019 5:25:36	7d4bc09a-352...
VERSION_READY	L-ANA Orb 1a	060812_blank...	51 MB	02/12/2019 3:33:33	02/12/2019 3:34:17	02/12/2019 3:33:32	90a8e07d-a15...
VERSION_READY	L-ANA Orb 1a	060812_blank...	51 MB	02/12/2019 2:19:14	02/12/2019 2:20:00	02/12/2019 2:19:14	e8370542-beb...
ARCHIVED	L-ANA Orb 1a	060812_blank...	51 MB	02/07/2019 12:03:49	02/07/2019 12:04:34	02/07/2019 12:03:49	3da1e582-0ff9...

Showing 1 to 7 of 7 rows 10 per page

# Ana2 Extends Beyond Mass Spectrometry



1,000,000+ ADFs on 90 instruments from 12+ departments as of 18 Sep 2019  
Proteomics – Genomics – DMPK – Discovery Chemistry – Microscopy – Antibody Engineering



Microscopes

Sequencers

Cell Counters

Liquid Handlers

# Making Laboratory Data FAIR Across Techniques

[www.nature.com/scientificdata](http://www.nature.com/scientificdata)

SCIENTIFIC DATA 

OPEN

SUBJECT CATEGORIES

» Research data

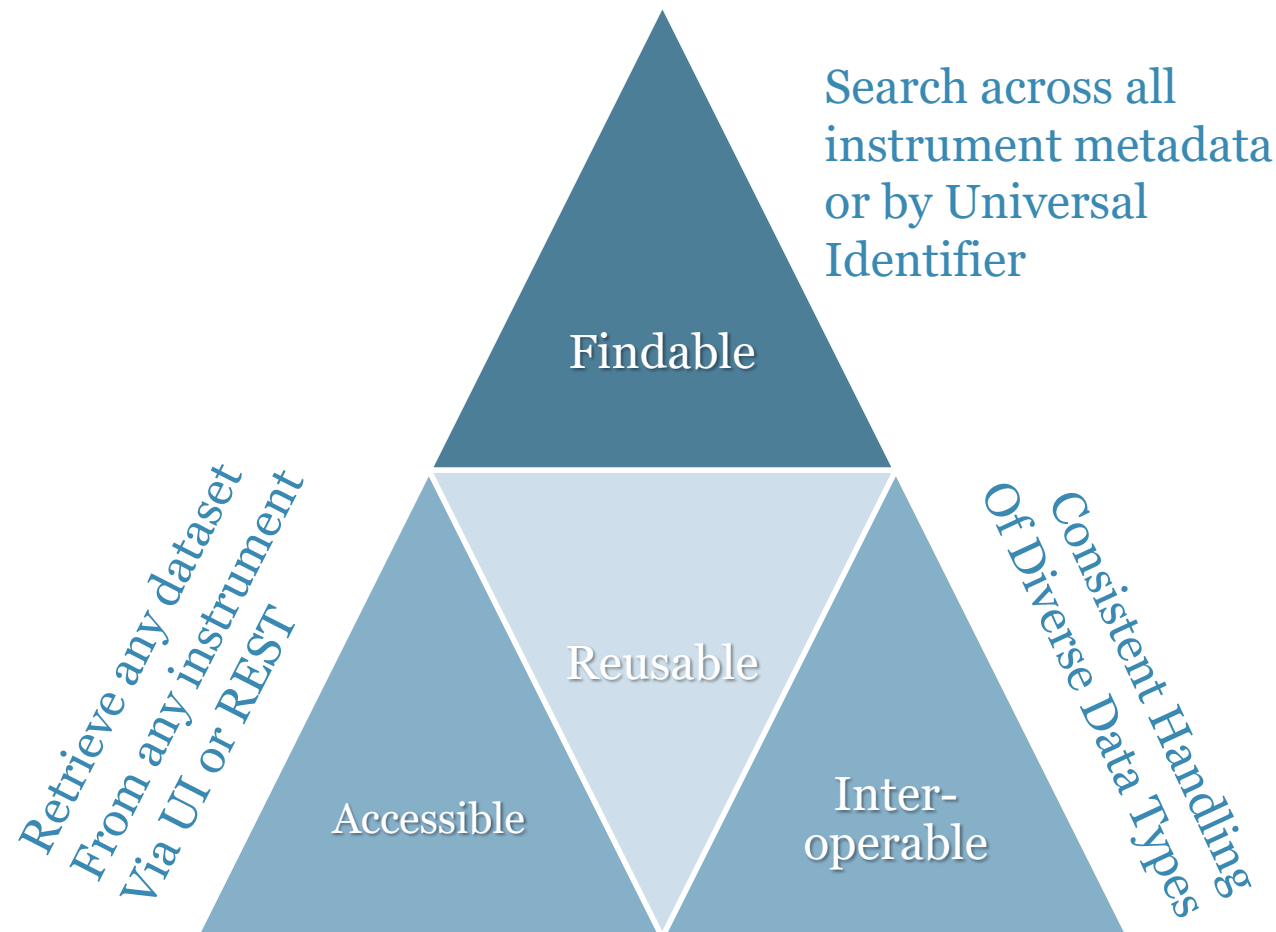
» Publication characteristics

**Comment:** The FAIR Guiding Principles for scientific data management and stewardship

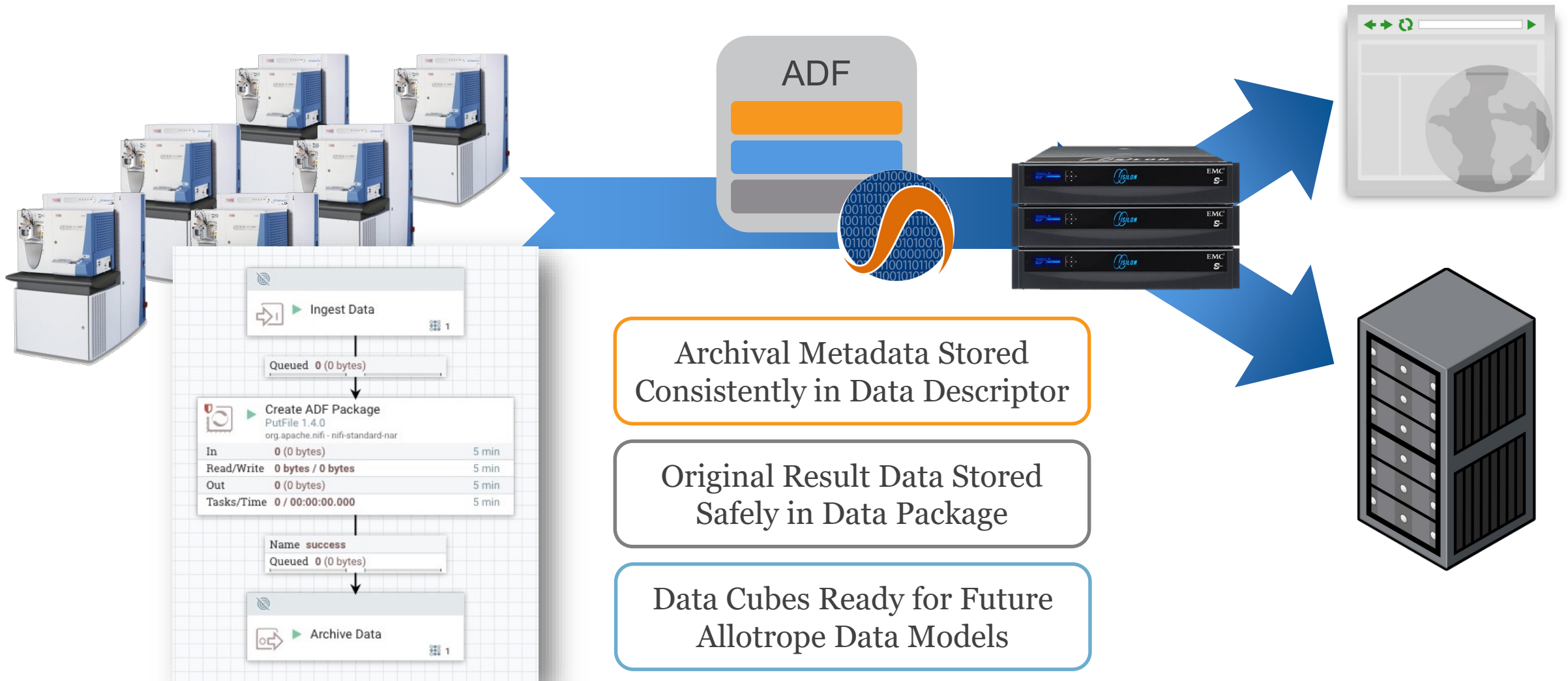
Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*3:160018 (2016).

