

# NMR Adapter Update

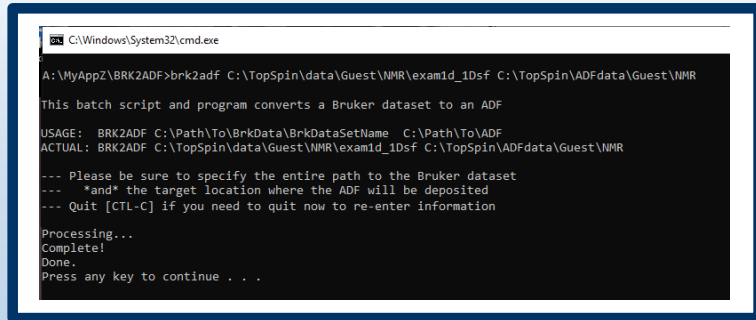
- A collaboration between Boehringer Ingelheim & Pfizer

October, 2019



# NMR Converter – Bruker to ADF

- There is fully-functional NMR converter to convert Bruker NMR data to ADF
  - Supports: TopSpin, XWIN-NMR, UXNMR, UXNMR/P data
- Jointly developed by Boehringer Ingelheim & Pfizer
- MestreLab is taking the BI-Pfizer converter and is carrying it forward.



```
C:\Windows\System32\cmd.exe

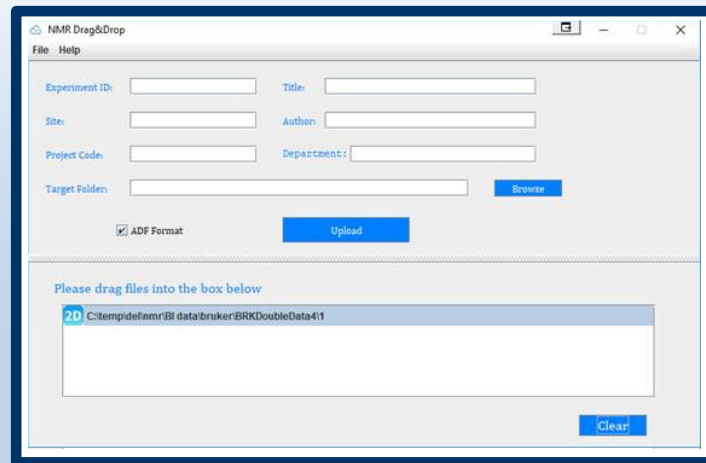
A:\MyAppZ\BRK2ADF>brk2adf C:\TopSpin\data\Guest\NMR\exam1d_1DsF C:\TopSpin\ADFdata\Guest\NMR

This batch script and program converts a Bruker dataset to an ADF

USAGE: BRK2ADF C:\Path\To\BrkData\BrkDataSetName C:\Path\To\ADF
ACTUAL: BRK2ADF C:\TopSpin\data\Guest\NMR\exam1d_1DsF C:\TopSpin\ADFdata\Guest\NMR

--- Please be sure to specify the entire path to the Bruker dataset
--- *and* the target location where the ADF will be deposited
--- Quit [CTL-C] if you need to quit now to re-enter information

Processing...
Complete!
Done.
Press any key to continue . . .
```



