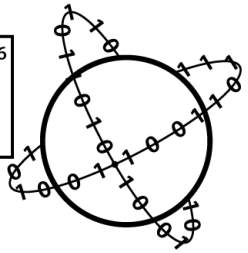
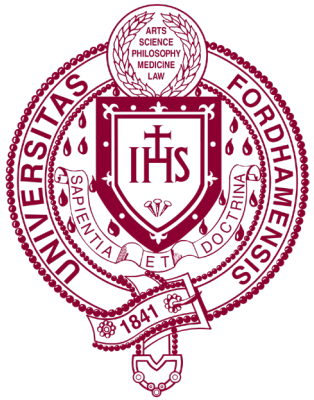
99 [254]	20 40.078	57 138.905	52 127.6
Es	Ca	La	Te

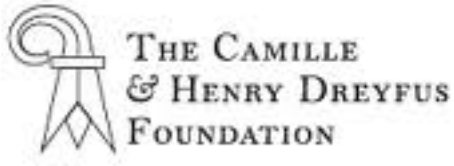


Creating Complex Scientific Workflows that Reach Into the Real World

Joshua Schrier, Fordham University



Thanks to our sponsors…



“We” is a deliberate pronoun choice

Alexander Norquist (Haverford College)
—solid state chemist & crystallographer.

Emory Chan (Molecular Foundry, LBL)
—inorganic lab automation

Technicians:

- Phillip Nega->A123
- Maggie Zelle → U. Delaware
- Rodolfo Keesey ‘20→Washington U. St. Louis
- Becca McAlpern ’23 (exchange student from U. Edinburgh)

Collaborators

- Löic Roch (Atinary)
- Alex Cristafaro & Ben Gordon & Ben Garcia (MIT/Broad)
- Hamed Eramian (Netrias)
- Scott Novotney (TwoSix)
- Nick Leiby (TwoSix)
- Francis Motta (Florida Atlantic)
- Gemma Moran (Columbia)
- Joe Allen (TACC)

An army of undergraduate students…
Even a few high school students…

Funding

DARPA SD2 (HR001118C0036)

Other senior collaborators

- Sorelle Friedler (Haverford)
- Tonio Buonassisi (MIT)

Postdocs

- Ian Pendleton→Relay Pharma.
- Venkateswaran Shekar →Brookhaven NL
- Mansoor Ani Najeeb Nellikkal→Univ. Southh
- Zhi Li→Adv. Semiconductors
- Mina Kim
- Shijing Sun→Toyota Research Institute

