LA-UR-16-26794
Approved for public release; distribution is unlimited.

Title:	Efficient Workflows for Curation of Heterogeneous Data Supporting Modeling of U-Nb Alloy Aging
Author(s):	Ward, Logan Timothy Hackenberg, Robert Errol
Intended for:	This is a presentation summarizing a graduate research associates summer project. This request is to clear it to be presented to his thesis advisor and research group at his home institution (Northwestern University).
Issued:	2016-09-07

Efficient Workflows for Curation of Heterogeneous Data Supporting Modeling of U-Nb Alloy Aging

Logan T. Ward, Robert E. Hackenberg

Sigma Division

Los Alamos National Laboratory

31 August 2016

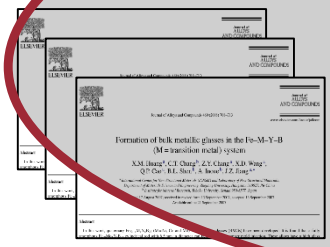
Los Alamos NATIONAL LABORATORY EST. 1943

National Nuclear Security Administration

Data Challenges in Materials

Materials Informatics: *Extract Knowledge from Materials Data*

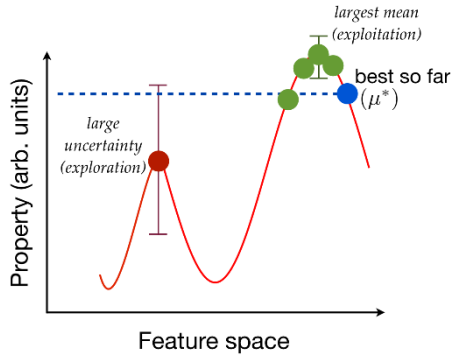
Gathering Data



Building Models

$$f(Z_A, Z_B)$$

Uncertainty Quantification



Balachandran *et al*. Sci Rprts. (2016)

Computing Infrastructure

hpc.lanl.gov

Aging in U-Nb Alloys

U-Nb alloys become hard and brittle with aging

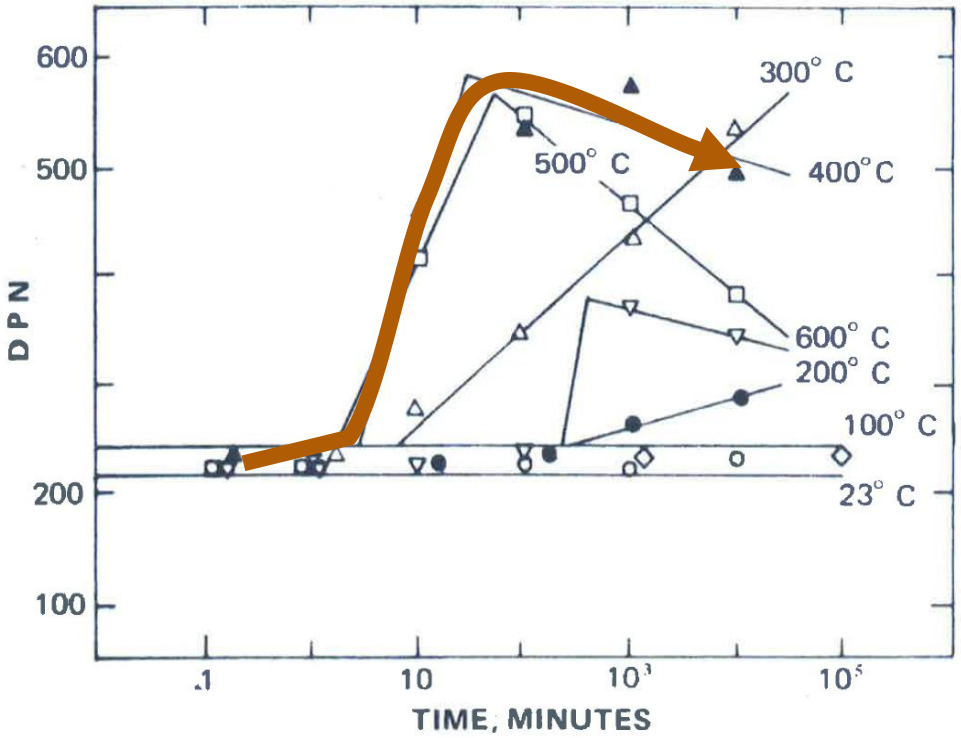


Figure G10-a. Room Temperature Hardness Value for the U-4.5 wt% Nb Alloy. The samples were held 30 minutes at 850°C and then isothermally aged for the indicated times and temperatures.

Our Goal: Be able to predict hardening

Building an Aging Model

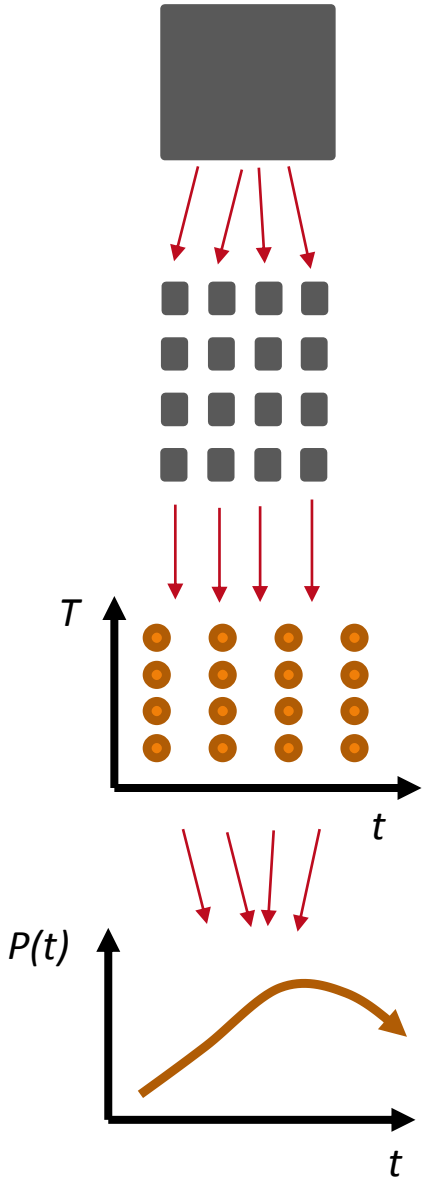
Typical Aging Experiment:

1. **Cast** a single batch of alloy
2. **Divide** into smaller samples
3. **Age** them for different times, Temp.
4. **Fit** a property model

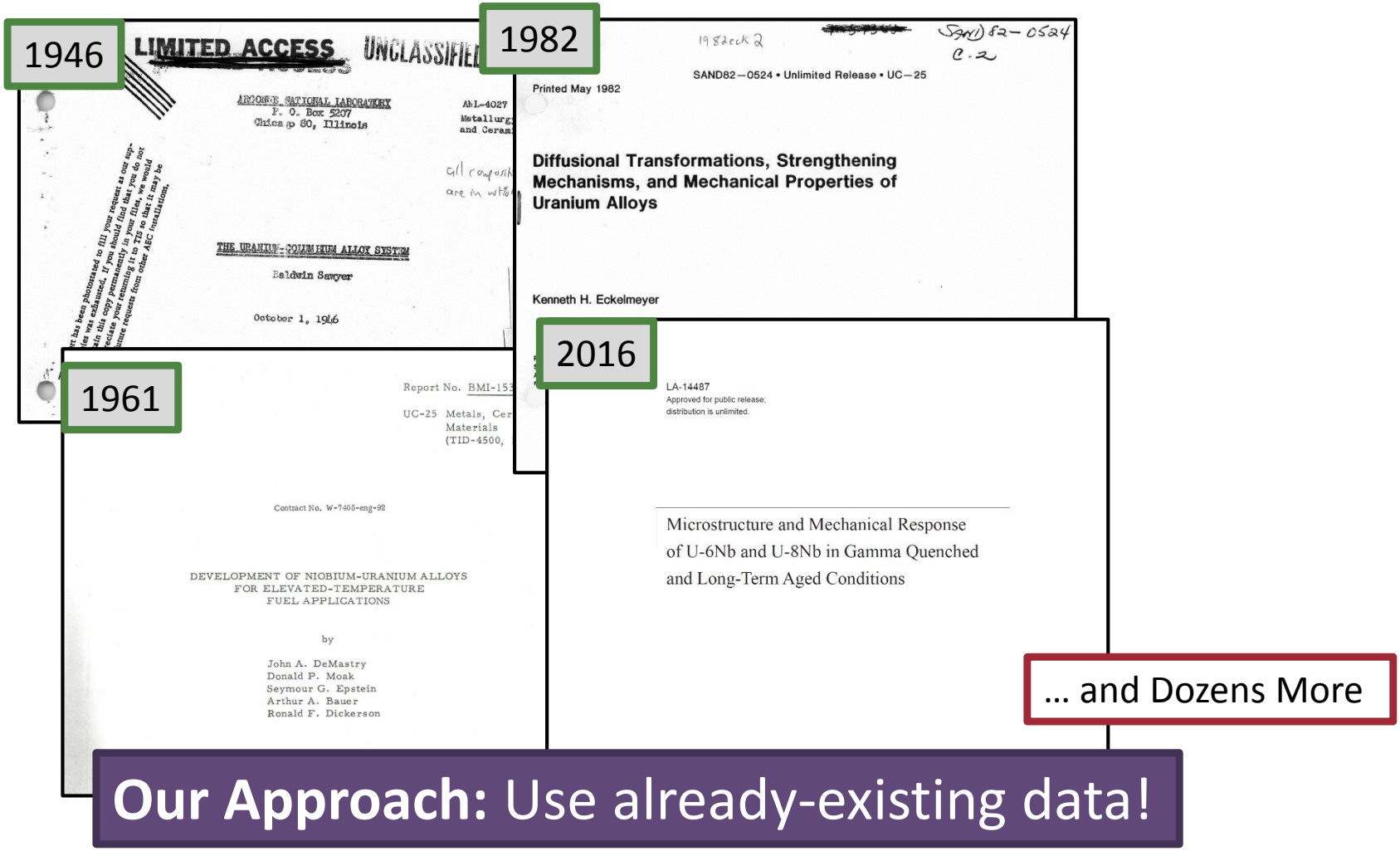
A few aging studies are practical,
but there are many factors to consider…

- Alloy composition
- Grain size
- Statistical variation
- *… enough to become costly*

Problem: Cannot run all the experiments
Potential Solution: The literature!