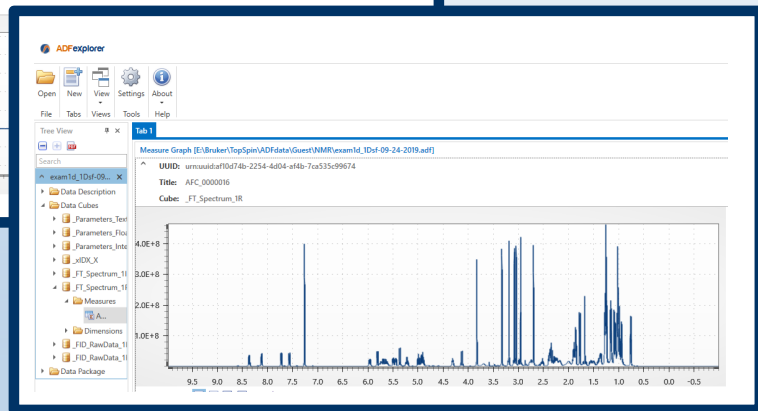
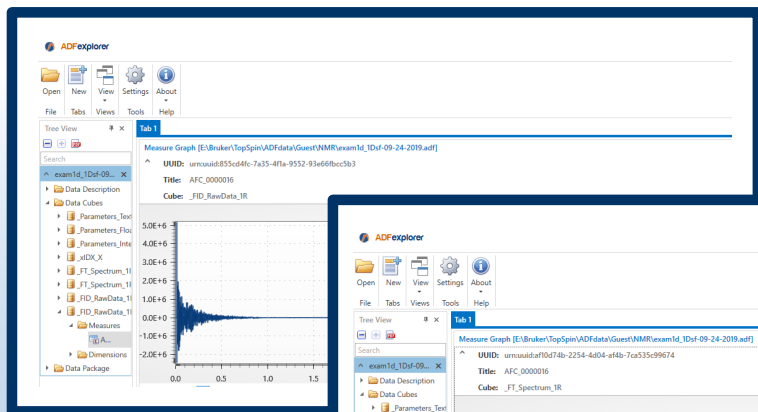
Command-line or GUI mode!



# Data Cubes & Data Package



- Raw Data (FID's) and Processed Data (Spectra) in the Data Cubes
- Proprietary-format vendor files in the Data Package



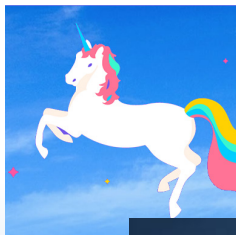
```
f92ff00-9f51-49a2-bad4-9a8cb9e7de8_uxnmr.par - Notepad
File Edit Format View Help
##TITLE= Parameter file, TOPSPIN                               Version
##JCAMPODX= 5.0
##DATATYPE= Parameter Values
##NPOINTS= 193 $$ modification sequence number
##ORIGIN= Bruker BioSpin GmbH
##OWNER= nmrsu
$$ 2007-09-17 17:33:19.750 +0200 Administrator@APOLLON
$$ C:\Bruker\ts21b10\conf\instr\spect\uxnmr.par
$$ process C:\Bruker\ts21b10\prog\mod\hconfserver.exe
##$ACB= 0
##$ACBTAT= (0..15)
0 66 66 68 50 0 0 0 0 0 0 0 0 0 0 0
##$ACBTBA= (0..15)
0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
##$ACBTLK= (0..15)
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##$ACBCTL= (0..15)
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##$ACBTDI= (0..15)
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##$ACBTDI= (0..15)
0 102 102 104 73 0 0 0 0 0 0 0 0 0 0 0
##$ACBTDI= (0..15)
0 10 10 10 0 0 0 0 0 0 0 0 0 0 0 0
##$ACBTFHI= (0..15)
0 600 600 243 0 0 0 0 0 0 0 0 0 0 0 0
```



# Data Description

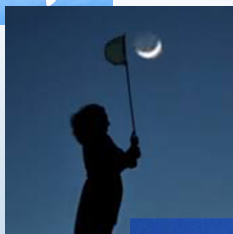
Using a monolithic full graph ontology to describe all NMR metadata is ...

➤ Unrealistic



13,589 possible unique metadata values as of July 2019 and this does not include up to thousands more unique metadata values related to peaks, assignments, etc.

➤ Unachievable



Varian/Agilent & JEOL metadata not yet included in the model. Most vendor-specific hardware & software metadata are irreconcilably incompatible

➤ Unsustainable



Deprecated values are repurposed & new values are continually added; definitions, datatypes & relationships evolve over time

# Data Description

... this is further complicated by the fact that all NMR metadata are not created equal ...