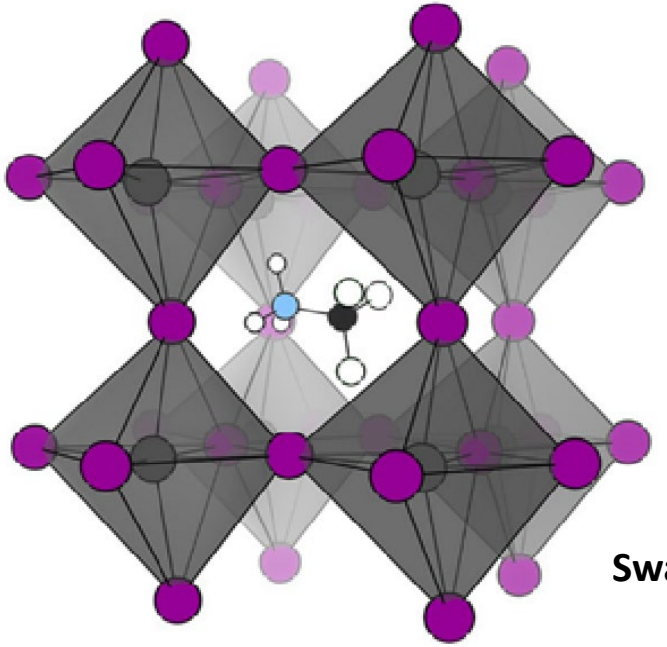
Software Engineers

- Gary Cattabriga → Healthcare startup
- Joe Panizzo

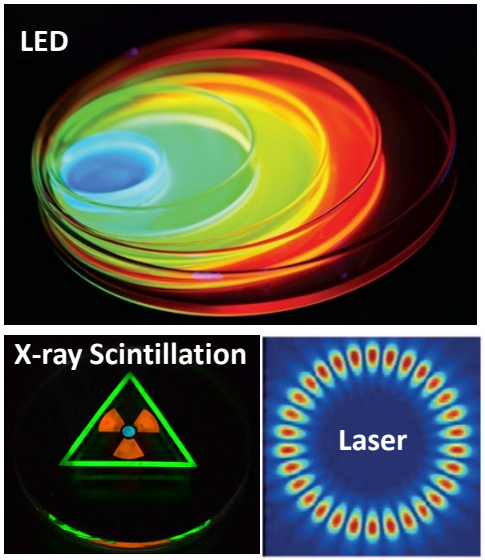
Master’s & Undergrad students

- Michael Tynes ‘20 → LANL/UChicago
- Aaron Dharna ‘20→NJIT/UBC
- Nicole Smina ‘20 →NYU
- Yuanqing Tang
- Jan Estrada ‘20 →Stanford
- Mary Caucci ‘20→Penn State
- Allysa Sherman→Boeing
- Liana Alves→UCSD
- Peter Cruz Parilla→UW Madison
- Victor Ghosh (Bronx HS Science)→Harvard
- +…

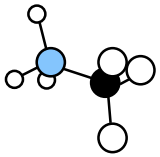
Perovskites: tunable materials for high-performance, low-cost optoelectronics



Applications



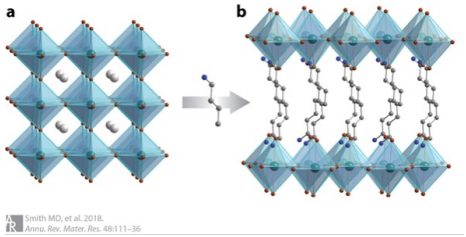
Swapping organic "amines" results in diverse structures/properties

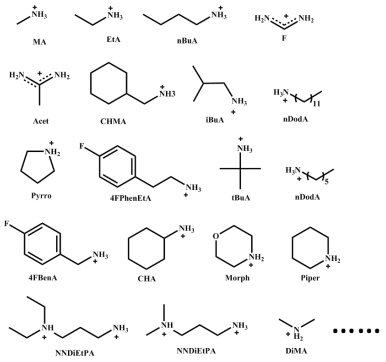


Amine

Metal

Halide



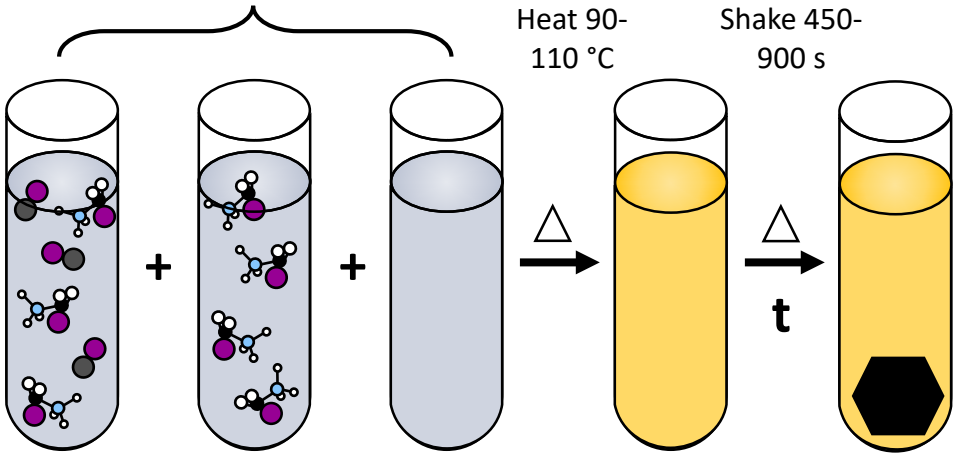


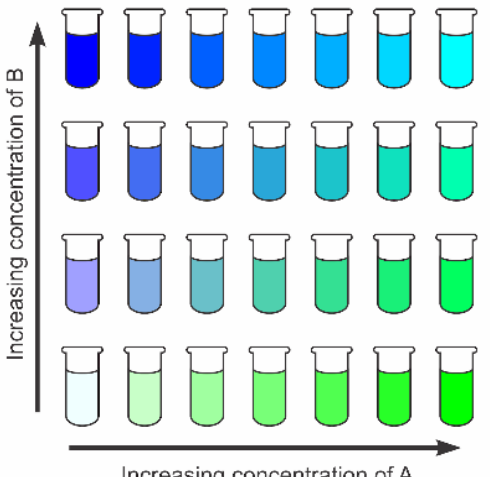
The Challenge: How do you make a new material?