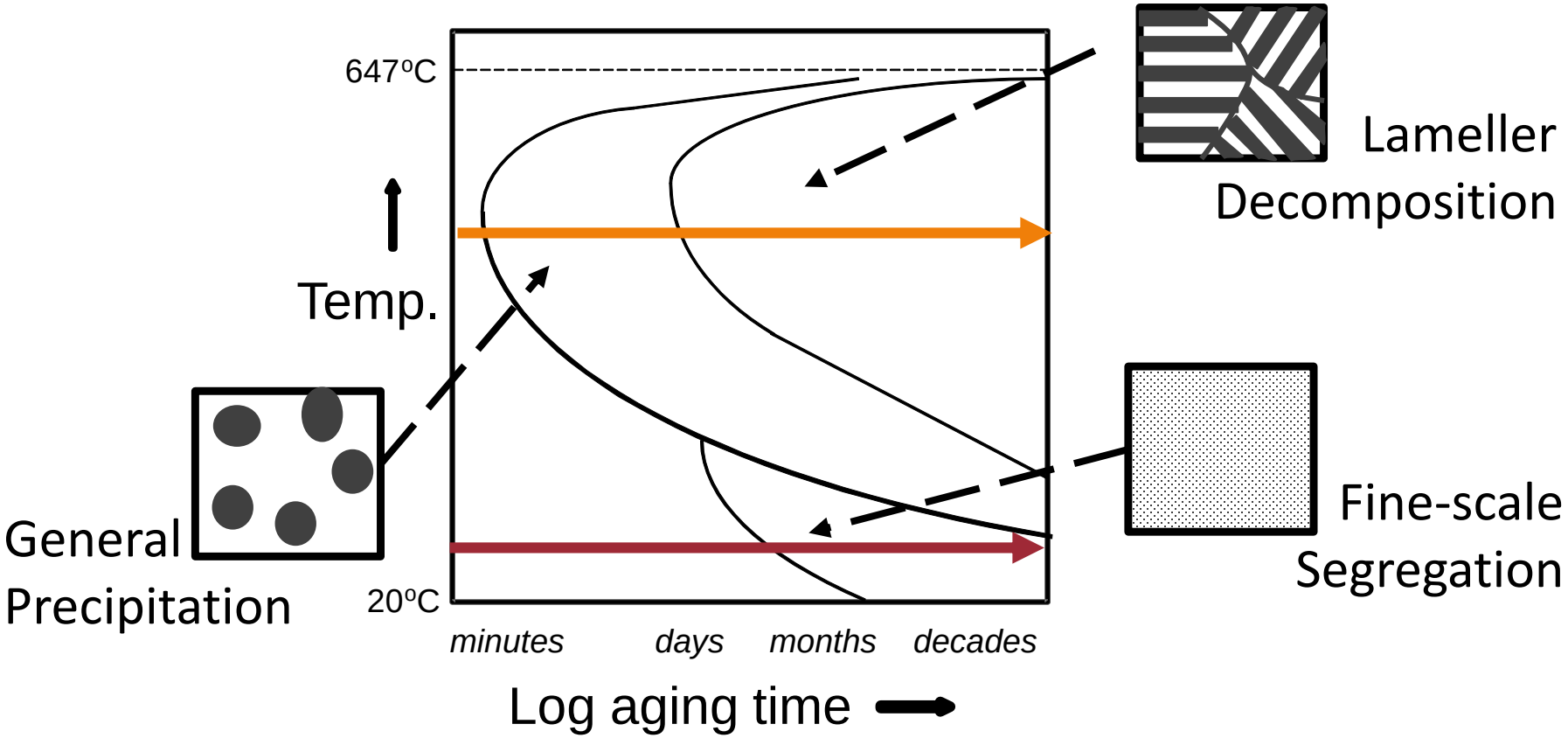
The Literature



Our Approach: Use already-existing data!

Different Phase Trans. in U-Nb

Problem: Different aging mechanisms, require different models



Another Need: Images of aged microstructure

The Challenge

Input:
Papers

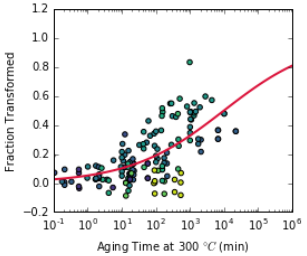
How do we store the data?

Images

What is the μ-structure?

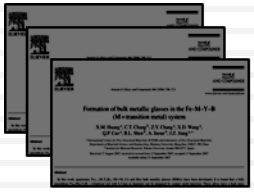
Hardness

Which data is relevant?



Output:
Aging Model

Storing Materials Data



Input: *Papers*

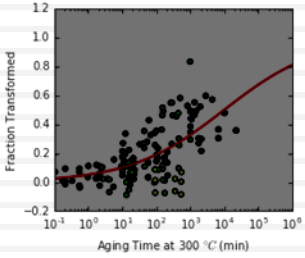
How do we store the data?

Images

What is the μ-structure?

Hardness

Which data is relevant?



Output: *Aging Model*

Example Data Source

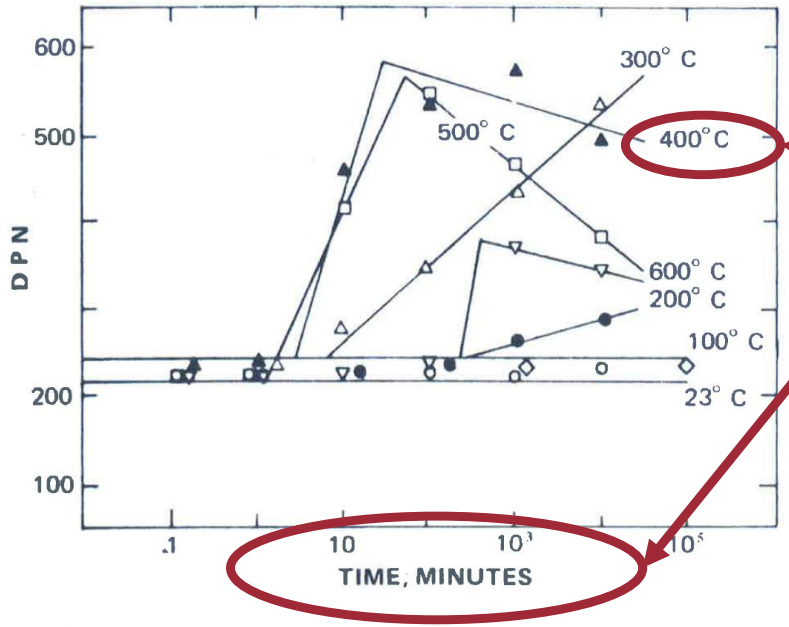


Figure G10-a. Room Temperature Hardness Value for the U-4.5 wt% Nb Alloy. The samples were held 30 minutes at 850°C and then isothermally aged for the indicated times and temperatures.

Main Variables:
- ✓ Aging Time
- ✓ Aging Temp.
- ✓ Composition

Other Considerations:
- Who did this?
- Processing history
- Banding
- …?

Two problems:
- How to organize this information?
- How to store it?

Organizing Data: What is a Schema?

Definition of Data Format / Structure

Why is this important?

- Simplify data processing software
- Ensure all relevant data is captured

What kinds of questions does a schema answer?

- What is the relevant data?

  "Hardness measurements have a scale and value"
- How are they related to each other?

  "Solutionizing is a processing step"
- What are the format requirements?

  "Temperature must be a real number"

What does a “XML Schema” look like?


```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" attributeFormDefault="unqualified" elementFormDefault="unqualified">
  <!--  Imported types  -->
  <xsd:include schemaLocation="physical-quantity.xsd"/>
  <!--  Custom types  -->
  <xsd:simpleType name="hardness-measurement-method-type">
    <xsd:restriction base="xsd:string">
      <xsd:enumeration value="vickers"/>
      <xsd:enumeration value="knoop"/>
      <xsd:enumeration value="brinell 3000kg"/>
      <xsd:enumeration value="brinell 500kg"/>
      <xsd:enumeration value="rockwell a"/>
      <xsd:enumeration value="rockwell b"/>
      <xsd:enumeration value="rockwell c"/>
      <xsd:enumeration value="rockwell d"/>
      <xsd:enumeration value="rockwell e"/>
      <xsd:enumeration value="rockwell f"/>
      <xsd:enumeration value="rockwell 15-N"/>
      <xsd:enumeration value="rockwell 30-N"/>
      <xsd:enumeration value="rockwell 45-N"/>
      <xsd:enumeration value="rockwell 30-T"/>
      <xsd:enumeration value="shore"/>
    </xsd:restriction>
  </xsd:simpleType>
  <!--  Actual Schema  -->
  <xsd:complexType name="hardness-measurement-type">
    <xsd:sequence>
      <xsd:element name="measurement-type" type="hardness-measurement-method-type"/>
      <xsd:element name="measurement" type="physical-quantity-type"/>
      <xsd:element name="applied-load" minOccurs="0" maxOccurs="1" type="xsd:double"/>
      <xsd:element name="microindenter" minOccurs="0" type="xsd:boolean"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:schema>
```