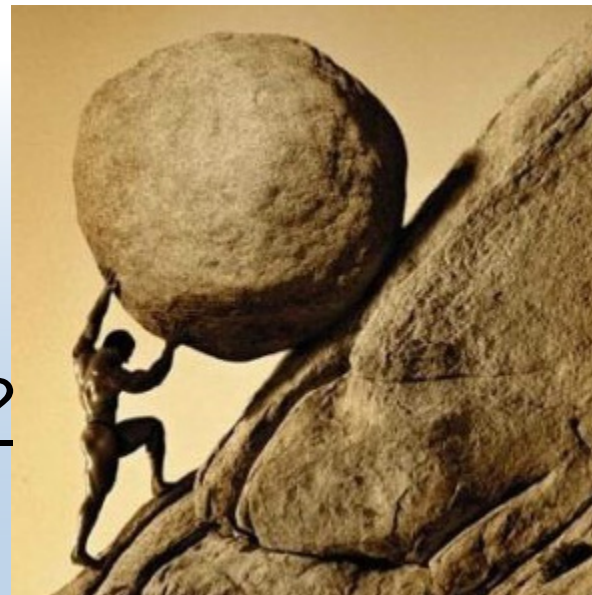
- Cross-vendor  Approximately 80 values
- Vendor-specific  Approximately 13,500 values
- Other Company  Approximately 10 values

# Data Description

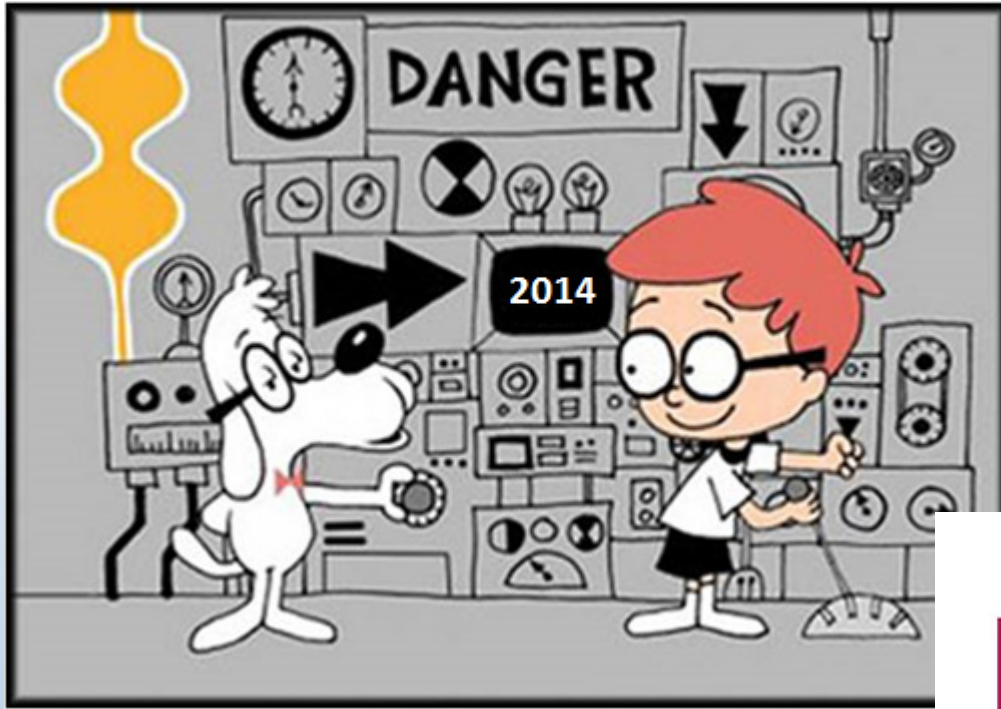
... so trying to force-fit a very large volume of diverse, fundamentally irreconcilable, evolving, vendor-specific metadata into a single monolithic NMR ontology does not seem like a good use of valuable resources.

*What are we to do?*

*How can we deliver value today?*



Long, long ago and (not so) far away....



# Allotrope from the WayBack Machine (2014)

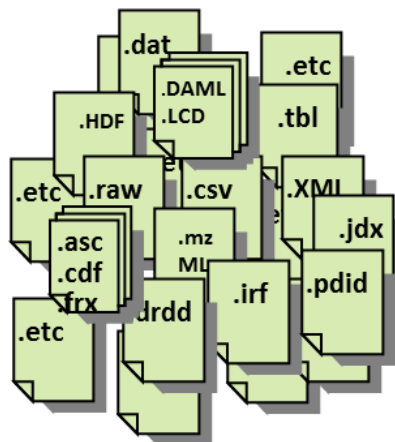
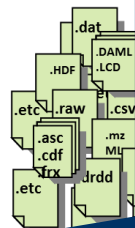
## What is Allotrope Creating?