The FAIR Guiding Principles:

- To be Findable
- To be Accessible
- To be Interoperable
- To be Reusable



# Allotrope Technologies Provide Essential Consistency

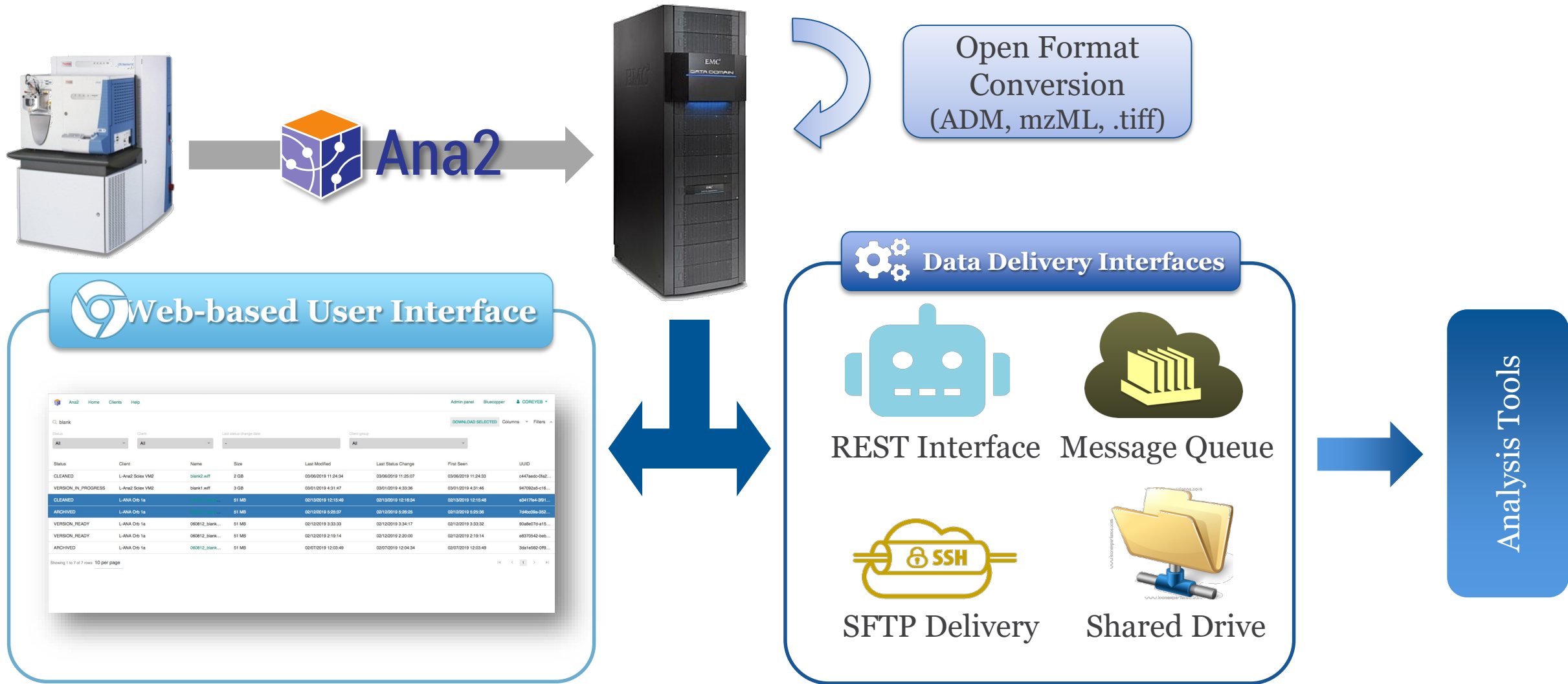


Archival Metadata Stored Consistently in Data Descriptor

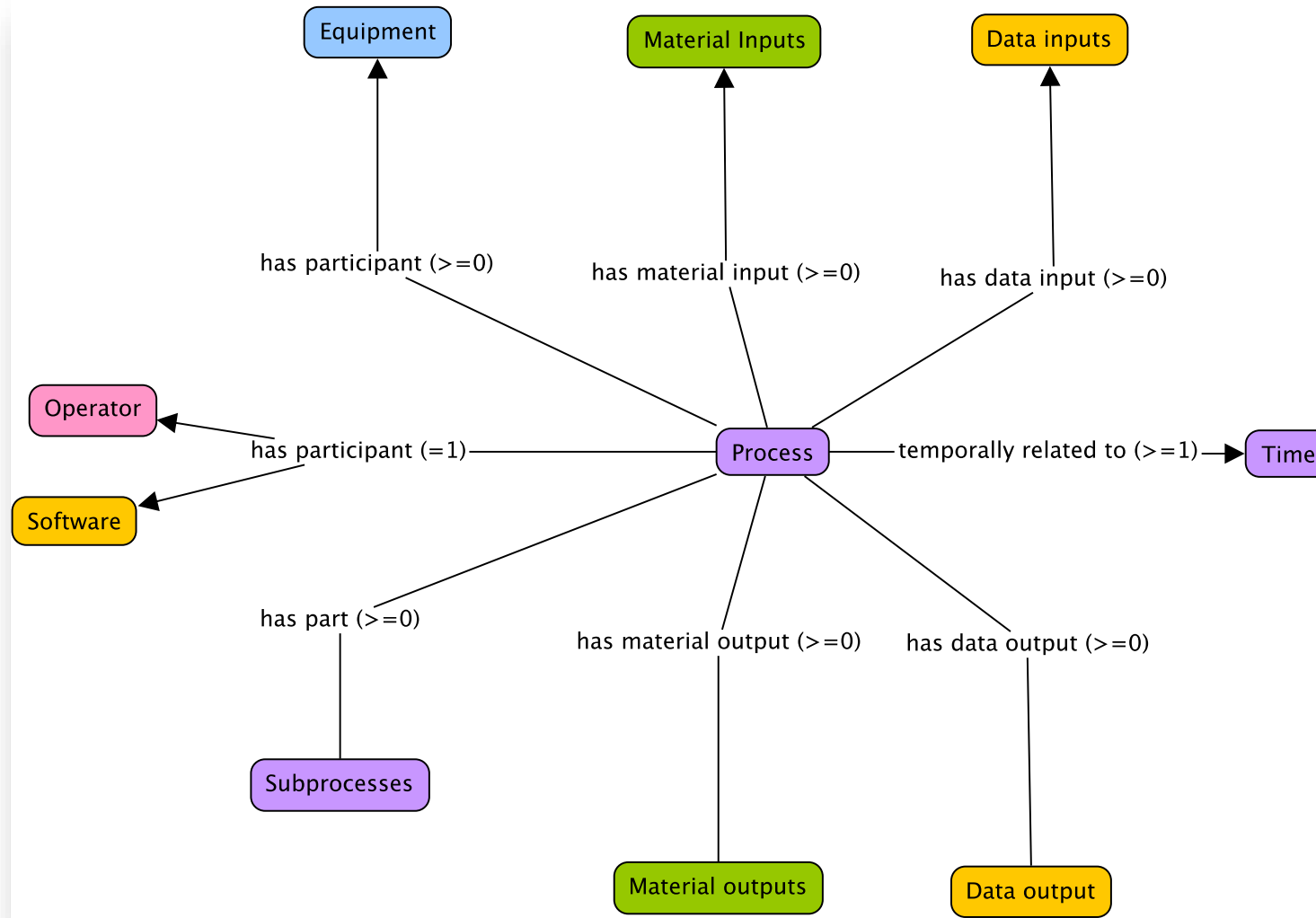
Original Result Data Stored Safely in Data Package