Import from other schemas

Hardness Measurement

Make fields optional

Enforce data types

Our Data Schema: Nice & Simple


```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:include schemaLocation="hardness-type.xsd"/>
  <xsd:include schemaLocation="note-type.xsd"/>
  <xsd:include schemaLocation="xray-diffraction-type.xsd"/>
  <xsd:include schemaLocation="material-composition.xsd"/>
  <xsd:include schemaLocation="material-thermomechanical-processing-type.xsd
  <xsd:include schemaLocation="citation-type.xsd"/>
  <xsd:include schemaLocation="micrograph-type.xsd"/>
  <xsd:element name="literature-data">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element minOccurs="0" name="id" type="xsd:string"/>
        <xsd:element minOccurs="0" name="citation" type="citation-type"/>
        <xsd:element maxOccurs="unbounded" minOccurs="0" name="note" type="note-type"/>
        <xsd:element minOccurs="0" name="material">
          <xsd:complexType>
                <xsd:annotation>
                  <xsd:documentation>Who made this alloy</xsd:documentation>
                </xsd:ann
              </xsd:eleme
              <xsd:elemen                                    pe="material-composition"/>
              <xsd:elemen                                    ype="material-composition"/>
              <xsd:element minOccurs="0" name="processing" type="material-thermomechanical-processing-type"/>
            </xsd:sequence>
          </xsd:complexType>
        </xsd:element>
        <xsd:choice maxOccurs="unbounded" minOccurs="0">
          <xsd:element name="hardness" type="hardness-type"/>
          <xsd:element name="diffraction" type="xray-diffraction-type"/>
          <xsd:element name="micrograph" type="micrograph-type"/>
        </xsd:choice>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

No “reinventing wheel”

Bottom Line: *Schemas Are Not The Hardest Step*

S hole
on process

☺ Last Slide with XML ☺

Can fit micrographs, XRD patterns, and hardness

Storing Data: MDCS

Materials Database: Materials Data Curation System

https://github.com/usnistgov/MDCS

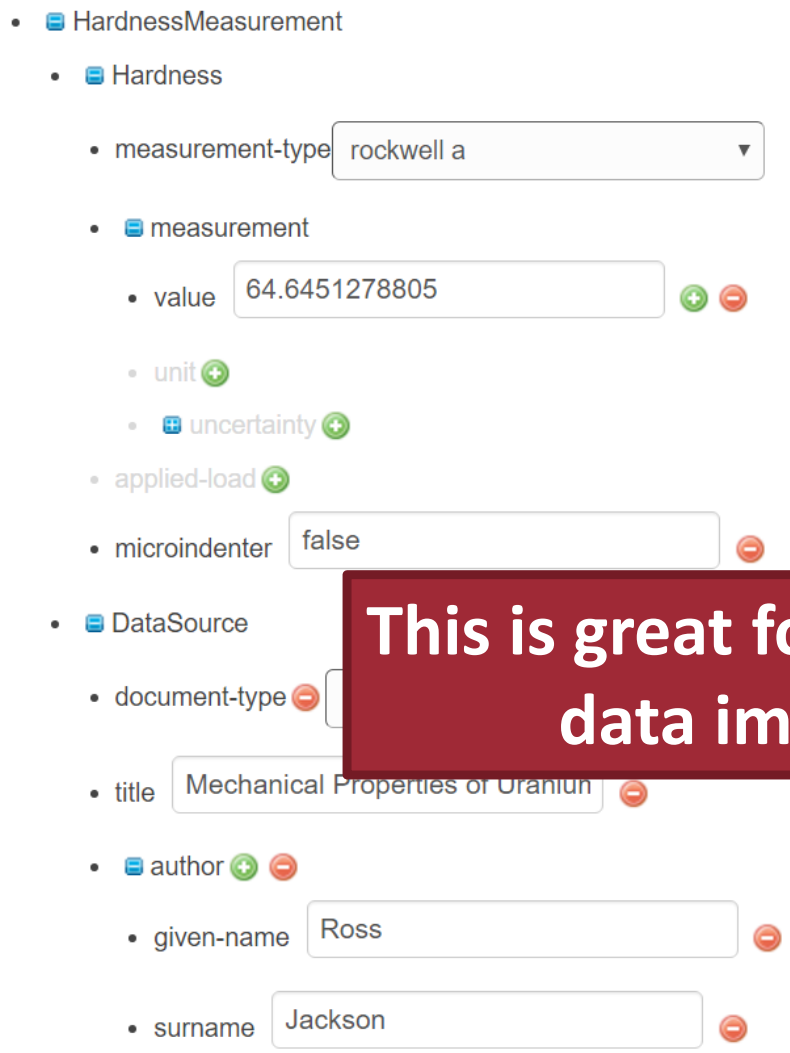
Advantages:

- ✓ Already built / under active development
- ✓ Browser-based inputting, editing, viewing data
- ✓ REST interface for integrating with other tools

To-Do List:

- ✓ Designing schema
- × Efficiently importing data

Problem w/ MDCS: Slow Data Entry