Standard  
File Format

File format for any  
technique or instrument

File format  
technique or i



© 2014 Allotrope Foundation

AF0045674	NMR Characterization	DRX600-SN/10234567
AF0034558	IR Fingerprinting	IS10 FTIR/-SN/341980





# Allotrope from the WayBack Machine (2014)

## What is Allotrope Creating?



Open  
Metadata  
Repository

Standard vocabulary &  
structure for metadata

Unit that enables  
of the standards &  
metadata in software  
development

Project	Test	Instrument
AF 0012354	IR Fingerprinting	QC Lab #33B 380 FT-IR
AF0012764	Bulk & Tapped Density	ASTM Standard Sieve #6
AF 12989	NMR Characterization	AM500
	Tapped & Bulk Density	Sieve XXX
AF0045674	Characterization with	Nouvelle DRX600
AF 0034558	IR	iS10 FT-IR

With the Metadata Repository

Project	Test	Instrument
AF0012354	IR Fingerprinting	380 FTIR/-SN/145453
AF0012764	Bulk and Tapped Density	ASTM Sieve-SN/3452
AF0012989	NMR Characterization	AM500-SN/0034578
AF0013142	Bulk and Tapped Density	ASTM Sieve-SN/09783
AF0045674	NMR Characterization	DRX600-SN/10234567
AF0034558	IR Fingerprinting	iS10 FTIR/-SN/341980

Standard Vocabulary &  
Metadata

Instrument
QC Lab #33B 380 FT-IR
ASTM Standard Sieve #6
AM500
Sieve XXX
Nouvelle DRX600
iS10 FT-IR

the Metadata Repository

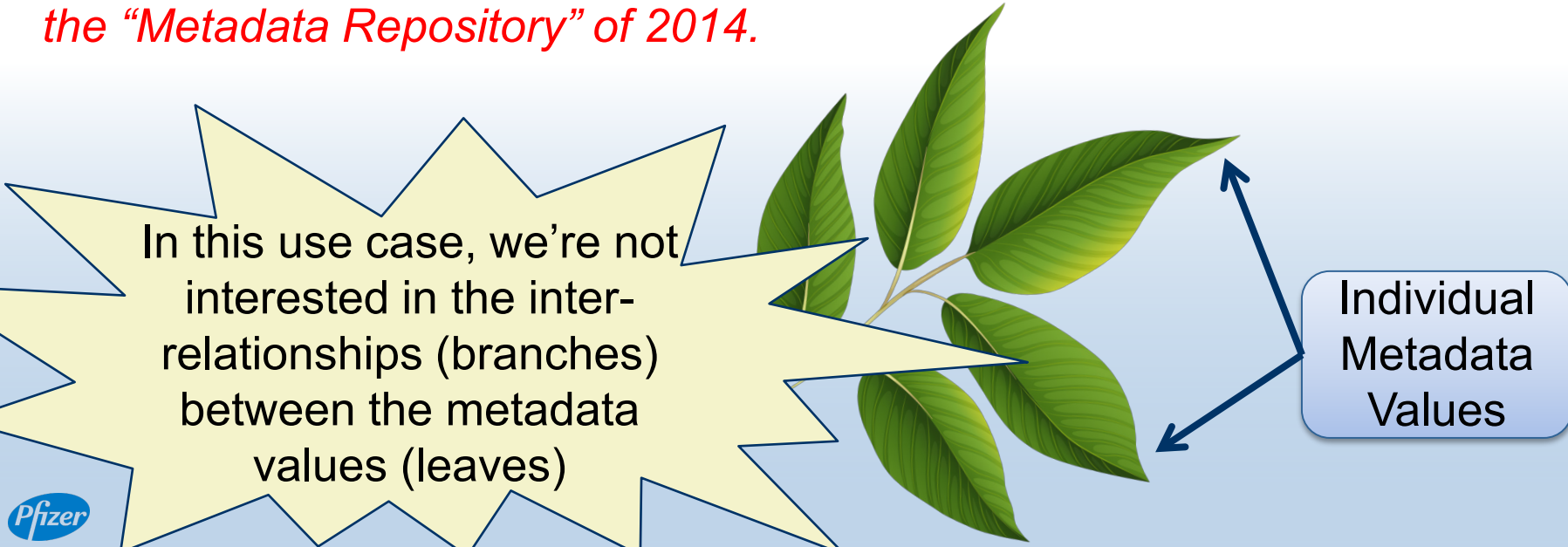
Instrument
380 FTIR/-SN/145453
ASTM Sieve-SN/3452
AM500-SN/0034578
ASTM Sieve-SN/09783
DRX600-SN/10234567
iS10 FTIR/-SN/341980

