

A decorative graphic consisting of four horizontal lines of varying colors and thicknesses: a thin light blue line, a medium-thick orange line, a thick dark blue line, and another medium-thick orange line.

# **The Empower CDS Allotrope Community Project**

2019 Fall Allotrope Connect Meeting

*October 8, 2019*

# Who is in The Community

## Pharmaceutical Companies

- Baxter 
- Bayer 
- Boehringer Ingelheim 
- Genentech   
A Member of the Roche Group
- GSK 
- Merck   
INVENTING FOR LIFE
- Pfizer 

## Vendor

- TetraScience 

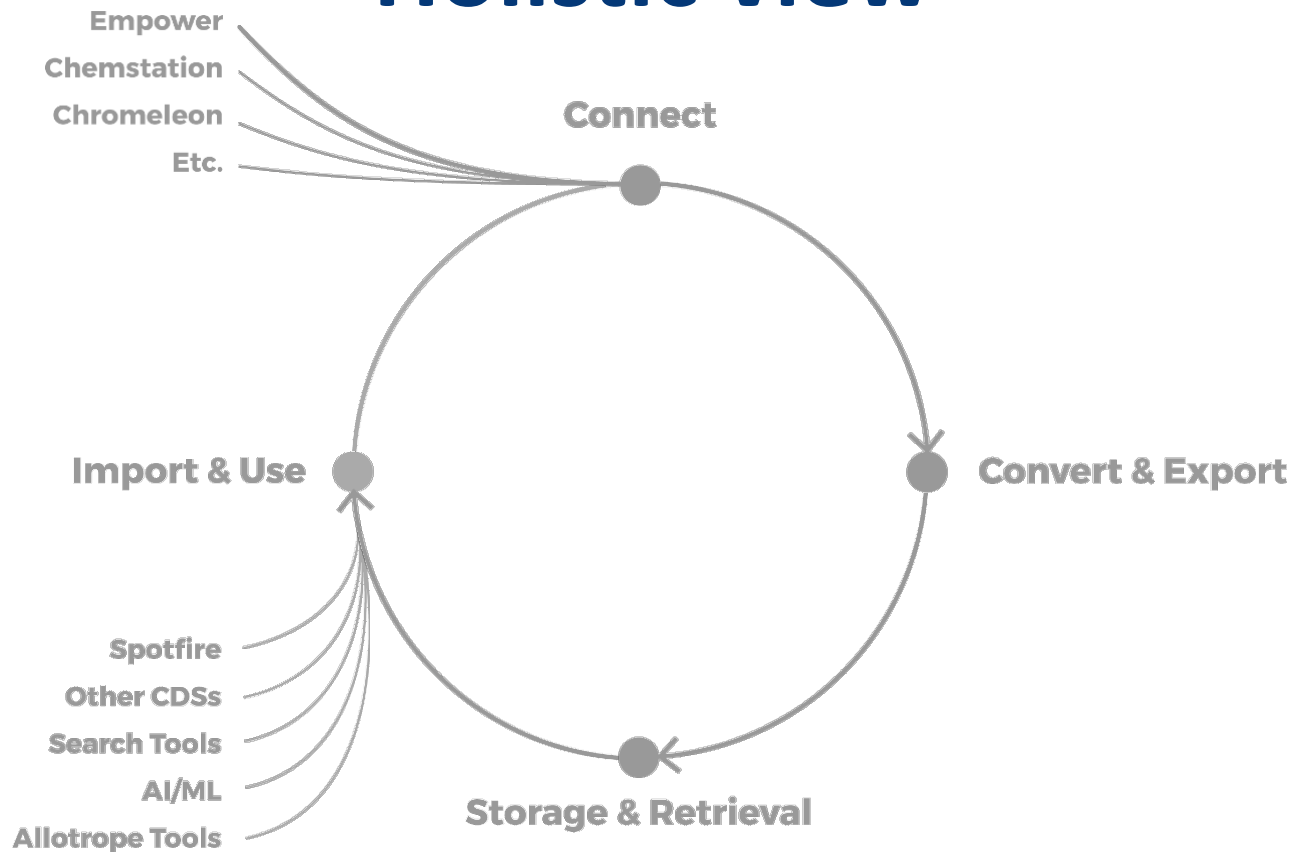


# Why are we doing this?

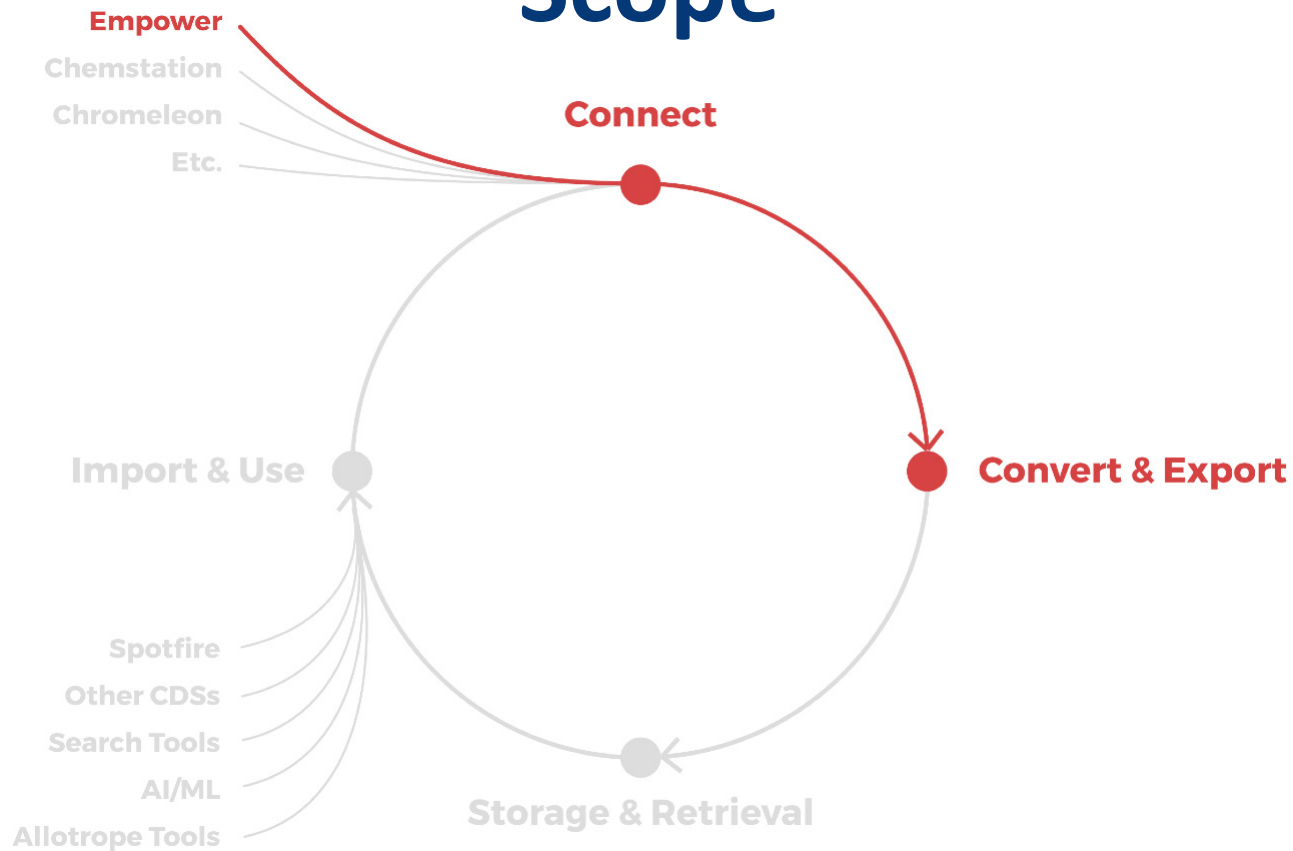
- Why do we want a CDS ADF?
  - Downstream analytics/integration: Spotfire, ACD, etc...
  - Long-term archiving: No software/version dependences
  - Vendor agnosticity: Mix and match CDS... Empower to Chromeleon to ChemStation...
  - Method development: Use/reuse the data from 100+ runs
- Many things we haven't even imagined yet will be possible once we able to access data freely...