MDCS Automatically Generates Data Entry Forms

This is great for a few data points, but… data import via form is slow

Getting Literature Data into MDCS

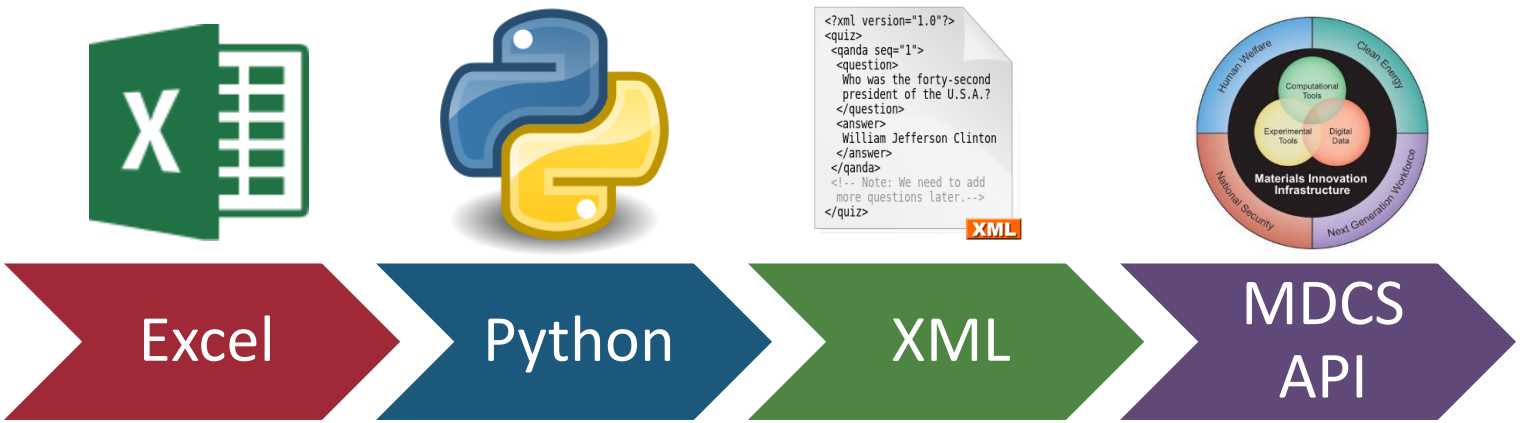
Goals:

- Provide a simple import method

  No writing XML by hand!

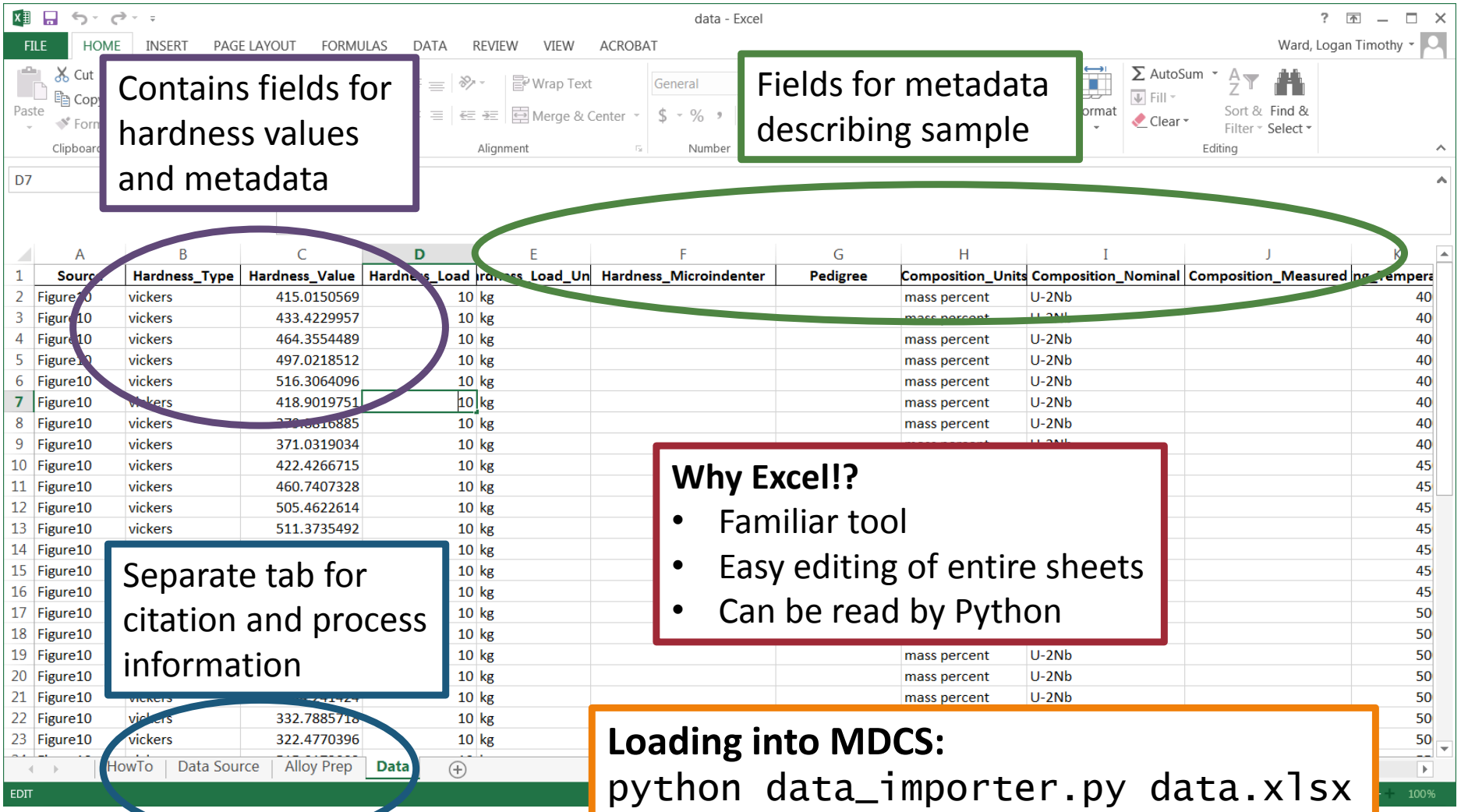
- Hide the MDCS API in a black box

Solution: Excel + Python → XML → MDCS



Logos from Wikimedia Commons, github.com/usnistgov/MDCS

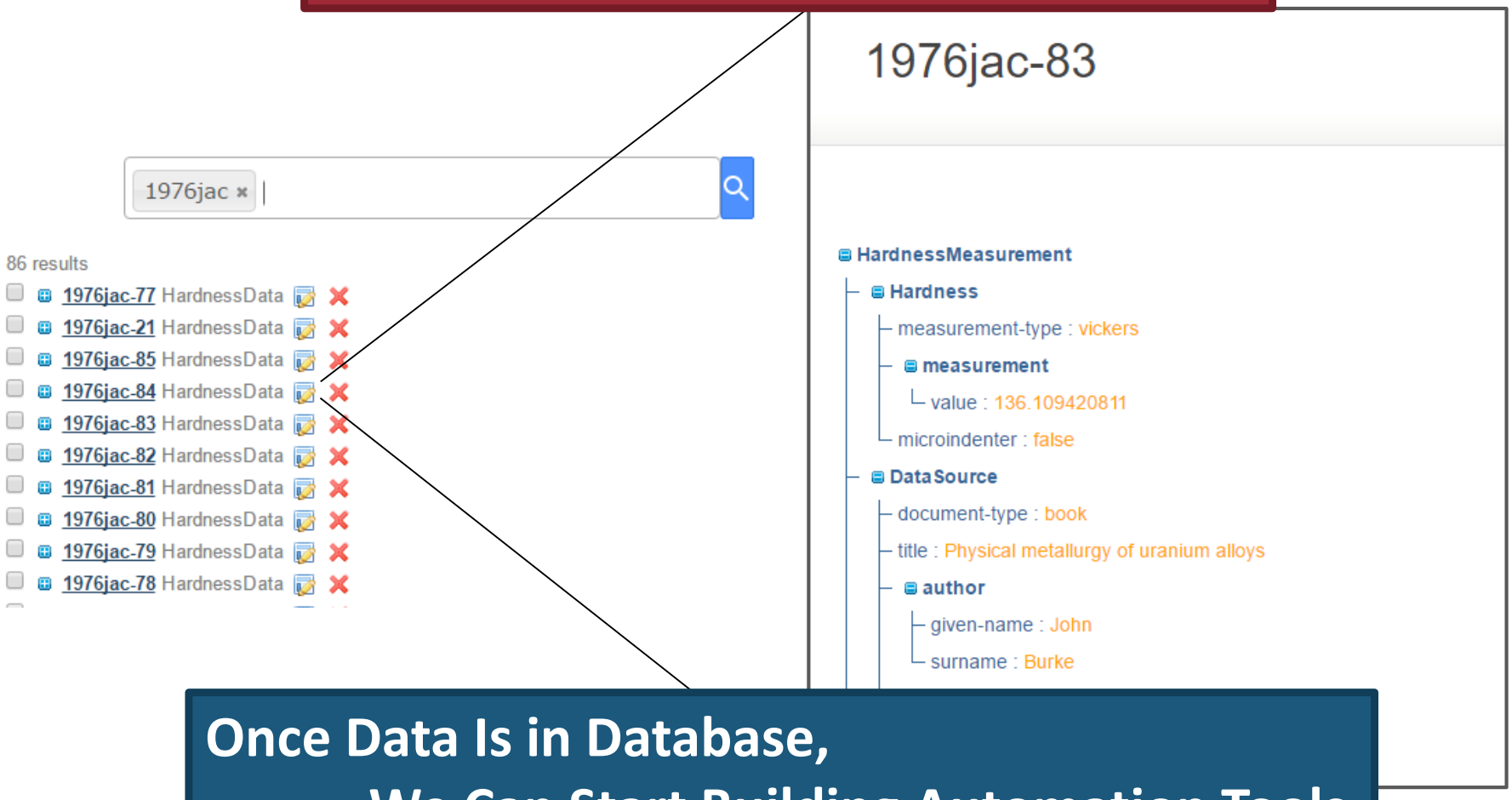
Staging Data in Excel Document



	A	B	C	D	E	F	G	H	I	J	K
1	Source	Hardness_Type	Hardness_Value	Hardness_Load	Hardness_Load_Un	Hardness_Microindenter	Pedigree	Composition_Units	Composition_Nominal	Composition_Measured	ng_Tempera
2	Figure10	vickers	415.0150569	10	kg			mass percent	U-2Nb		40
3	Figure10	vickers	433.4229957	10	kg			mass percent	U-2Nb		40
4	Figure10	vickers	464.3554489	10	kg			mass percent	U-2Nb		40
5	Figure10	vickers	497.0218512	10	kg			mass percent	U-2Nb		40
6	Figure10	vickers	516.3064096	10	kg			mass percent	U-2Nb		40
7	Figure10	vickers	418.9019751	10	kg			mass percent	U-2Nb		40
8	Figure10	vickers	378.8816885	10	kg			mass percent	U-2Nb		40
9	Figure10	vickers	371.0319034	10	kg						40
10	Figure10	vickers	422.4266715	10	kg						45
11	Figure10	vickers	460.7407328	10	kg						45
12	Figure10	vickers	505.4622614	10	kg						45
13	Figure10	vickers	511.3735492	10	kg						45
14	Figure10			10	kg						45
15	Figure10			10	kg						45
16	Figure10			10	kg						45
17	Figure10			10	kg						50
18	Figure10			10	kg						50
19	Figure10			10	kg			mass percent	U-2Nb		50
20	Figure10			10	kg			mass percent	U-2Nb		50
21	Figure10			10	kg			mass percent	U-2Nb		50
22	Figure10	vickers	332.7885718	10	kg						50
23	Figure10	vickers	322.4770396	10	kg						50

Final Result: MDCS Records

Paper to Database: 10-30 min per figure (or table)



Once Data Is in Database,
We Can Start Building Automation Tools

Analyzing Image Data

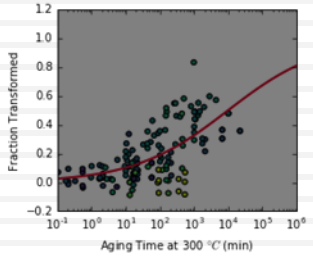
Input: *Papers*

Images

What is the μ-structure?

Hardness

Which data is relevant?

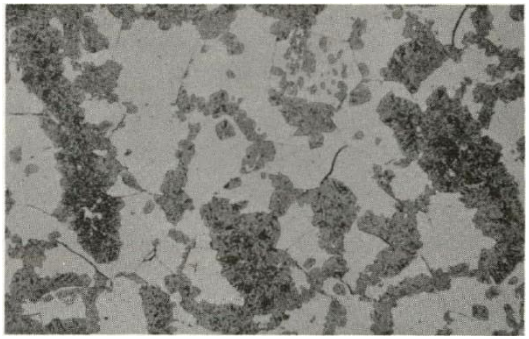


Output: *Aging Model*

Process for Making TTT Diagram

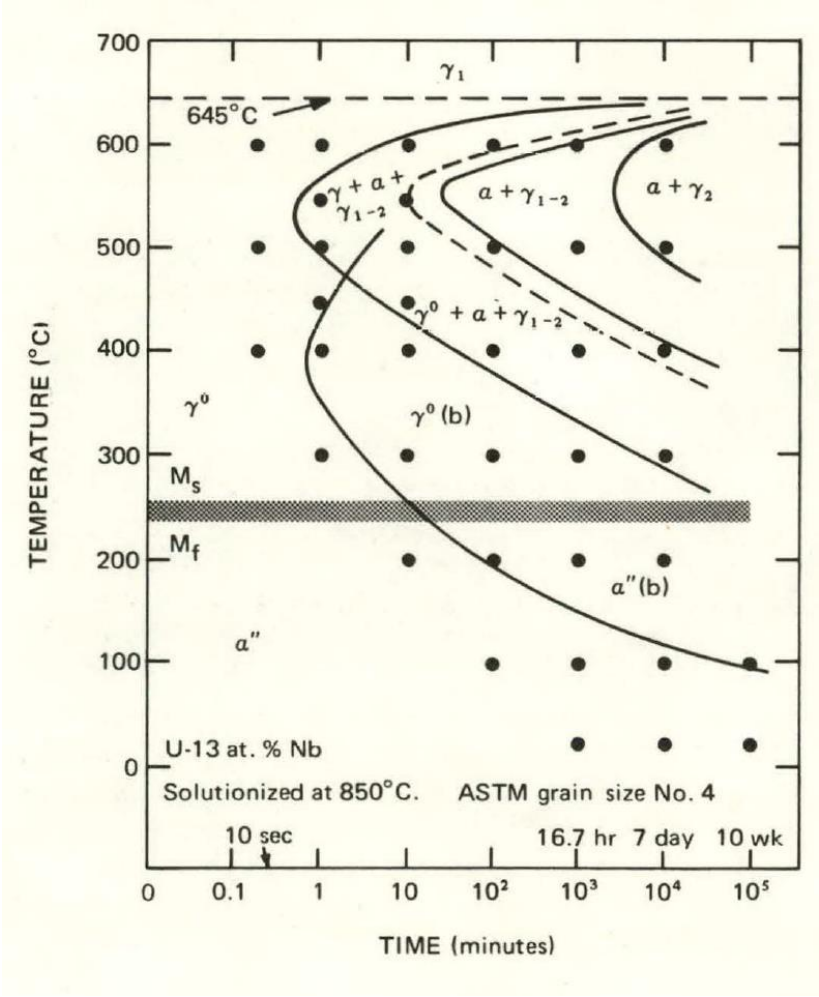
Procedure:

1. Age alloys at many times and temps
2. Evaluate fraction of transformed phase
3. Draw dividing boundaries



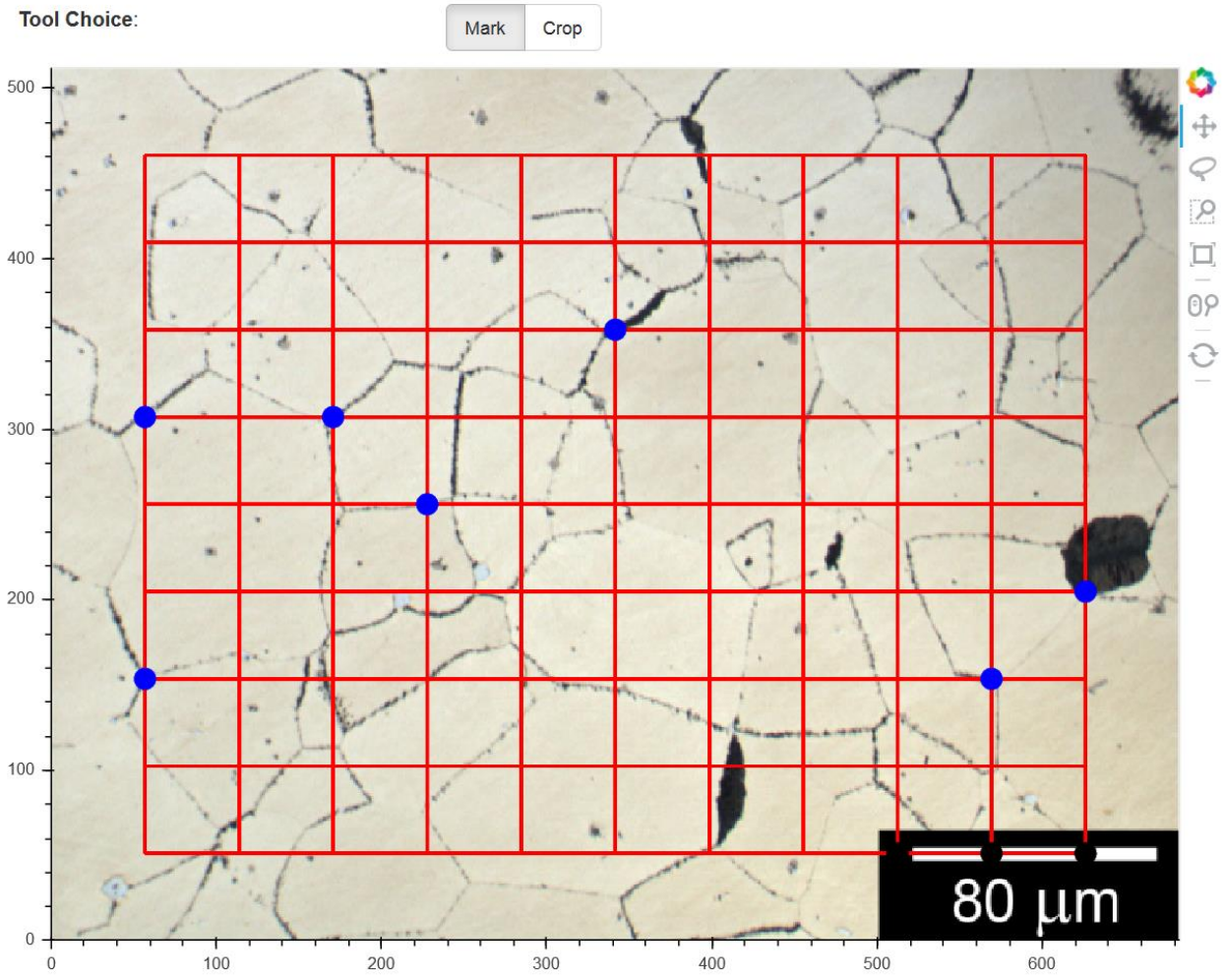
Ref: R. Jackson. RFP-1609.