© 2014 Allotrope Foundation



# Leaf Nodes in ADF

- We can use greatly simplified semantic constructs called “leaf nodes” to represent this metadata in a **standardized**, highly **structured** way so that the ambiguity is greatly reduced – *in effect delivering on the promise of the “Metadata Repository” of 2014.*



In this use case, we're not interested in the inter-relationships (branches) between the metadata values (leaves)

Individual  
Metadata  
Values

# Before ADF

- Each NMR experiment has thousands of instrument settings that are represented in the original data something like this:

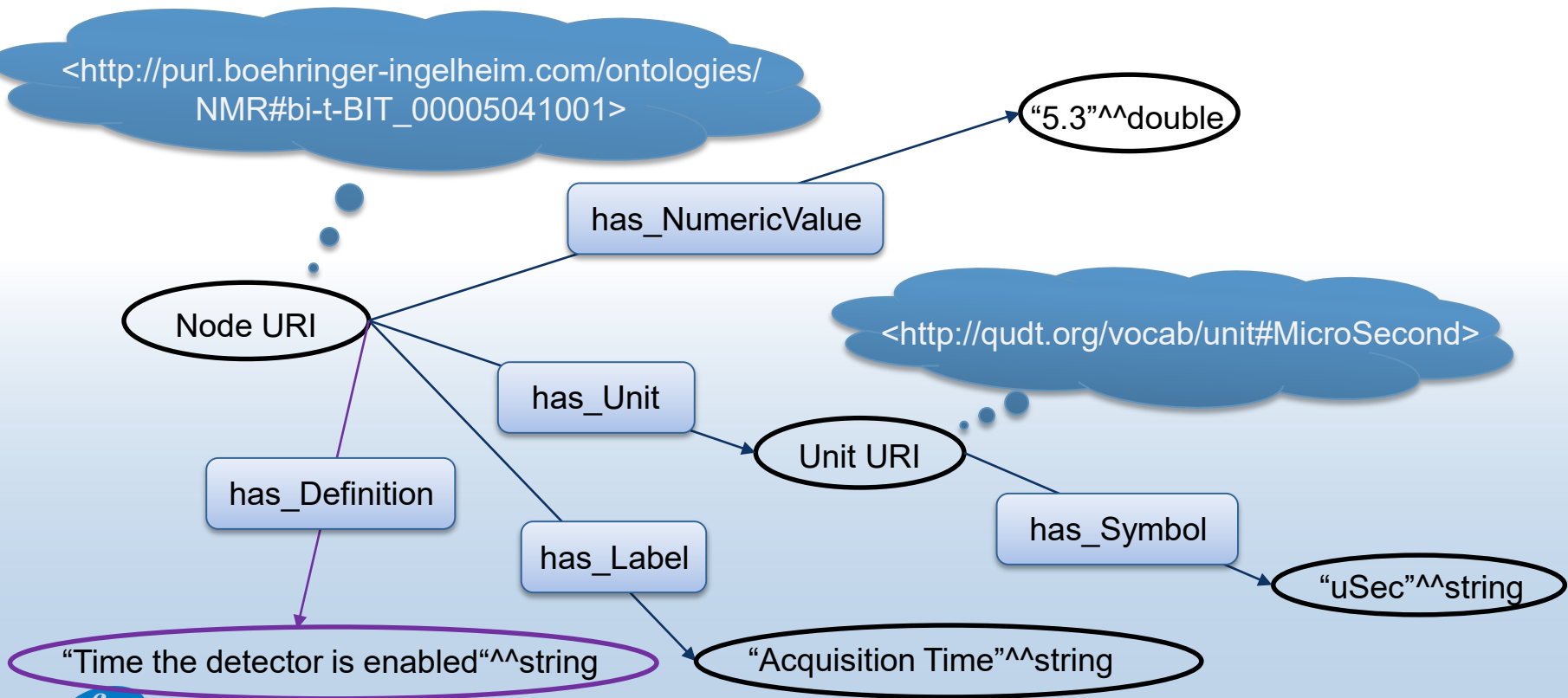
**Acquisition Time = 5.3**

- *What does “Acquisition Time” really mean?*
- *Is the “Acquisition Time” on one instrument the same thing as another?*
- *What are the units?*
- *What kind of a number is it?*
- *Is it a number at all, or is it really a text string that sometime has letters?*