Data Cubes Ready for Future Allotrope Data Models

# Beyond Archival: Automation and Data Delivery



# Process Focus Enables Scientific Knowledge Interchange



# Ana2 uses AFO to Deliver Consistent Result Data

```
"@id": "http://purl.allotrope.org/ontologies/process#AFP_0000748",
"@type": "process",
"prefLabel": "high performance liquid chromatography",
"has data output": {
  "@id": "http://ana2.gene.com/rest/dataset/ca0fb3ec-de57-49d1-8b13-9e06659b2470",
  "@type": "dataset",
  "original dataset location": "/path/to/original/location",
  "has version" : {
    "@id": "http://ana2.gene.com/rest/version/02c12605-c1ff-4910-bc13-0fbcbb1db43",
    "@type": "version",
    "last modified" : "2016-10-19T21:09:49Z"
  }
},
"device": {
  "@id": "http://ana2.gene.com/rest/client/1234543543-1231241-12124324-14123534252",
  "device type" : {
    "@type": "http://purl.allotrope.org/ontologies/equipment#AFE_0000224",
    "prefLabel" : "high-performance liquid chromatograph"
  },
  "metadata": {
    "device manufacturer" : "Agilent",
    "device model" : "1200"
  },
  "location" : "41-3040",
  "prefLabel" : "Agilent HPLC 2"
},
"participant" : {
  "name": "coreyeb",
  "@type": "person",
  "role": {
    "@type": "http://purl.allotrope.org/ontologies/role#AFRL_0000003",
    "prefLabel": "analyst"
  }
}
}
```



```
{ "@context": {
  "version": {
    "@id": "http://purl.mpl.gene.com/term/version"
  },
  "dataset": {
    "@id": "http://purl.org/linked-data/cube#DataSet"
  },
  "process": {
    "@id": "http://purl.obolibrary.org/obo/BFO_0000015"
  },
  "device": {
    "@id": "http://purl.mpl.gene.com/term/client"
  },
  "has data output": {
    "@id": "http://purl.allotrope.org/ontologies/property#AFX_0000544"
  },
  "has version": {
    "@id": "http://purl.allotrope.org/ontologies/property#AFX_0002737"
  },
  "analyst" : {
    "@id": "http://purl.allotrope.org/ontologies/role#AFRL_0000003"
  },
  "last modified" : {
    "@id": "http://purl.org/dc/terms/modified",
    "@type": "http://www.w3.org/2001/XMLSchema#dateTime"
  },
  "original dataset location": {
    "@id": "http://purl.allotrope.org/ontologies/result#AFR_0000919"
  },
  "location": {
    "@id": "http://purl.allotrope.org/ontologies/spatial#AFSP_0000001"
  },
  "prefLabel": {
    "@id": "http://www.w3.org/2004/02/skos/core#prefLabel"
  },
  "metadata": {
    "@id": "http://purl.allotrope.org/ontologies/property#AFX_0002802"
  },
  "device manufacturer": {
    "@id": "http://purl.allotrope.org/ontologies/result#AFR_0001258"
  },
  "device model": {
    "@id": "http://purl.obolibrary.org/obo/IAO_0000017"
  },
  "device type": {
    "@id": "http://purl.allotrope.org/ontologies/equipment#AFE_0000354"
  },
  "name": {
    "@id": "http://purl.allotrope.org/ontologies/result#AFR_0001063"
  },
  "participant": {
    "@id": "http://purl.obolibrary.org/obo/RO_0000057"
  }
}}
```