Goal: Create software to automate/assist this process



Ref: R. Jackson. RFP-1609.

Bottleneck #1: Image Analysis



Fitting a TTP Boundary

Current State: Many images, and fraction transformed

What We Need: $f(T, t, x_{Nb})$

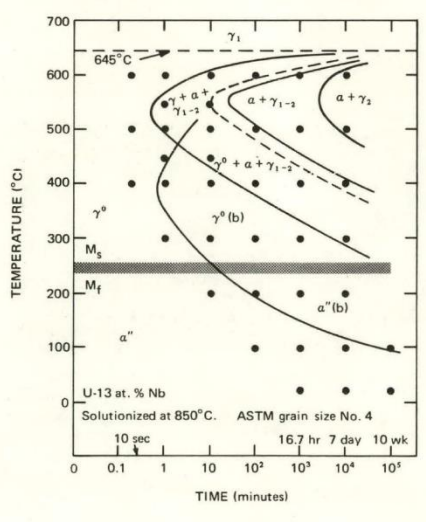
We can fit a TTT curve with *all* our data!

Problem: Conventionally, all experiments have same x_{Nb}

Solution: Combing Kirkaldy and JMAEK Models

Kirkaldy: $t_{f=\beta\%} \propto \frac{\Delta T^{-3}}{2^{N/8}} e^{\frac{Q}{kT}} \alpha x_{Nb}$

JMAEK: $f(t) = 1 - e^{-kt^n}$

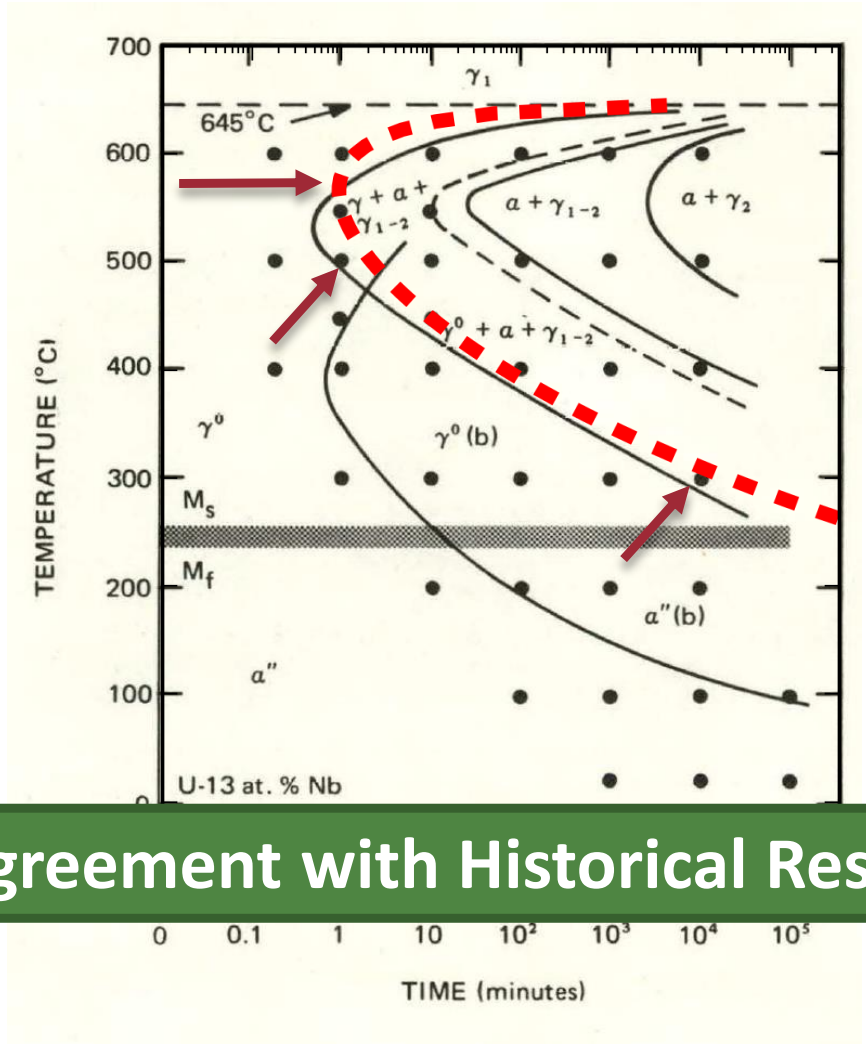


Ref: R. Jackson. RFP-1609.

Result: $f(T, t, x_{Nb}) = 1 - e^{-kt'^n}$

where: $t' = \frac{x_{Nb,ref}}{x_{Nb}} \left(\frac{\Delta T}{\Delta T_{ref}}\right)^3 \exp\left(\frac{Q}{k}\left(\frac{1}{T_{ref}} - \frac{1}{T}\right)\right) t$

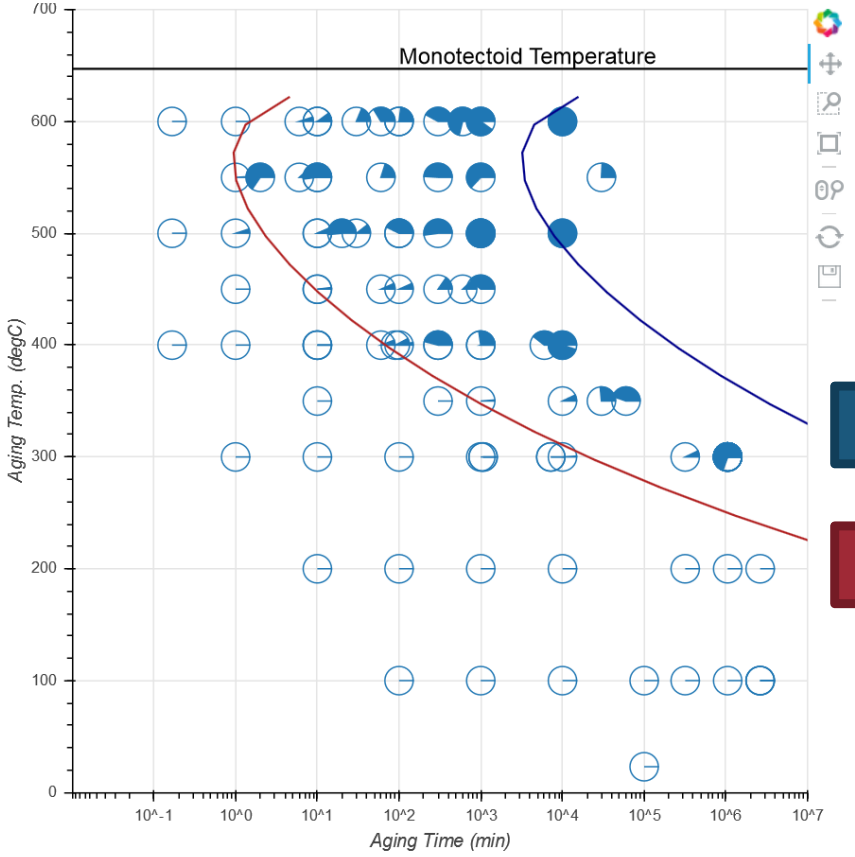
Fitting a TTP Curve: Comparable Results!



Good Agreement with Historical Results

Ref: R. Jackson. RFP-1609.

How Does it Compare to Our Data?



95% transformed

5% transformed

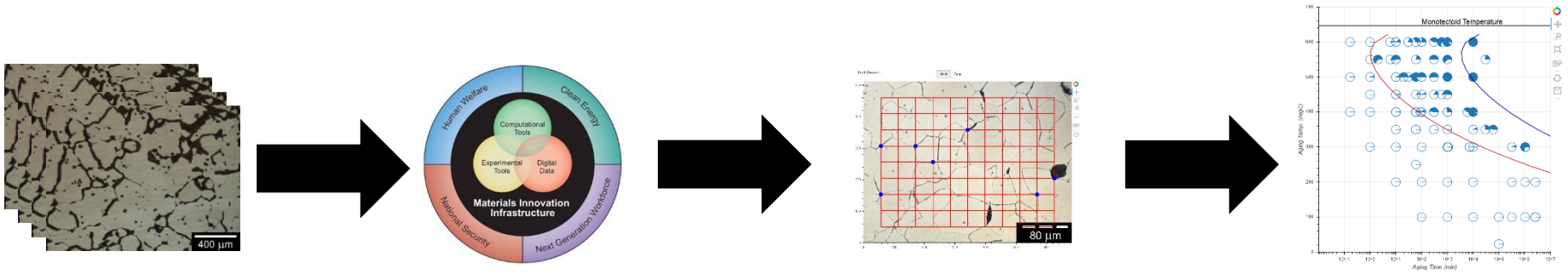
Nominal Nb wt%: 5.5

Nb wt% tolerance:

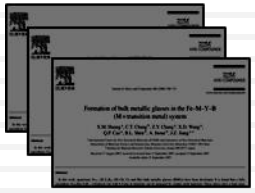
0.5

Image Analysis Workflow

Summary So Far: Simple Procedure from Images to TTT Curve

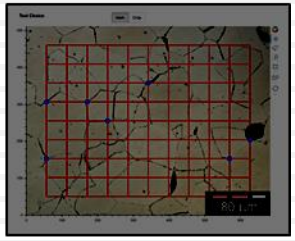


Curating Hardness Records



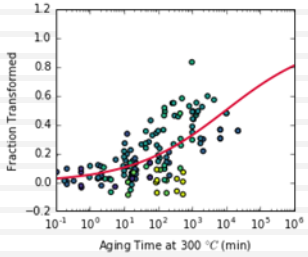
Input: *Papers*

Images



Hardness

Which data is relevant?



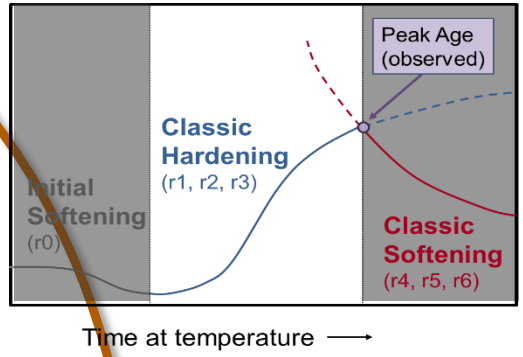
Output: *Aging Model*

Hardness Data: Two Key Decisions

INPUT

Hardness: 100
Scale: Rockwell A
Alloy: U-6wt%Nb
Aging Time: 1 hr
Aging Temp.: 350 C

Before Peak Age?



YES

Focus on two decisions

$f_1(t, T)$

How can we automate, or simplify this process?

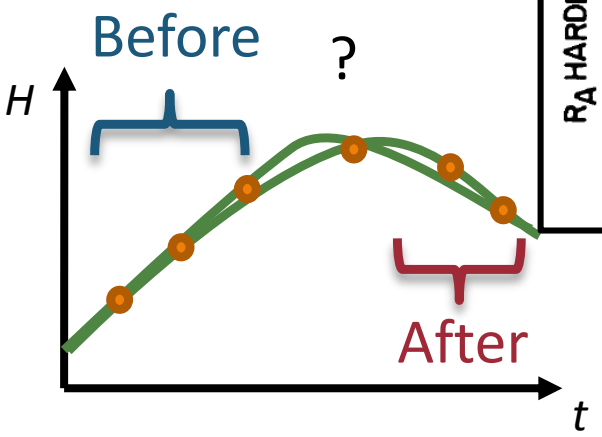
Which Transformation?

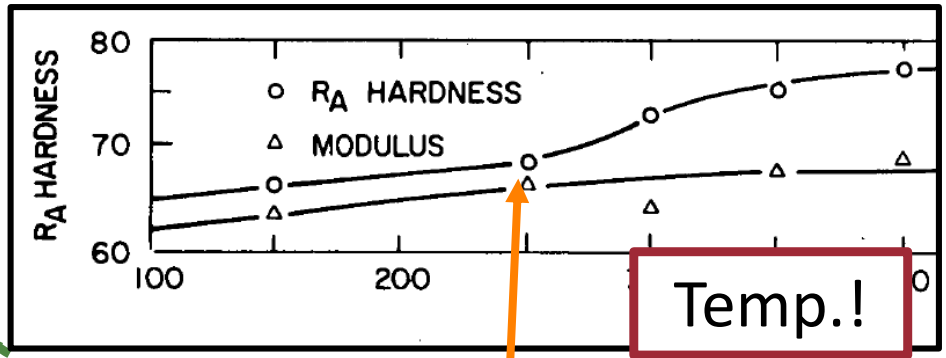
$f_3(t, T)$

Different Transformation
→ Different Physics
→ Different Model

Q1: Before Peak Age? - Automation

Step #1: Automation

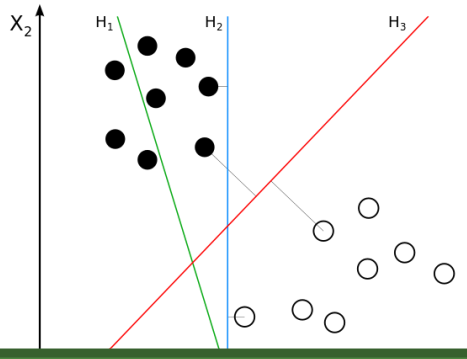


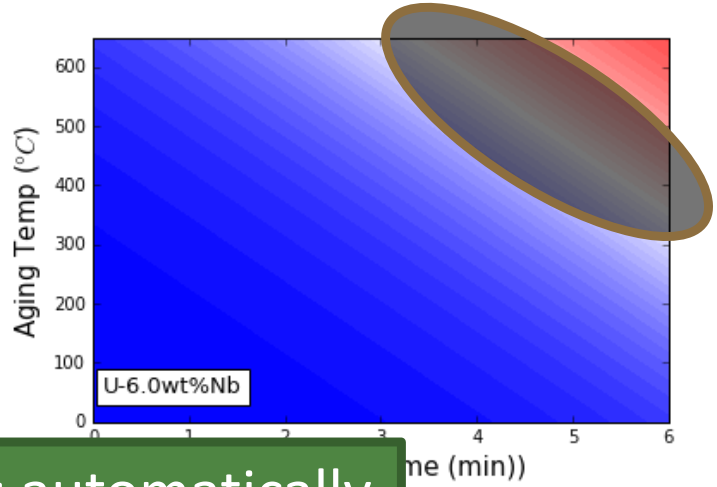


Temp.!

Before or after peak *time*?

Step #2: SVC Model





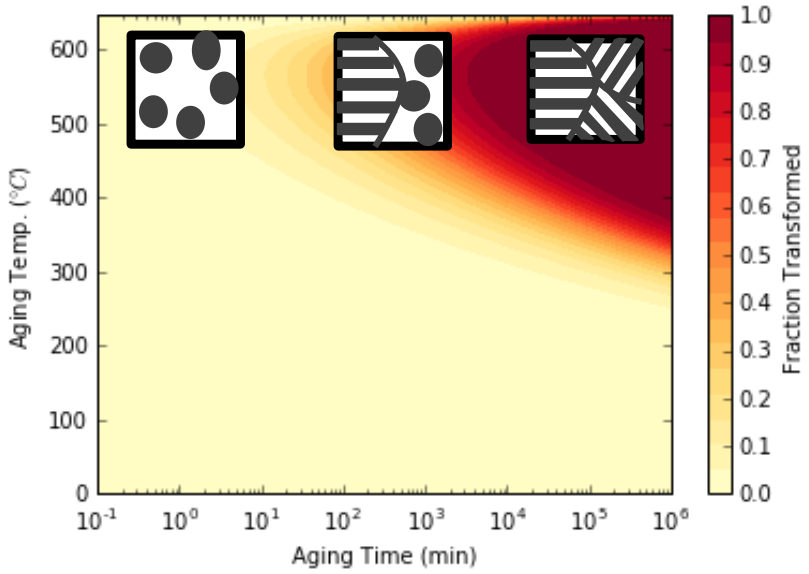
Make the “easy” decisions automatically

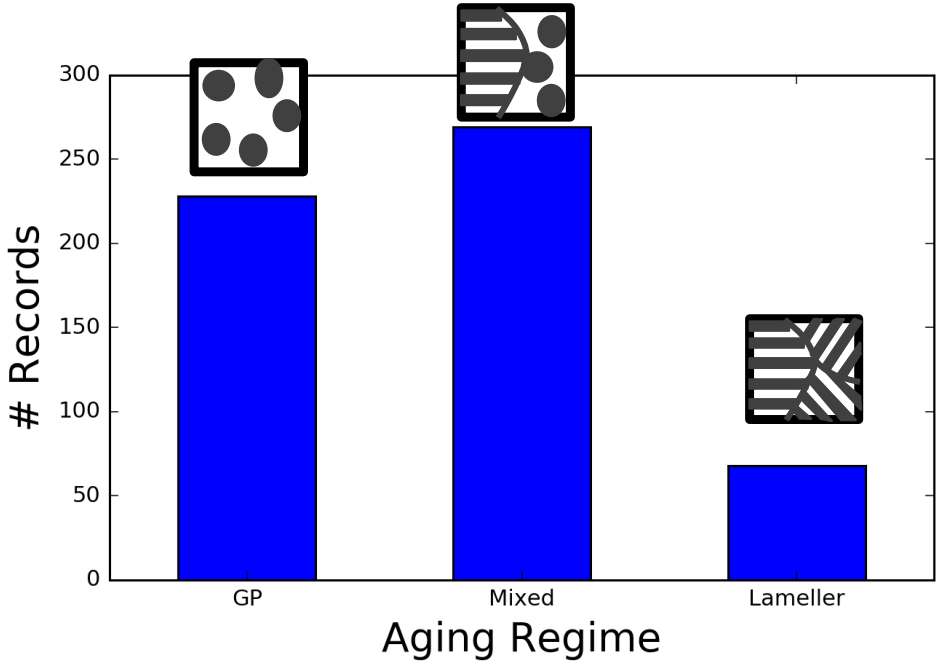
Q1: Before Peak Age? - Interactive Viz

Q2: Which transformation?

Automatically Linked to Image Analysis:

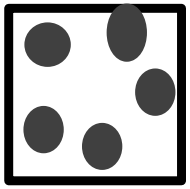