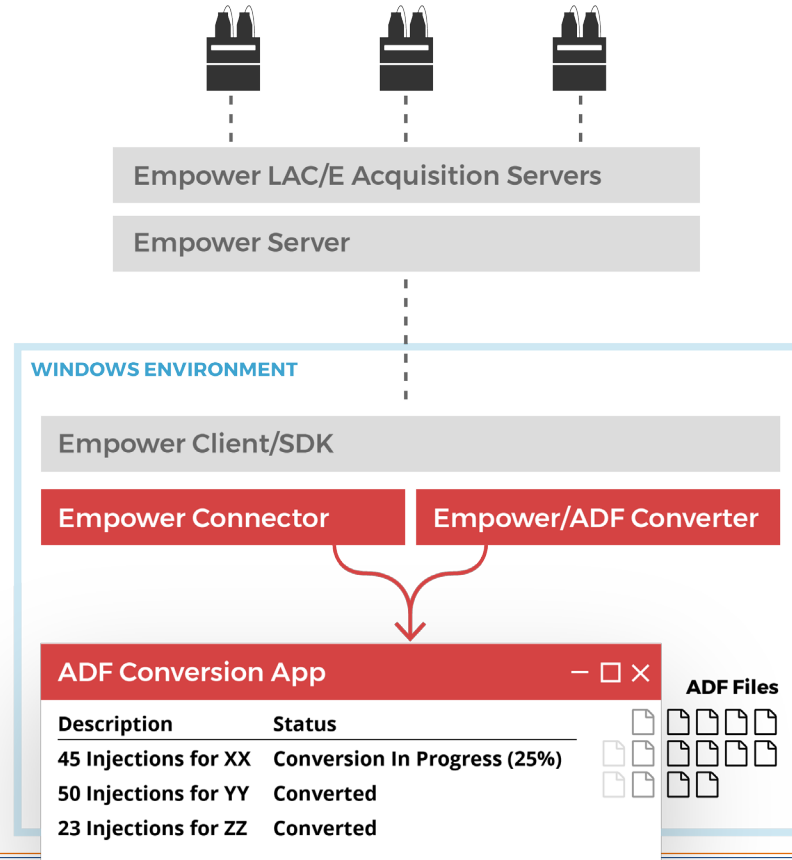
# Holistic view



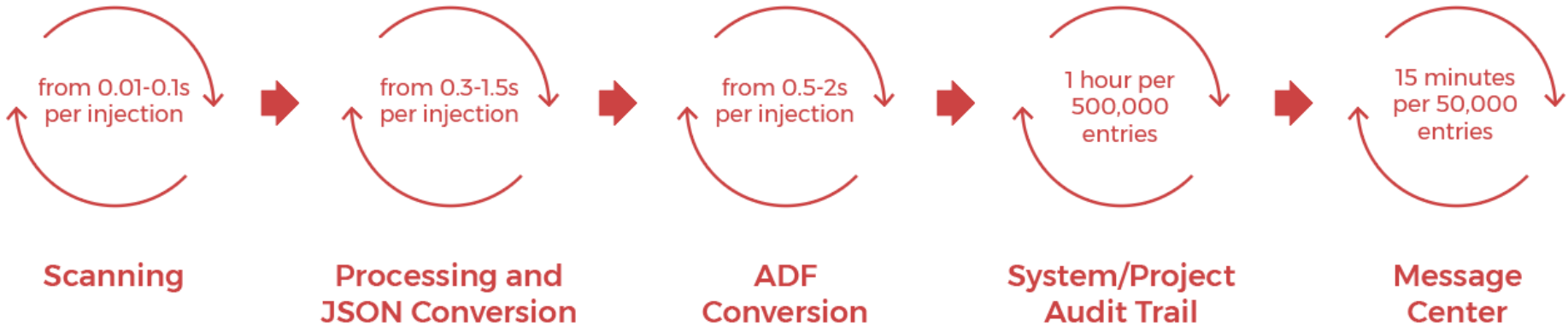
# Scope



# What is in this Project?



# How does it work?



# What's the purpose?





## **Two practical use-cases unlocked by the project**

1. Backup/Long-term-storage

2. Scientific data retrieval

# Backup/Long-term Storage



Backup	Project	Backup Status	Last Backup Date	Message	Open
<input type="checkbox"/>	ALLTestProj	Done	4/3/2019 4:25:06 PM		Open
<input type="checkbox"/>	DKP_Inhalation_0718	Error	4/3/2019 4:03:41 PM	Backup project DKP_Inhalation_0718 error: One or more Project Integrity error(s) were detected in Please go to the 'DKP_Inhalation_0718' project Properties. View the 'Integrity' page and get 'Details' on the most recent Integrity report for details.	Open
<input type="checkbox"/>	Defaults				Open

