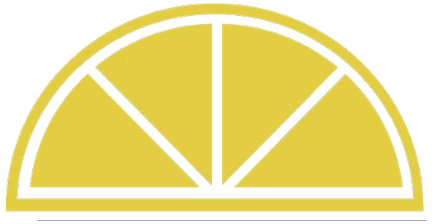


Citrine Informatics

The data analytics platform for the physical world

Experimental design, PIF details, and data workshop

Chris Borg

26 February 2018

Today's goals

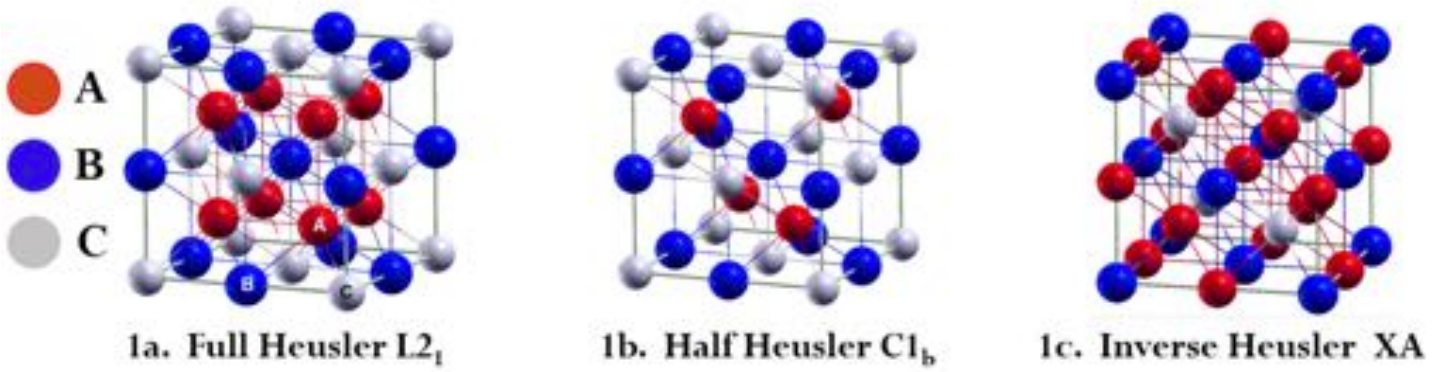
1. Learn how to use the experimental design endpoint.

2. Use the PIF schema to help organize and analyze data.

3. Organize data in a way that is amenable to ML.

Designing new Heusler compounds

THE UNIVERSITY OF ALABAMA | Center for Materials for Information Technology | Heusler Database



1a. Full Heusler $L2_1$ 1b. Half Heusler $C1_b$ 1c. Inverse Heusler XA

Inputs: Chemical composition

Outputs: Formation energy, Heusler type, stability

Electronic structure calculations for 576 full Heuslers, 378 half Heuslers and 405 inverse Heuslers.

Experimental design

Set user-defined ranges for properties to return inputs that map to those ranges.

Define targets:

”Minimize formation energy”

Set constraints:

”Restrict search to Full Heuslers that are stable.”

Analyze results:

“Do the returned results make sense?”

Working with a small design space

$Ba_{0.8\text{-}0.9}Co_{0.2\text{-}0.4}Fe_{0.2\text{-}0.4}Y_{0.05\text{-}0.1}Zr_{0.2\text{-}0.4}O_3$

"default" + input constraints

"This view" (must create sparse dataset)

Return compounds found in ICSD

formula	Property Position of 110 p...
Ba1.0Co0.0Fe0.12118Y0.22...	29.906498
Ba1.0Co0.0Fe0.10368Y0.21...	29.960784999999998
Ba0.97502Co0.0Fe0.85517...	30.8673291904762