Use TTT Model from Earlier

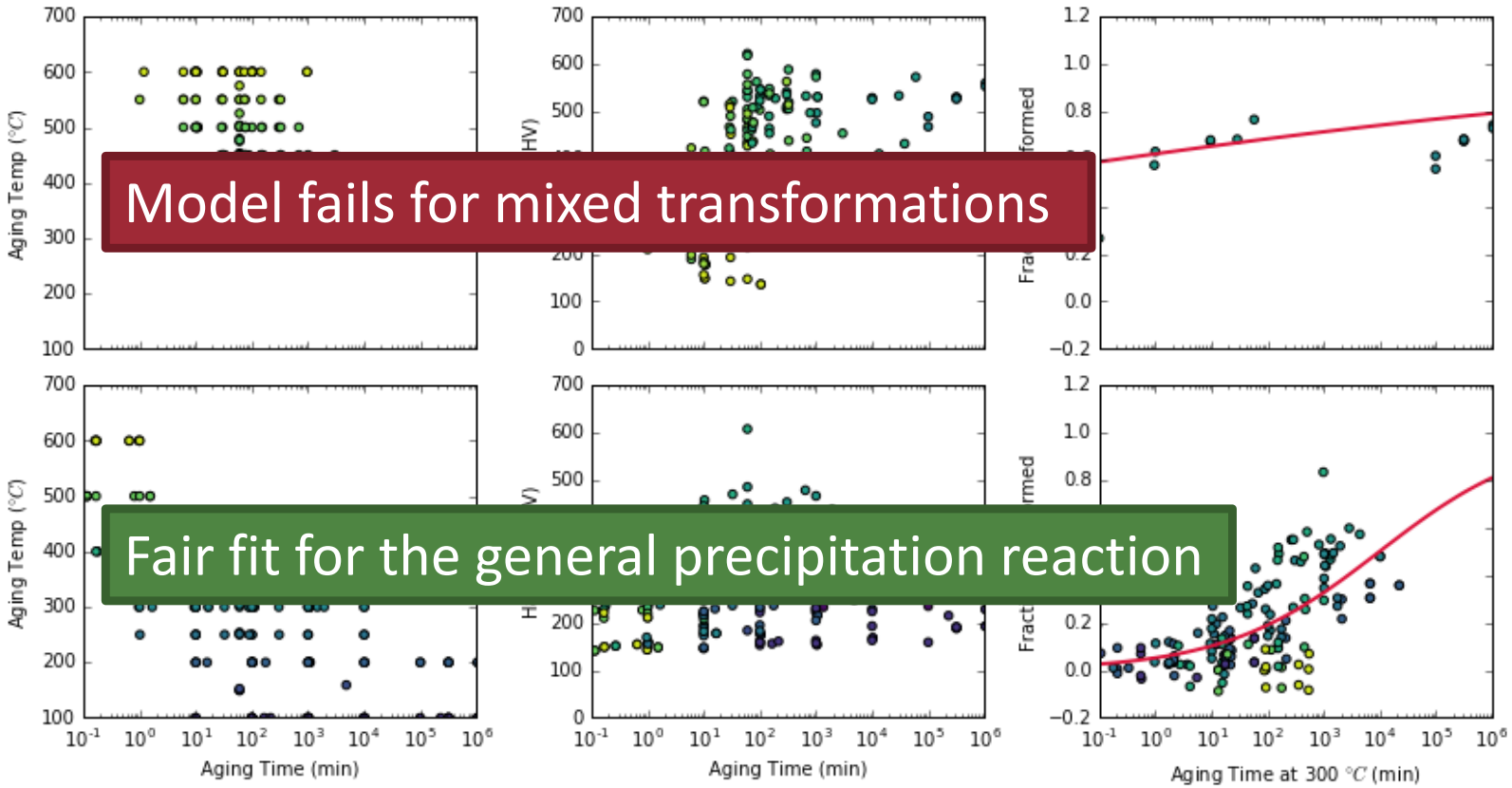




Microstructure-Informed Model

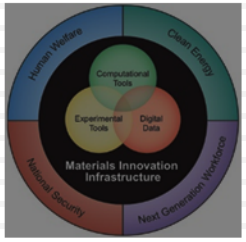
Now we can curate by “before peak” and “transformation type”



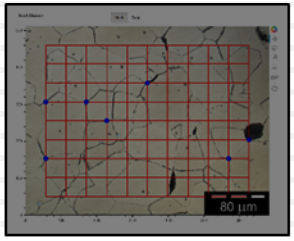


Tracking the Entire Process

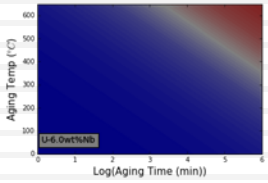
Input: *Papers*

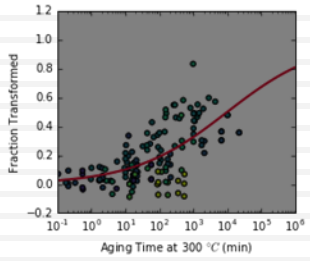


Images



Hardness

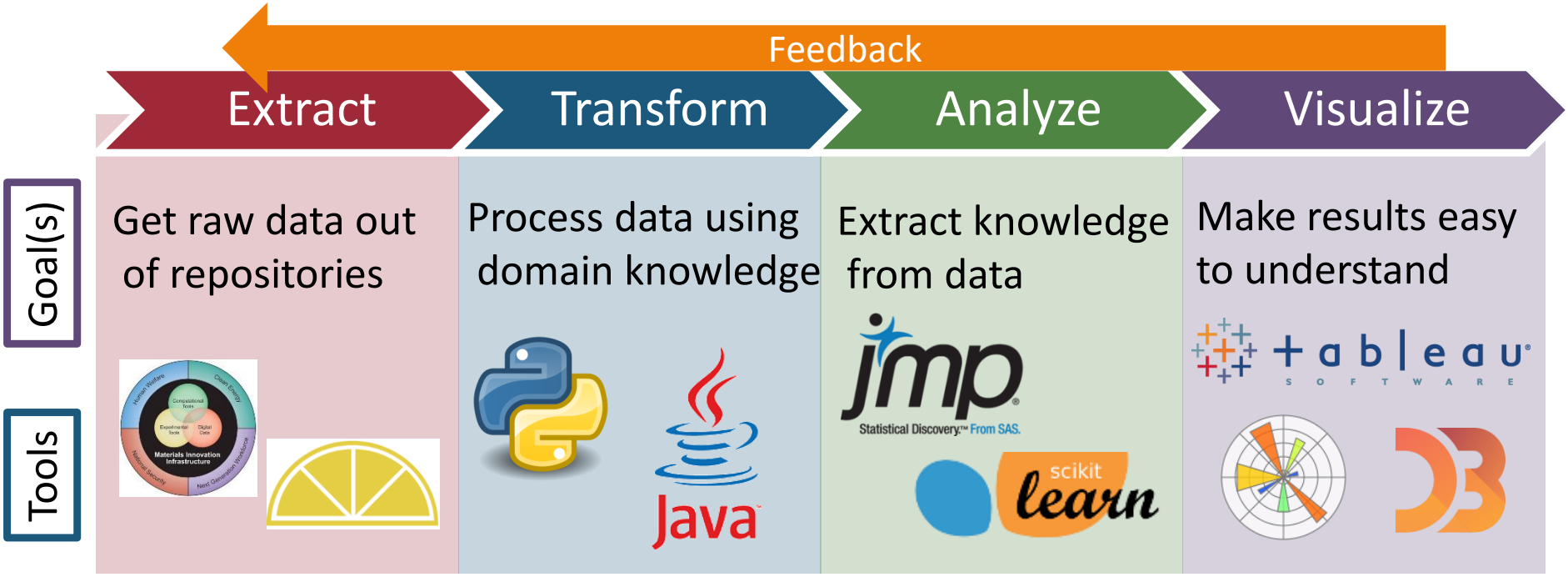




Output: *Aging Model*

General Problem with Property Models

Building models often has many steps, and linking different tools



Problems:

- Updating model with new data
- Keeping track of decisions

Solution: *Pinyon*

Pinyon: Toolkit for Managing Model Creation

Guiding Principle: Store enough information to recreate a model

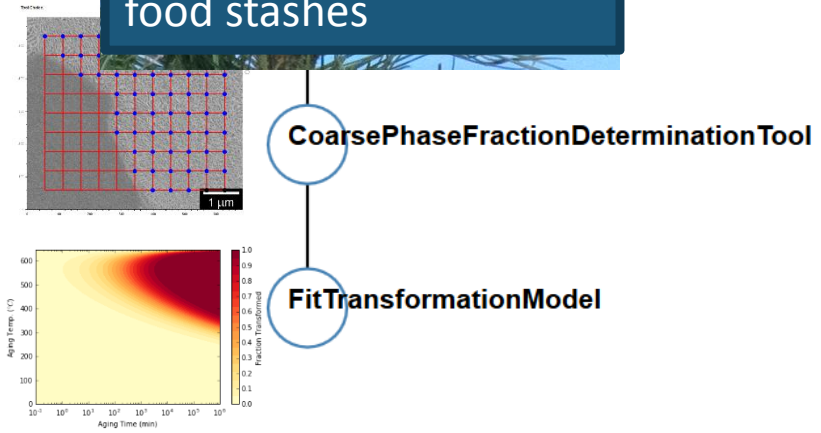
Goal: Collect all steps in model-building process

- Individual calculations
- How they link together
- "Post-It Notes", Documentation

Approach: Database with web interface

- Records all in one place
- Updated automatically
- Can be annotated

Pinyon Jay: NM bird known for remembering food stashes



mongoDB®

Tracking Individual Decisions

Entry Selection Table

Click on submit button to edit entry

Show 10 entries

Search:

	EntryTitle	NominalComposition	AgingTemp	AgingTime	FractionLameller	Edit
260	1967jac2-Figure10a-0	U-8.4Nb	250.0	6.000000e+01	0	Edit
261	1967jac2-Figure10b-1	U-8.4Nb	350.0	6.000000e+01	0	Edit
262	1967jac2-Figure10c-2	U-8.4Nb	500.0	6.000000e+01	0.676767676768	Edit
263	1967jac2-Figure11a-0	U-4.6Nb	250.0	6.000000e+01	0	Edit
264	1967jac2-Figure8a-0	U-4.6Nb	350.0	6.000000e+01	0	Edit
265	1967jac2-Figure8b-1	U-4.6Nb	400.0	6.000000e+01	0.489583333333	Edit
266	1967jac2-Figure9a-2	U-6.4Nb	350.0	6.000000e+01	0	Edit
267	1967jac2-Figure9b-3	U-6.4Nb	400.0	6.000000e+01	0.03125	Edit
268	1967jac2-Figure9c-4	U-6.4Nb	500.0	6.000000e+01	0.833333333333	Edit
341	1967jac2-Figure11b-1	U-4.6Nb	350.0	6.000000e+01	0	Edit

Showing 1 to 10 of 64 entries

Previous 1 2 3 4 5 6 7 Next

Jupyter: Docs & Code in One File!

Fit a TTP Curve Using All Available Data ¶

Fit a TTP Curve Using All Available Data ¶

In [7]:
```
def time_adjust_factor(temp, nb, temp_ref, nb_ref, temp_monotectoid, Q):
    """Compute the multiplicative factor to convert a transformation time measured for one
    alloy at a certain temperature to a time for a reference alloy and aging condition