File Explorer window showing the contents of the folder: Empower-Agent > ProjectBackup > ALLTestProj > 20190403162400 > ALLTestProj

Name	Date modified	Type	Size
ALLTestProj.EXP	4/3/2019 4:24 PM	EXP File	4,192 KB
ALLTestProj.INF	4/3/2019 4:24 PM	Setup Information	1 KB
ALLTestProj_EXP.CKS	4/3/2019 4:24 PM	CKS File	2 KB
ALLTestProj_INF.CKS	4/3/2019 4:24 PM	CKS File	2 KB
backup.log	4/3/2019 4:24 PM	Text Document	6 KB
d1030.dat	3/16/2018 2:28 PM	DAT File	2,744 KB
d1031.dat	3/16/2018 2:28 PM	DAT File	2,744 KB
d1032.dat	3/16/2018 2:28 PM	DAT File	818 KB
d1044.dat	3/16/2018 2:28 PM	DAT File	2,558 KB
d1045.dat	3/16/2018 2:28 PM	DAT File	2,558 KB
d1046.dat	3/16/2018 2:28 PM	DAT File	754 KB
d1058.dat	3/16/2018 2:28 PM	DAT File	2,574 KB
d1059.dat	3/16/2018 2:28 PM	DAT File	2,574 KB



Backup	Project	Backup Status	Last Backup Date	Message	Open
<input type="checkbox"/>	ALLTestProj	Done	4/3/2019 4:25:06 PM		<input type="button" value="Open"/>
<input type="checkbox"/>	DKP_Inhalation_0718	Error	4/3/2019 4:03:41 PM	Backup project DKP_Inhalation_0718 error: One or more Project Integrity error(s) were detected in Please go to the 'DKP_Inhalation_0718' project Properties. View the 'Integrity' page and get 'Details' on the most recent Integrity report for details.	<input type="button" value="Open"/>
<input type="checkbox"/>	Defaults				<input type="button" value="Open"/>
<input type="checkbox"/>	Demo				<input type="button" value="Open"/>

File Explorer window showing the contents of the folder 1920.

Address bar: << 20190403162400 > ADF > NIF\_RS\_WN\_101514C\_1822 > 1920

Name	Date modified	Type	Size
20141015142229.adf	4/3/2019 3:41 PM	ADF File	25,146 KB

Left sidebar (Navigation pane):

- Quick access
- Desktop
- Downloads
- Documents
- Pictures
- bin
- nxlog
- nxlog
- nxlog
- This PC
- Empower Data (D:)

Bottom left: 1 item



# Scientific Data Retrieval



User Request		Request History		Request ADF			
Project	Selected	IsSelected	Sameset/Injection	Last Requested date	Message		
ALLTestProj	3 Injections	<input type="checkbox"/>	NIF_SP_WP_101414A				
CIA_Default		<input type="checkbox"/>	NIF_WN_WP_101514E				
Defaults		<input checked="" type="checkbox"/>	NIF_RS_WN_101514C				
Defaults_FAT		<input type="checkbox"/>	1872		Last Converted 4/3/2019 3:41 PM		
▶ Demo		<input type="checkbox"/>	1880		Last Converted 4/3/2019 3:41 PM		
Dissolution_Default		<input checked="" type="checkbox"/>	1888		Last Converted 4/3/2019 3:41 PM		
GPCVLS_Default		<input type="checkbox"/>	1896		Last Converted 4/3/2019 3:41 PM		
ME_Method_DB_Imp_Team		<input type="checkbox"/>	1904		Last Converted 4/3/2019 3:41 PM		
		<input checked="" type="checkbox"/>	1912		Last Converted 4/3/2019 3:41 PM		
		<input checked="" type="checkbox"/>	1920		Last Converted 4/3/2019 3:41 PM		
		<input checked="" type="checkbox"/>	1823		Last Converted 4/3/2019 3:41 PM		
		<input type="checkbox"/>	1840		Last Converted 4/3/2019 3:41 PM		
		<input type="checkbox"/>	1848		Last Converted 4/3/2019 3:41 PM		
		<input type="checkbox"/>	1856		Last Converted 4/3/2019 3:41 PM		
		<input type="checkbox"/>	1864		Last Converted 4/3/2019 3:41 PM		