Return chemistries in training set

Edit composition input based on material type

inorganics

BaTiO3, CeO2, …

formula	Property Position of 11
Ba0.9241Co0.87788Fe0.0Y…	31.118078000000004
Ba0.92688Co0.88167Fe0.0…	31.108259000000004

alloys

Cr 1 (wt. %), balance Ni

composition	Property Ultimate tensile
Ti.375Al.425Ni37.95Fe39.5C…	69.97628
Ni64.92Ce.03Mo10Cr25C.05	311.478610765889

organics

CC(=O)O, C1CCCCC1, …

Property SMILES	Property Melting point
N1=CC2=C([Se]C=3C=C(NC…	83.66371681415929
N1=CC(=CC=C1C2=CC=C…	81.14159292035399

Today's goals

1. Learn how to use the experimental design endpoint.

2. Use the PIF schema to help organize and analyze data.

3. Organize data in a way that is amenable to ML.

System level PIF objects

uid

names

ids

properties

preparation

subSystems

references

contacts

ChemicalSystem is a specialization of System that includes information about chemical composition

chemicalFormula

composition

CSV-template ingester

http://help.citrination.com/knowledgebase

	A	B
1	FORMULA	PROPERTY: Band gap (eV)
2	MgO2	7.8

Citrine: Template CSV

Converts a specialized CSV/XLSX or XLS file to a physical information file.

MgO_2

Chemical formula: MgO_2

Band gap: 7.8 eV


```
{
    "category": "system.chemical",
    "chemicalFormula": "MgO2",
    "properties": {
        "units": "eV",
        "name": "Band gap",
        "scalars": [
            {
                "value": 7.8
            }
        ]
    }
}
```

PIF properties are materials properties

- Mechanical, physical, electronic, etc.
- Responses to stimuli

Common properties:

Band gap
Electrical resistivity
Tensile strength
Hardness

Common conditions:

Temperature
Time
Applied field
Wavelength

PIF conditions are test conditions

PROPERTY: Ultimate tensile strength (MPa)

400

CONDITION: Temperature (K)

873

CONDITION: Atmosphere

inert

PIF ProcessSteps are sample conditions

- Gives context to how a sample was prepared

Common preparation step names:

Solution treatment
Calcination
Gas carburization
Surface polishing

Common preparation step details:

Heating time
Heating temperature
Number of heating cycles
Atmosphere
Solution type

PIF ProcessSteps are sample conditions

PREPARATION STEP NAME

Solution treatment

PREPARATION STEP DETAIL: Temperature (K)

1323

PREPARATION STEP DETAIL: Time (s)

1500

PIF SubSystems are equivalent systems

System

Name: Fuel cell device
Subsystems: [list of systems]

ChemicalSystem

Name: Cathode material

ChemicalSystem

Name: Anode material

PIF References and contacts

REFERENCE: doi, author, title, year, etc.

10.1016/j.jallcom.2008.07.087

CONTACT: Name

Chris Borg

CONTACT: Email

cborg@citrine.io

PIF identifiers

UID = unique identifier for Citrination record. Will be automatically assigned.

Ids = List of material identifiers (CAS #, internal sample number).

Names = Common name of material or system. Useful if formula is unknown.


```
{
    "name": "CAS",
    "value": "1335-26-8"
}
```

NAME	FORMULA	ID: CAS number	ID: SMILES	ID: inChI Key
Magnesium peroxide	MgO2	1335-26-8	[Mg+2].[O-][O-]	SPAGIJMPHSUYSE-UHFFFAOYSA-N