    :param temp: float, temperature of measurement in degC
    :param nb: float, Nb fraction of alloy being measured
    :param temp_ref: float, reference temperature in degC
    :param nb_ref: float, Nb fraction of reference alloy
    :param temp_monotectoid: float, monotectoid transformation temperature
    :param Q: flaot, "activation energy" fitting parameter
    :return: multiplicative adjustment factor"""

    return nb_ref / nb \
        * np.power((temp_monotectoid - temp) / (temp_monotectoid - temp_ref), 3) \
        * np.exp(Q * (1.0/(temp_ref + 273.0) - 1.0 / (temp + 273.0)))
```

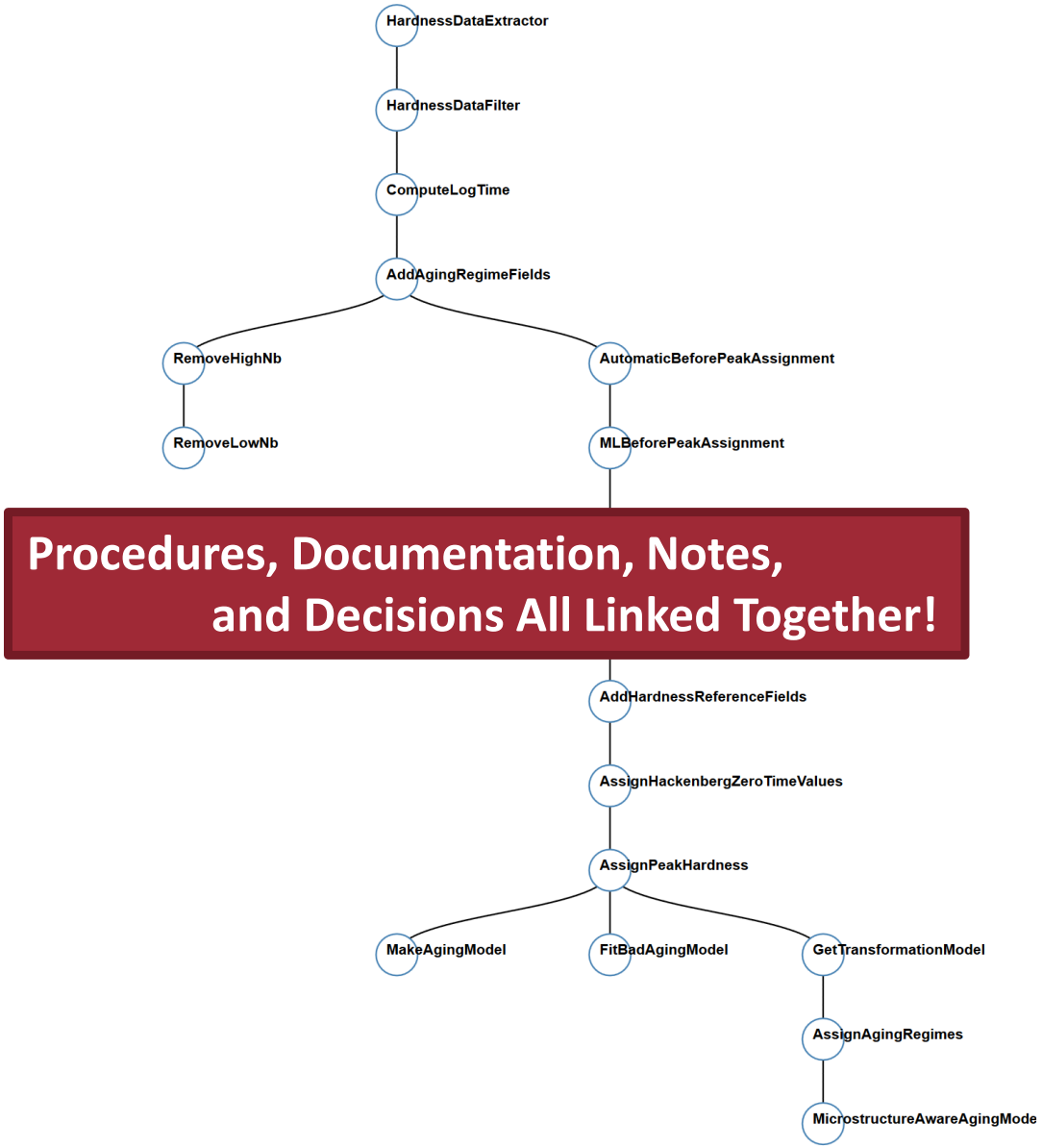
Jupyter: Ensure Whoever Inherits this Code Knows What It Does (and why!)