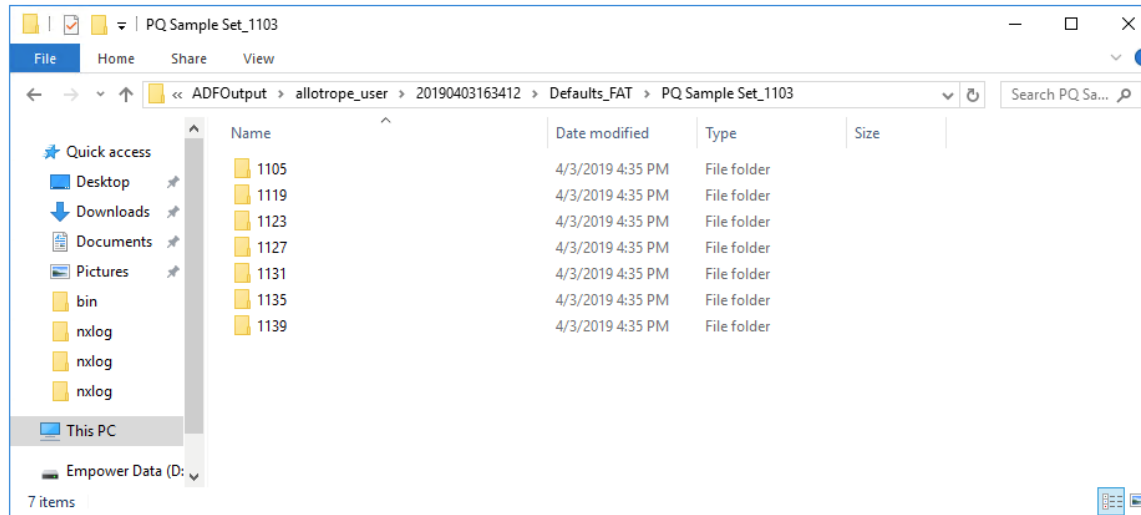


User Request Request History

Refresh

Drag a column header and drop it here to group by that column

	Request Date	Description	Completion Date	Status	Open
>	4/3/2019 4:16:53 PM	PQ sample set injections		Conversion In Progress	Open
	4/3/2019 4:32:46 PM		4/3/2019 4:33:15 PM	Converted	Open
	4/3/2019 4:34:12 PM	Defaults FAT PQ Sample Set	4/3/2019 4:35:15 PM	Converted	Open



# Live Demo of a working ADF...





# Project results



1050+ fields extracted from Empower and available for conversion to ADF

41 data fields are mapped and converted to ADF. Most are SHACL validated.



# What this means

Archival format?

Enabling search/analytics?



## Fields Include the following:

### Peaks Pattern

```
Results[.]{ AdjustedTotalArea: "2788322" }
Results[.]{ resultId: "2622" }
Results[.].Peaks[.]{ Amount: "3733.949" }
Results[.].Peaks[.]{ Area: 778910 }
Results[.].Peaks[.]{ curveld: "2389" }
Results[.].Peaks[.]{ endHeight: 0.0 }
Results[.].Peaks[.]{ endTime: 3.325 }
Results[.].Peaks[.]{ name:"Propiophenone" }
Results[.].Peaks[.]{ width: 21.5 }
Results[.].Peaks[.]{ height: 104391.0 }
Results[.].Peaks[.]{ peakLabel: "3" }
Results[.].Peaks[.]{ Resolution: NO_EXAMPLE }
Results[.].Peaks[.]{ RetentionTime: 1.249 }
Results[.].Peaks[.]{ Selectivity: NO_EXAMPLE }
Results[.].Peaks[.]{ SignalToNoise: NO_EXAMPLE }
Results[.].Peaks[.]{ startHeight: 0.0 }
Results[.].Peaks[.]{ startTime: 1.167 }
Results[.].Peaks[.]{ USPSPlateCount: NO_EXAMPLE }
Results[.].ProcessingMethods[.].SysSuitParamters.{ TangentPercentUSPPlateCount: "61" }
```