Today's goals

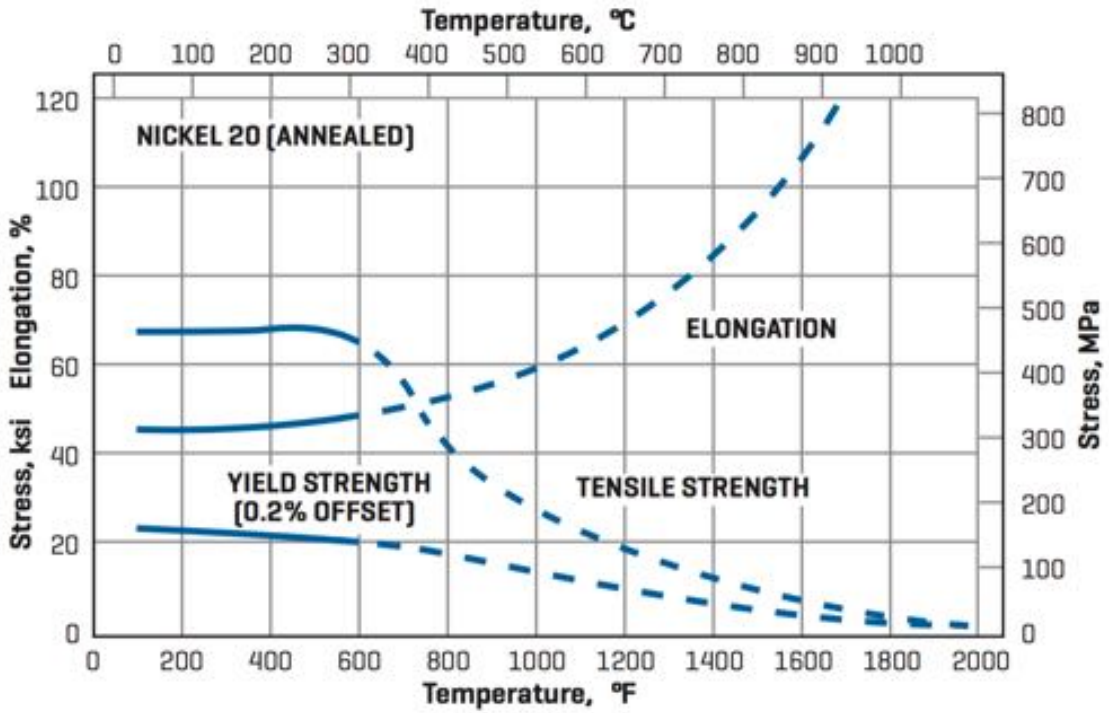
1. Learn how to use the experimental design endpoint.

2. Use the PIF schema to help organize and analyze data.

3. Organize data in a way that is amenable to ML.

Structuring material properties

NICKEL 200



LIMITING CHEMICAL COMPOSITION, %	Ni[a].....99.0 min.	Mn.....0.35 max.	S.....0.01 max.
	Cu.....0.25 max.	C.....0.15 max.	
	Fe.....0.40 max.	Si.....0.35 max.	
	[a]Plus Co.		

PHYSICAL CONSTANTS AND THERMAL PROPERTIES	Density, lb/in^3	0.321
	g/cm^3	8.89
	Melting Range, °F	2615 - 2635
	°C	1435 - 1446
	Specific Heat, Btu/lb•°F	0.109
	J/kg•°C	456
	Curie Temperature, °F	680
	°C	360
	Permeability	Ferromagnetic
	Coefficient of Expansion, 70 - 200°F, 10^{-6} in/in•°F	7.4
	21 - 93°C, µm/m•°C	13.3
	Thermal Conductivity, Btu • in/ft^2•h•°F	487
	W/m•°C	70.2
	Electrical Resistivity, ohm•circ mil/ft	58
	µΩ•m	0.096