# Allotrope: Paying Dividends for Ana2

## Process does not Define Product

- Using standards like AFO, ADM, ADF allow innovation and modularization
- Open standards ensure data remains accessible (now and in the future)

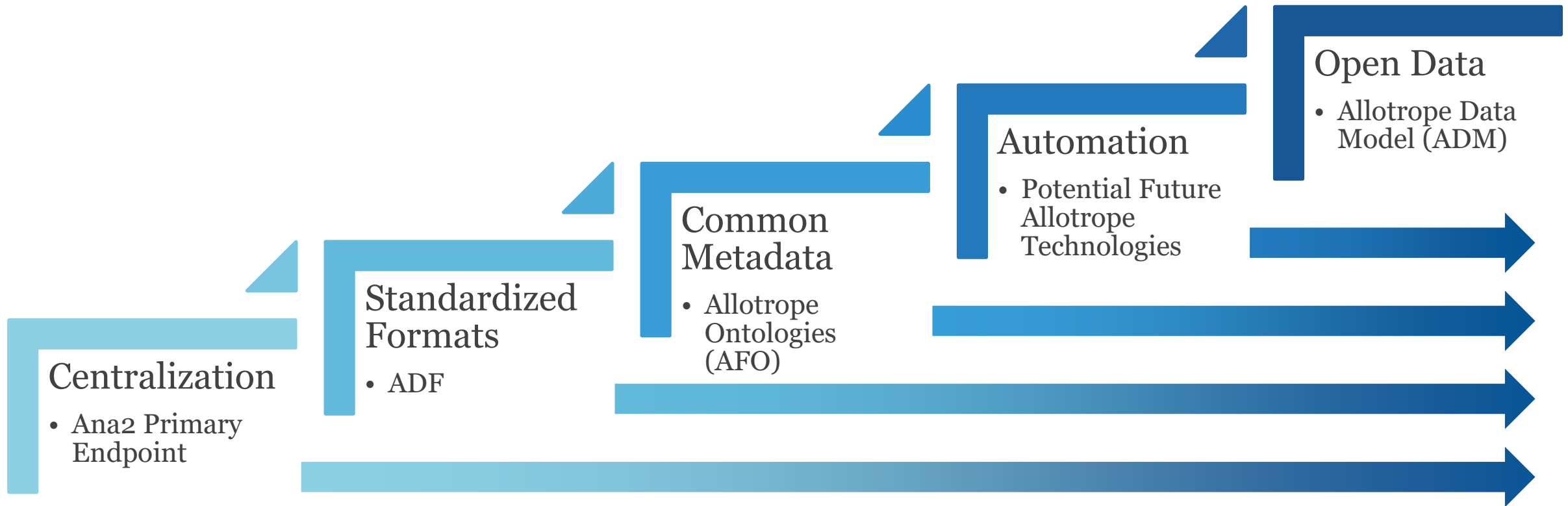
## Open Standards Ensure Accessibility

- Now: allow centralization and automation
- Future: provide compatibility for existing data
- Renewed focus on the science