## Fields Include the following:

### Mobile Phase Pattern

```
InsturmentMethod.Setups[.]{ SolventA: "Water" }  
InstrumentMethod.Setups[.]{ SolventB: "Methanol" }
```

### Gradient Table Pattern

```
InsturmentMethod.Setups[.]{ PctA: "40" }  
InsturmentMethod.Setups[.]{ PctB: "60" }
```

### Injection Pattern

```
{ ID: 1191 }  
{ DateAcquired: "9/17/1997 5:37:56 PM EDT" }  
{ InjectionVolume: 20.0 }  
{ Label: "U0104" }  
{ ProjectName: "A166033" }  
{ SampleSetId: 1342 }
```

### Sample Pattern

```
{ sampleName: "Blank" }  
{ Altered: "Yes" }  
{ Superseded: "No" }
```



## Fields Include the following:

### Submitter Pattern

```
{ acquiredBy: "test_user" }
```

### Software Agent Pattern

```
Results[].{ SourceSoftwareInfo }
```

### Sequence Pattern

```
SampleSet.{ OriginalSampleSetId }  
SampleSet.{ SampleSetId }
```

### Chromatogram Detection Pattern

```
Channels[].{ ChanelId: 1192 }  
Channels[].{ Channel: "486" }  
Channels[].{ SamplingRate: "2.00" }  
InstrumentMethod[].Setups[].{ Wavelength: "254" }  
InstrumentMethod[].Setups[].{ Channel1Offsets: NO_EXAMPLE }  
InstrumentMethod[].Setups[].{ Channel1Bandwidth: NO_EXAMPLE }
```



Pattern	Pattern Definition	Mapping coverage (Empower graph vs Instance data)
Peaks	100%	90%
Injection	100%	80%
Sequence	100%	100%
Data Processing	100%	40%
HPLC Run	100%	100%
Sample	100%	100%
Gradient Table	100%	90%
Mobile Phase	100%	90%
Chromatogram Detection	0%	N/A
HPLC System	0%	N/A
Baselines	0%	N/A



# Only SHACL validation missing

```
[
  a
  sh:ValidationResult ;
  sh:focusNode          _:b0 ;
  sh:resultMessage      "Less than 1^^http://www.w3.org/2001/
                        XMLSchema#integer values have shape
                        0eaf968141a7f38390cf6865c22ae96f";
  sh:resultPath         af-x:AFX_0000544 ;
  sh:resultSeverity     sh:Violation ;
  sh:sourceConstraintComponent sh:QualifiedMinCountConstraintComponent ;
  sh:sourceShape        _:b1
]
```





# Only SHACL validation missing

```
[
  a
  sh:ValidationResult ;
  sh:focusNode          _:b0; Not sure what this refers to
  sh:resultMessage      "Less than 1^^http://www.w3.org/2001/
                        XMLSchema#integer values have shape
                        0eaf968141a7f38390cf6865c22ae96f";
  sh:resultPath         af-x:AFX_0000544 ;
  sh:resultSeverity     sh:Violation ;
  sh:sourceConstraintComponent sh:QualifiedMinCountConstraintComponent ;
  sh:sourceShape        _:b1
]
```