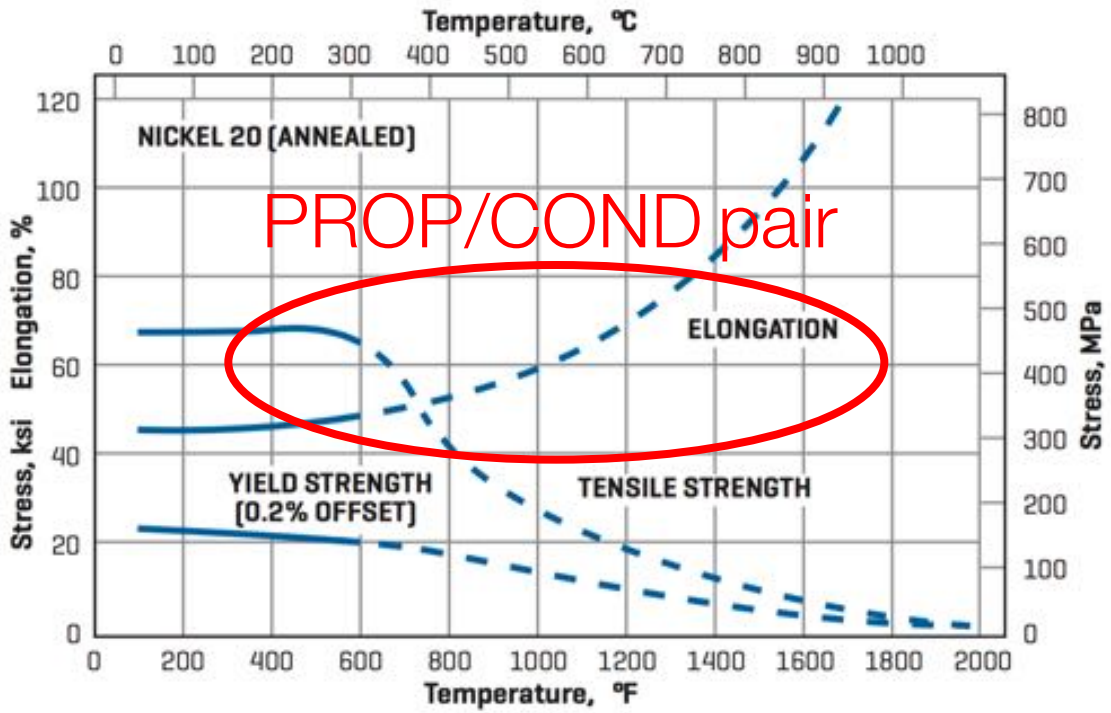
TYPICAL MECHANICAL PROPERTIES	**(Annealed)**	
	Tensile Strength, ksi	67
	MPa	462
	Yield Strength (0.2% Offset), ksi	21.5
	MPa	148
	Elongation, %	47

Structuring material properties

NAME

COMPOSITION

NICKEL 200



LIMITING CHEMICAL COMPOSITION, %	Ni[a].....99.0 min.	Mn.....0.35 max.	S.....0.01 max.
	Cu.....0.25 max.	C.....0.15 max.	
	Fe.....0.40 max.	Si.....0.35 max.	
	[a]Plus Co.		

PHYSICAL CONSTANTS AND THERMAL PROPERTIES	Density, lb/in³	0.321
	g/cm³	8.89
	Melting Range, °F	2615 - 2635
	°C	1435 - 1446
	Specific Heat, Btu/lb•°F	0.109
	J/kg•°C	456
	Curie Temperature, °F	680
	°C	360
	Permeability	Ferromagnetic
	Coefficient of Expansion, 70 - 200°F, 10⁻⁶ in/in•°F	7.4
	21 - 93°C, µm/m•°C	13.3
	Thermal Conductivity, Btu • in/ft²•h•°F	487
	W/m•°C	70.2
	Electrical Resistivity, ohm•circ mil/ft	58
	µΩ•m	0.096

PROPERTIES

TYPICAL MECHANICAL PROPERTIES	**(Annealed)**	
	Tensile Strength, ksi	67
	MPa	462
	Yield Strength (0.2% Offset), ksi	21.5
	MPa	148
	Elongation, %	47

Composition and identifiers

NICKEL 200

LIMITING CHEMICAL COMPOSITION, %	Ni[a].....99.0 min.	Mn.....0.35 max.	S.....0.01 max.
	Cu.....0.25 max.	C.....0.15 max.	
	Fe.....0.40 max.	Si.....0.35 max.	
	[a]Plus Co.		

```
{
  "element": "Ni",
  "idealWeightPercent": {
    "value": "99"
  }
},
```