# Anatomy of a Leaf Node (simplified view)

Acquisition Time = 5.3



# Anatomy of a Leaf Node (simplified view)

Non-Allotrope data usually has just this

<http://purl.boehringer-ingelheim.com/ontologies/  
NMR#bi-t-BIT\_00005041001>

"5.3"^^xsd:double

has\_NumericValue

Node URI

has\_Unit

<http://qudt.org/vocab/unit#MicroSecond>

Unit URI

has\_Definition

has\_Symbol

has\_Label

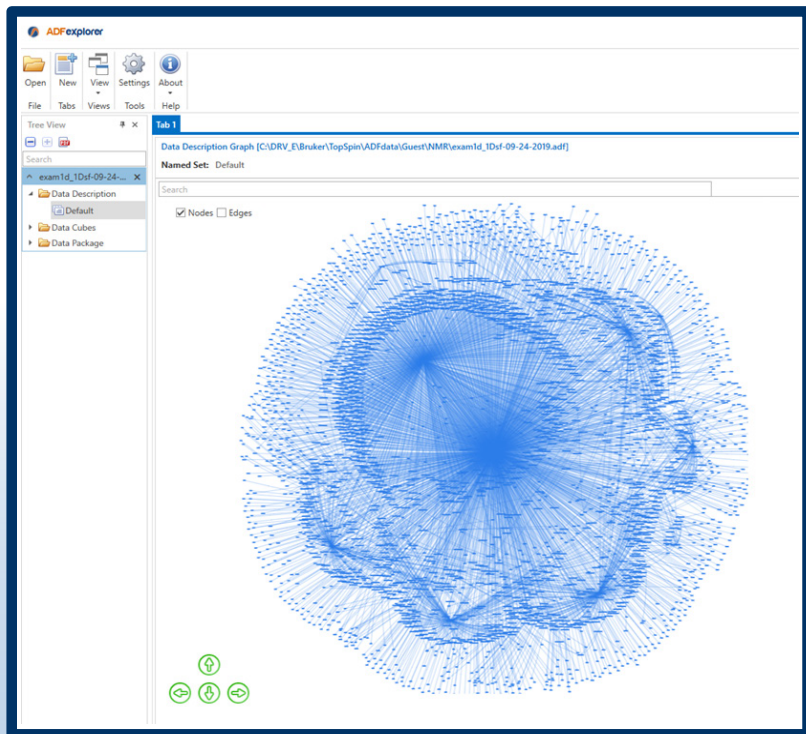
"uSec"^^string

"Time the detector is enabled"^^string

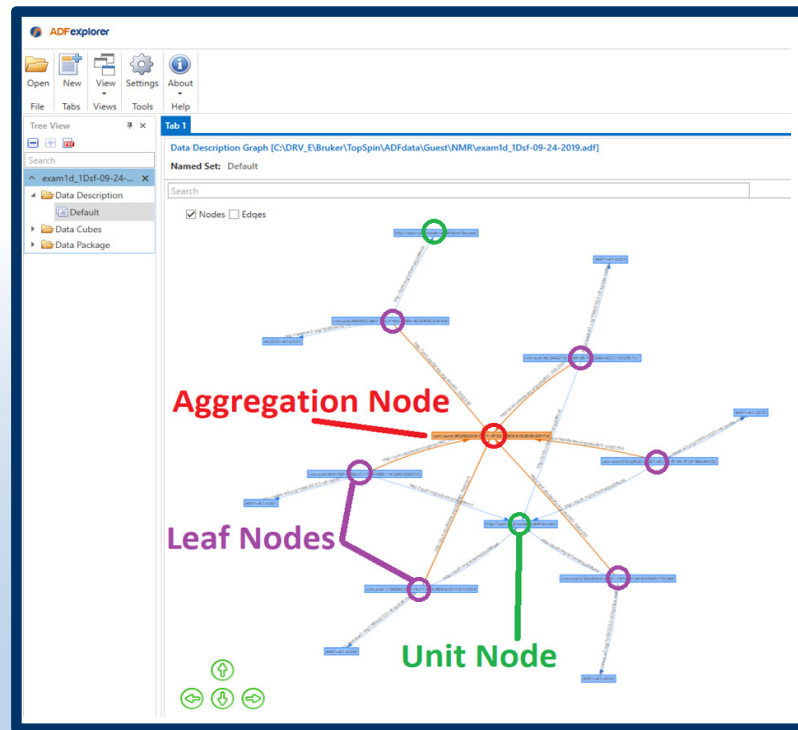
"Acquisition Time"^^string



# What does this look like in a real NMR example?



We have a lot of metadata!



A very simple case (6 LN)



# The Net Result

- Leaf nodes provide a structured, lightweight, Allotrope-compliant way to instantiate NMR metadata in the Data Description
  - This still allows us to use the Allotrope “semantics infrastructure” to automatically apply constraints to the allowed values in an ADF and to execute standardized querying & validation of the metadata via the Allotrope API’s
  - Leaf nodes can be implemented to yield complete, functional ADF’s in days compared to the months... years... it can take to develop comprehensive, usable ontologies
- By a judicious choice of where the terms come from (Allotrope, Vendor or Member Company) we can simultaneously and easily accommodate the different types of metadata that are present in real-world NMR experiments (Cross-vendor, Vendor-Specific and Member Company)

# KISS and the NMR ADF



So tying this all together into a process to permit rapid creation of ADFs with a minimum of overhead ... **and delivering value at each step along the way** ...



Phase 1: Instantiate all metadata values in ADF via a controlled vocabulary and a Member Company (or other suitable placeholder) URI namespace using a leaf node construct for speed & ultimate flexibility.