## Complementary Technologies Working Together

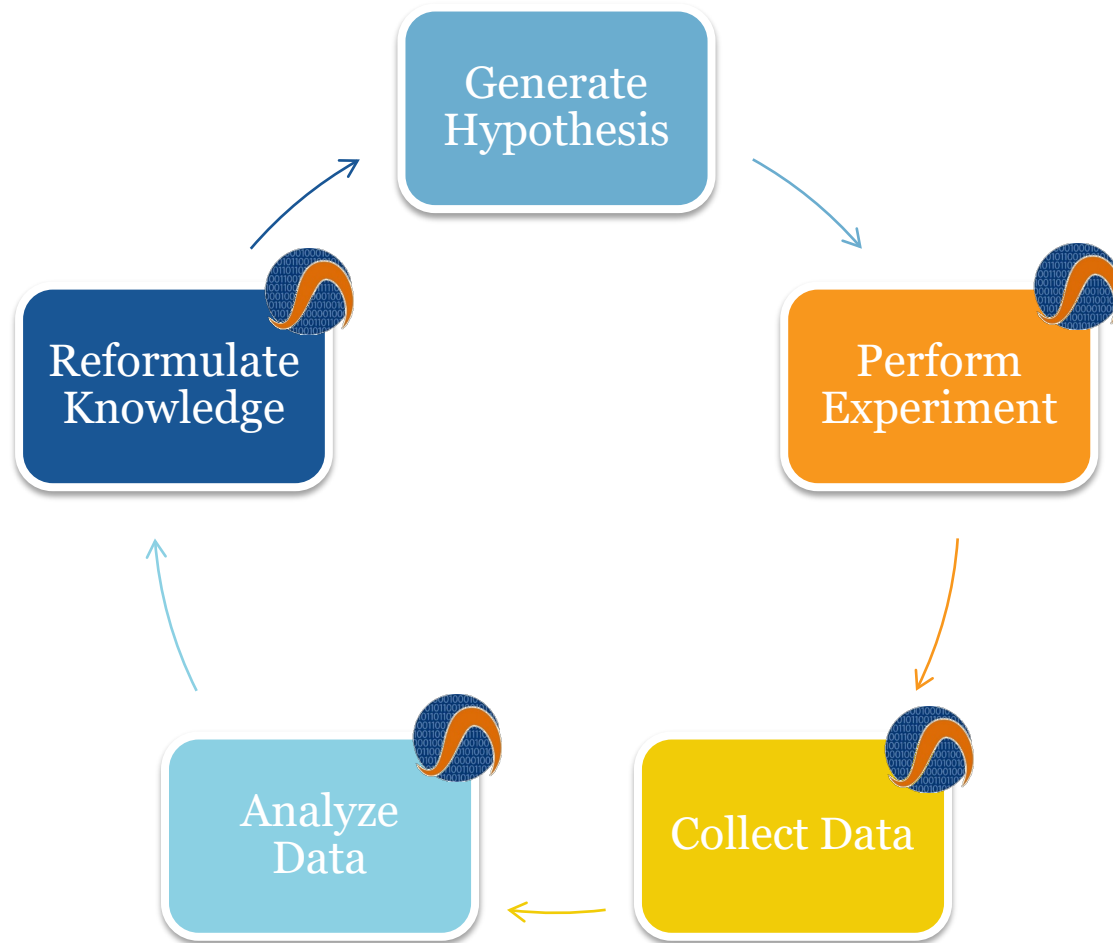
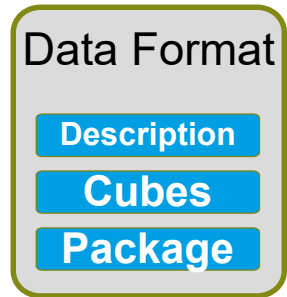
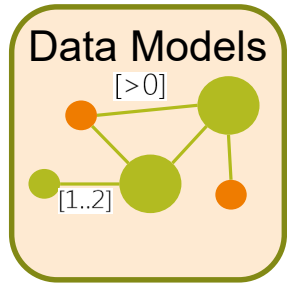
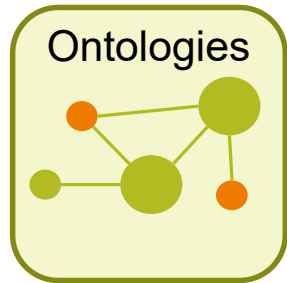
- Using Allotrope Data Format together with Ontologies and Data Models is a perfect fit for data archival and transfer
- Deploying them separately in other use cases (search, delivery) retains common ecosystem

# A Layered, Evolutionary Approach to Digital Laboratory Transformation



Deploying ADF widely allows us to handle data consistently, at scale, today – and allows us to enrich it in the future

# Allotrope and Ana2 Manage the Scientific Knowledge Lifecycle



- The FAIR Guiding Principles:
- To be Findable
  - To be Accessible
  - To be Interoperable
  - To be Reusable



**Genentech**

*A Member of the Roche Group*

[bakalarski.corey@gene.com](mailto:bakalarski.corey@gene.com)