FORMULA
Ni0.99Cu0.25Fe0.40Mn0.35C0.15Si0.35S0.01

IDEAL COMPOSITION: Ni (wt. %)	IDEAL COMPOSITION: Cu (wt. %)	IDEAL COMPOSITION: Fe (wt. %)
99	0.25	0.4

Properties

Density, lb/in^3	0.321
g/cm^3	8.89
Melting Range, °F	2615 - 2635
°C	1435 - 1446
Specific Heat, Btu/lb•°F	0.109
J/kg•°C	456
Curie Temperature, °F	680
°C	360
Permeability	Ferromagnetic
Coefficient of Expansion, 70 - 200°F, 10^{-6} in/in•°F	7.4
21 - 93°C, µm/m•°C	13.3
Thermal Conductivity, Btu • in/ft^2•h•°F	487
W/m•°C	70.2
Electrical Resistivity, ohm•circ mil/ft	58
µΩ•m	0.096

PIF and csv-template support ranges and LaTeX formatted units.

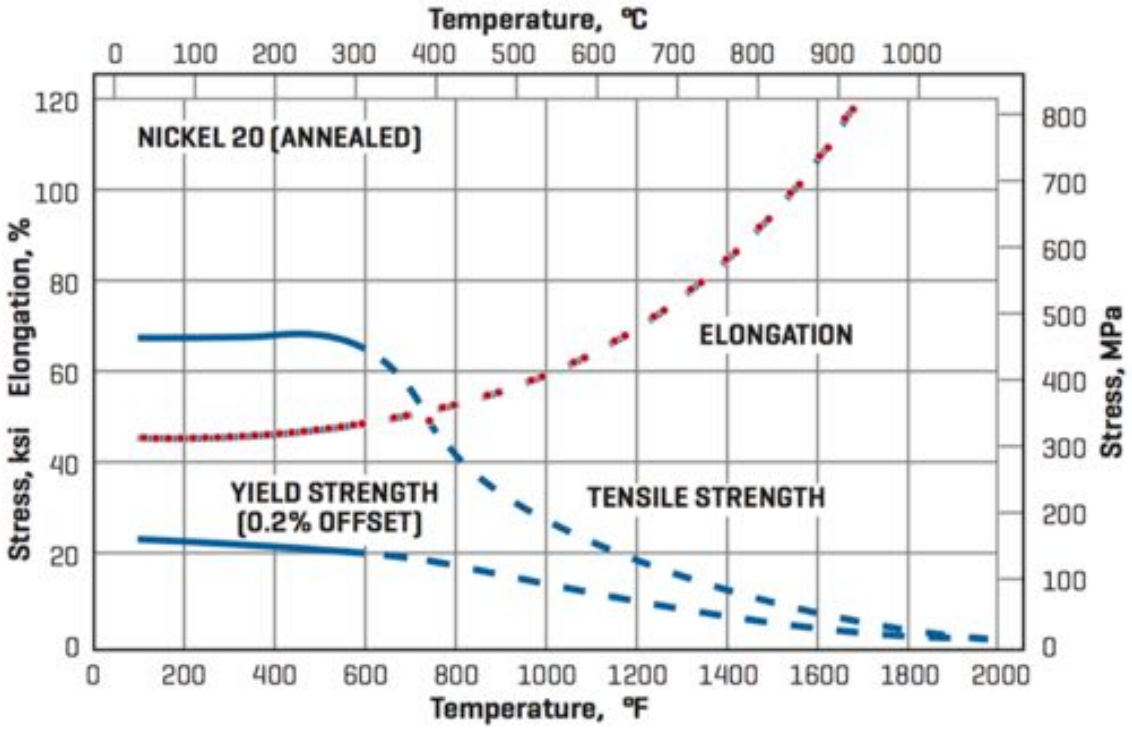
PROPERTY: Density (g/cm^3)
8.89

PROPERTY: Specific heat (J/kg$^\circ$C)
456

PROPERTY: Melting range ($^\circ$C)
1435-1446

Extracting prop/cond from plots

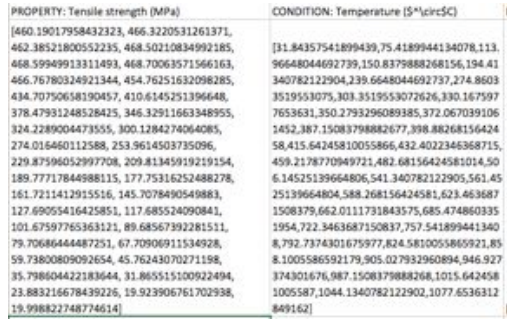
PROPERTY



CONDITION

x, y

43.22, 46.07
58.48, 45.95
73.73, 45.93
88.99, 45.98
104.25, 46.00
119.50, 46.08
134.76, 46.13
150.01, 46.26
165.26, 46.39
180.52, 46.55
195.77, 46.73
211.02, 46.95
226.27, 47.19
241.52, 47.45
256.77, 47.75
272.02, 48.06
287.26, 48.41
302.51, 48.81
313.47, 49.10
353.00, 50.38



PROPERTY: Tensile strength (MPa)	CONDITION: Temperature (S^\circ$C)
[460.19017958432323, 466.3220531261371, 462.38521800552235, 468.50210834992185, 468.59949913311493, 468.70063571566163, 466.76780324921344, 454.76251632098285, 434.70750658190457, 410.6145251396648, 378.47931248528425, 