- Phase 2: Promote all terms (except Member Company-specific ones) to a controlled vocabulary with a vendor-based URI namespace: ret simple leaf-node construct.

**(99% of all terms will stop here)**





# KISS and the NMR ADF



- Phase 3: Identify the small number of “universal” terms with community input; utilize community SME’s to draft consensus PrefLabels & ScopeNote definitions; Submit SME-curated universal terms to Allotrope governance for AFT reconciliation and ultimate promotion to an Allotrope namespace. Instantiate Allotrope-governed universal terms in ADF as leaf nodes.
- Phase 4: Elaborate relationships between “universal” terms only as needed to support *bona fide* use cases. Focused ontologies are only created out of a clearly and cogently articulated need in order to deliver demonstrable and quantifiable value to end users. *At all costs avoid unrewarded complexity and effort to create ontologies simply to support “what if...” use cases.*

# The NMR Recipe for Incremental and Iterative Value



2016-17

- Fill data cubes & data package; metadata literals in data description (DD)  
✓ Value: All data + metadata in enterprise in single, consistent, vendor-independent format

2019

- Upgrade DD literals to member-namespace leaf nodes (LN)  
✓ Value: Use semantics infrastructure of Allotrope API's to define, validate, query metadata
- Upgrade member-namespace LN to vendor-namespace LN  
✓ Value: Vendors get full control over metadata that apply only to their hardware/software
- Upgrade “universal” vendor-namespace LN to “universal” Allotrope LN  
✓ Value: Achieve vision of truly vendor-independent universal metadata (applies to all vendors)
- Build ontology to support use cases from “universal” Allotrope terms  
Pfizer Value: Data mining & advanced analytics

# Questions?

---