# Possible Next Steps

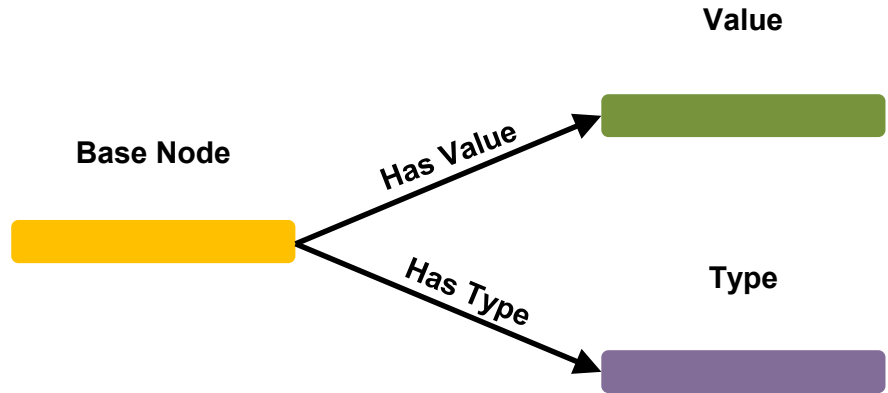


# Idea 1

## Leaf Nodes



All have the same pattern...



# Leaf nodes Pros/Cons

## Pros

<b>Simplicity</b>	Easier to use (less onions to peel) Easier to implement
<b>Human Readable</b>	Scientists can read the data more easily and potentially obtain value sooner
<b>Lossless</b>	Any data field can be captured, as it's not constrained to an ontology
<b>Quicker time to value</b>	Value can be realized sooner without building complex data-abstractions or huge learning curves for end-users

## Cons

More difficult to do data mining and advanced analytics	The relationships between terms are not fully defined (ex: harder to do joins and draw inferences)
Not significantly different than tag/value pairs	We will need to come up with patterns for more complex data structures (lists, nested objects, etc.)



# Idea 2

## Map more fields



41 data fields are mapped and converted to ADF

We can continue expanding this by:

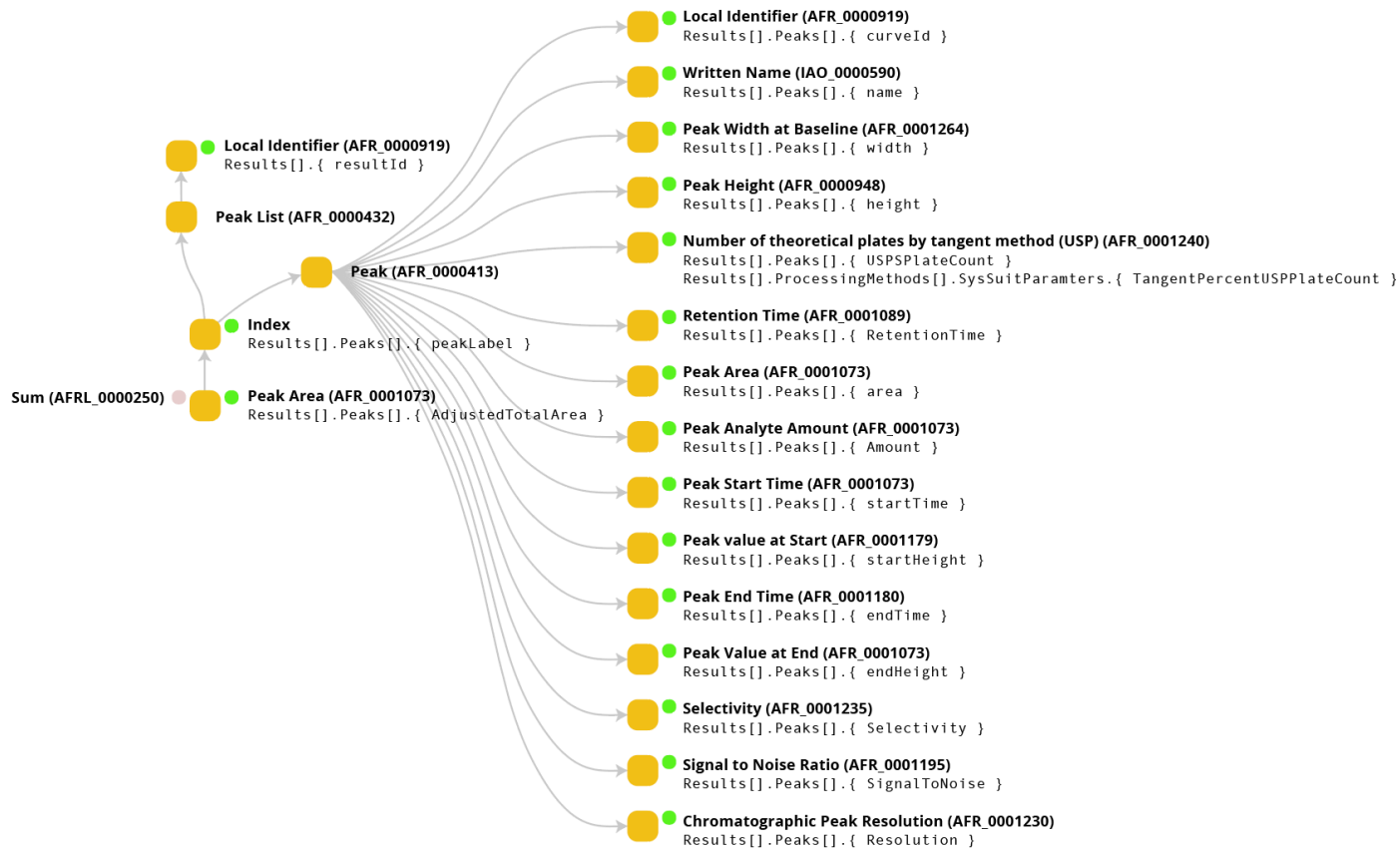
- 1) Expanding beyond the LC-UV reference**
- 2) Mapping against the GC model**