346.32911663348955, 324.22890044473555, 300.1284274064085, 274.0164601125 88, 253.96145037350 96, 229.87596052997708, 209.81345919219154, 189.77717844988115, 177.75316252488278, 161.72114129155 16, 145.70784905498 83, 127.69055416425851, 117.68552409084 1, 101.67597765363121, 89.68567392281511, 79.70686444487251, 67.70906911534928, 59.73800809092654, 45.76243070271198, 35.79860442218364 4, 31.86551510092249 4, 23.88321667843922 6, 19.92390676170293 8, 19.99882274877461 4]	[31.84357541899439,75.4189944134078,113. 96648044692739,150.8379888268156,194.41 3407821229 04,239.6648044692737,274.8603 3519553075,303.3519553072626,330.167597 7653631,350.2793296089385,372.067039106 1452,387.1508379888267 7,398.8826815642 4 58,415.6424581005586 6,432.4022346368715, 459.2178770949721,482.6815642458101 4,50 6.1452513966480 6,541.3407821229 05,561.45 2513966480 4,588.2681564245 81,623.463687 1508379,662.0111731843575,685.4748603 35 1954,722.3463687150837,757.5418994413 40 8,792.7374301675977,824.5810055865921,85 8.1005586592179,905.0279329608 94,946.927 3743016 76,987.1508379888268,1015.642458 1005587,1044.1340782122902,1077.6536312 849162]

Python Citrination client

https://github.com/CitrineInformatics/python-citrination-client

https://citrineinformatics.github.io/api-documentation/#

- Programmatically push and pull files from Citrination
- Query Citrination for properties relevant to your research
- Prediction queries on trained models