```
< 1
== 1
```

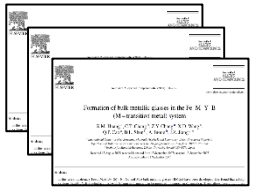
This reduces the equation to three fitting parameters: k, n, and Q.

```
assert time_adjust_factor(100, 6, 100, 6, t_monotectoid, 1000) == 1
```

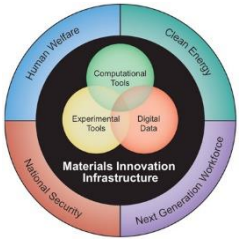
Hardness Analysis Workflow



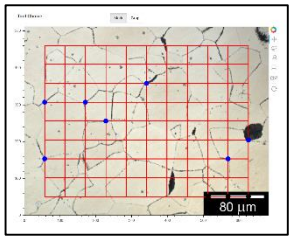
Workflow for Aging Models



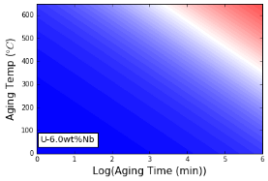
Input:
Papers

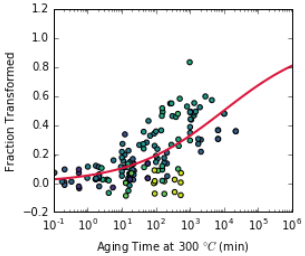


Images



Hardness





Output:
Aging Model

Conclusions

Interconnected Workflow for Aging Models

- ✓ Storing Materials Data: MDCS
- ✓ Hardness Curation: Semiautomated Decision Making
- ✓ Microstructure Data: Web-Based Analysis Tools

Pinyon: Better reproducibility through integration

Future Vision:

- Improving the microstructure-aware aging models
- Adapting these tools to other property models

Acknowledgments

Funding:

People:

- Zachary Trautt, NIST
- Francesca Samsel, UT-Austin and CCS-7