Inorganic 0.1 - 1.500 mmol

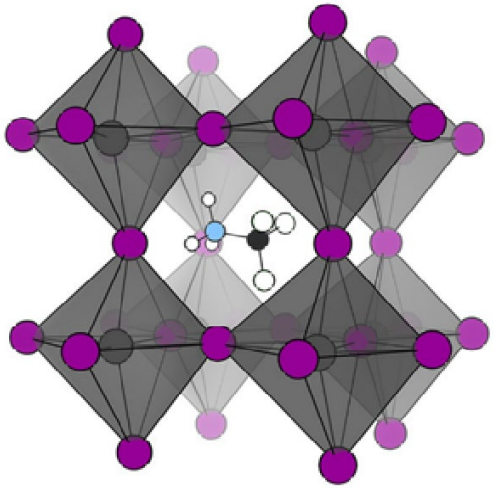
Organic 0.1 - 3.000 mmol

Acid 0.1 - 8.000 mmol

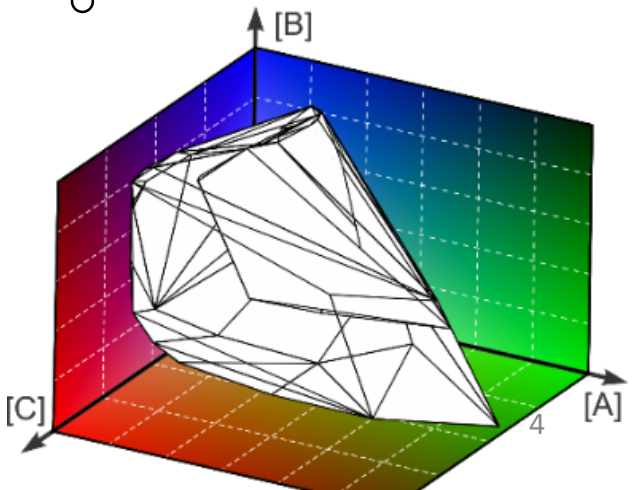




$>10^7$ conditions (too many to test)
Only some are productive




Amine Metal Halide



Can’t we just compute this?

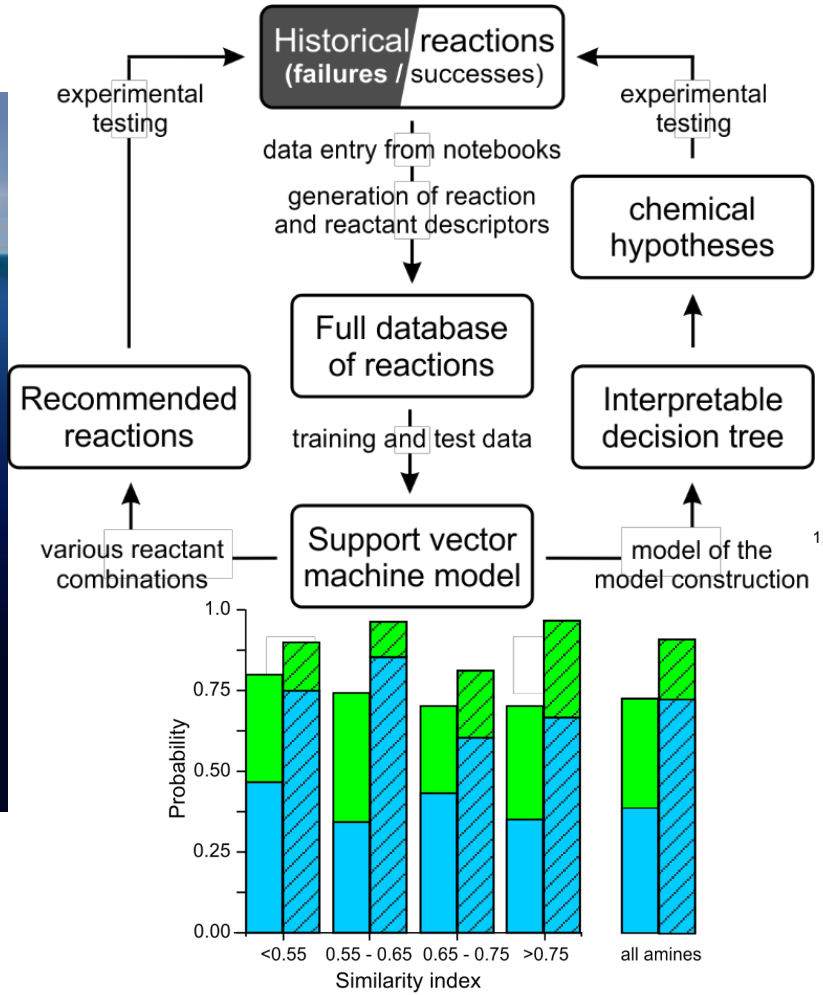
(from underlying physical theory?)

(from ML?)