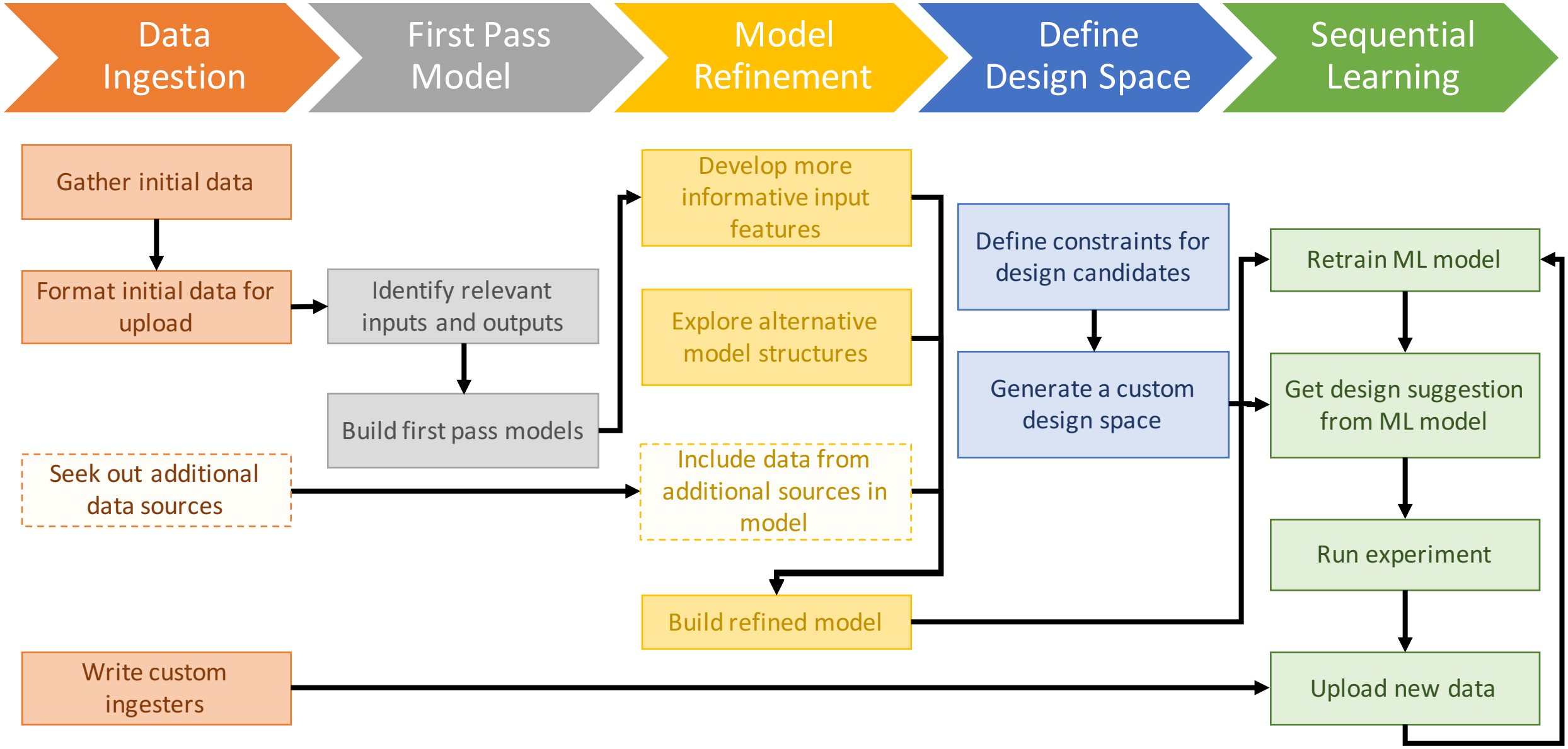
Data workshop

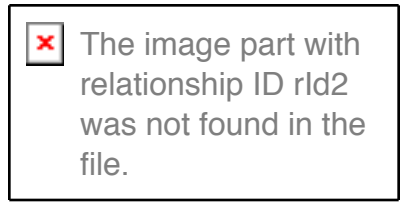
1. Identify any data you wish to digitize. If you do not have any, there is an Al alloy dataset shared with the MIDDMI team.

2. Using PIF fields, decide how to organize the relevant info in your dataset.

3. Upload to Citrination and present the organizational choices you made.

4. Consider where you are in the Al workflow stages.


Data Ingestion
First Pass Model
Model Refinement
Define Design Space
Sequential Learning
Gather initial data
Format initial data for upload
Identify relevant inputs and outputs
Build first pass models
Seek out additional data sources
Write custom ingesters
Develop more informative input features
Explore alternative model structures
Include data from additional sources in model
Build refined model
Define constraints for design candidates
Generate a custom design space
Retrain ML model
Get design suggestion from ML model
Run experiment
Upload new data



Citrine Informatics

The data analytics platform for the physical world

Thank you

cborg@citrine.io