# Expanding beyond LC-UV reference







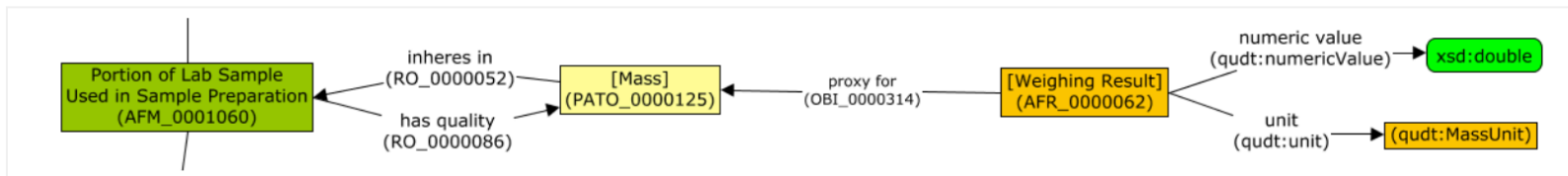
+ Additional fields



# Using the GC model



SampleWeight is described in the GC Sample materials class model (Note this is not specific at all to GC, the GC model is being used to augment the chromatography model with items that were not included in the initial release.

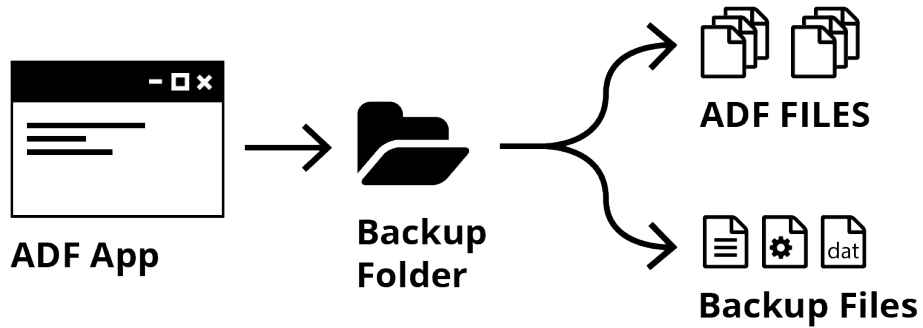


# Idea 3

## Search for ADFs and Backups



## CURRENT PROJECT



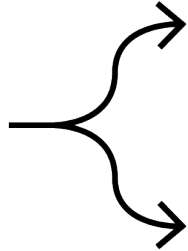
## CURRENT PROJECT



ADF App



Backup Folder



ADF FILES



Backup Files



## DOWNSTREAM (FUTURE)



Triple Store



Backup File Storage



Backup Search Tool



Backup Visualizers



# Thank You/Questions?