Good ML models require comprehensive experimental data of success and *failure*.

Journals do not publish "failed" or "marginally successful" experiments ("dark data")….

…but incorporating this data improves ML model predictions…

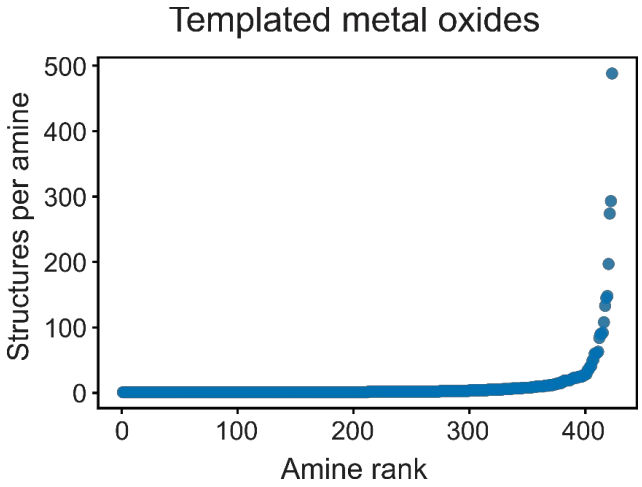


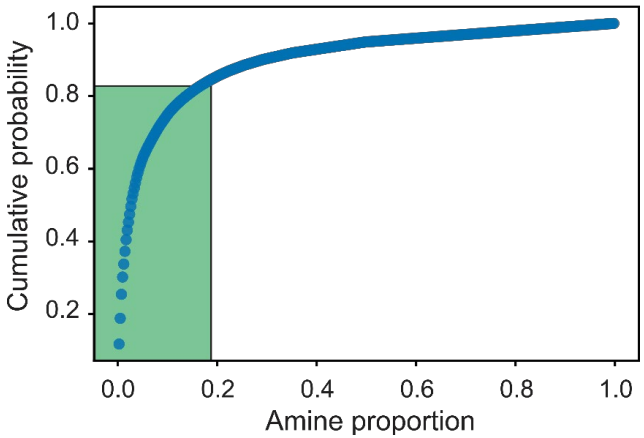
…and new chemical hypotheses can be extracted from this "dark data" in the data/model…

Nature **533**, 73-76 (2016)
doi:10.1038/nature17439

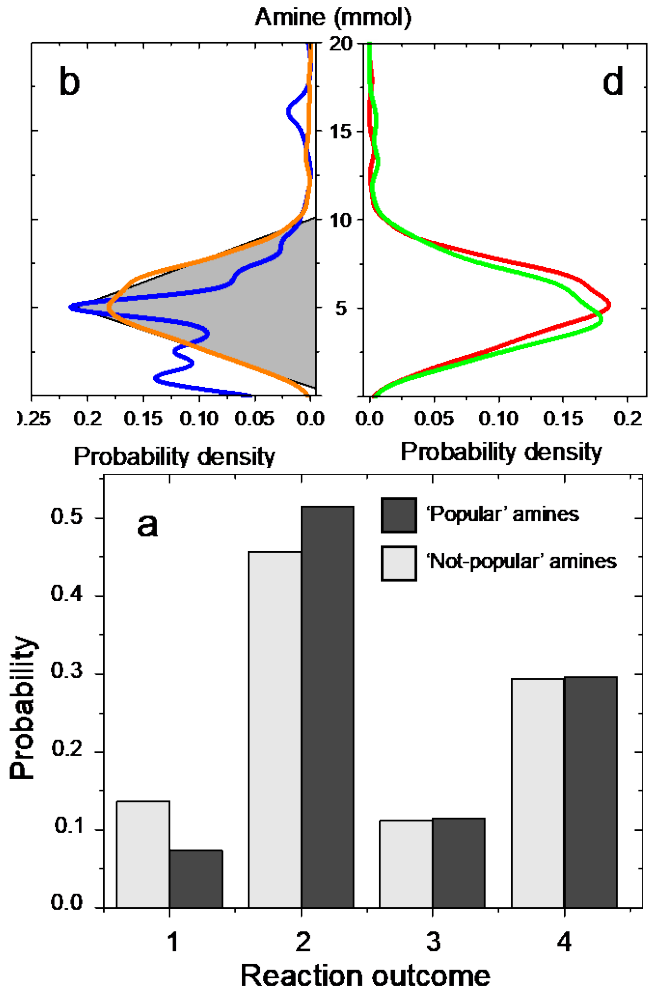
Humans get stuck in a rut…and pollute the data.

Some reagents and reaction conditions are over-represented in both published datasets and unpublished lab notebooks…

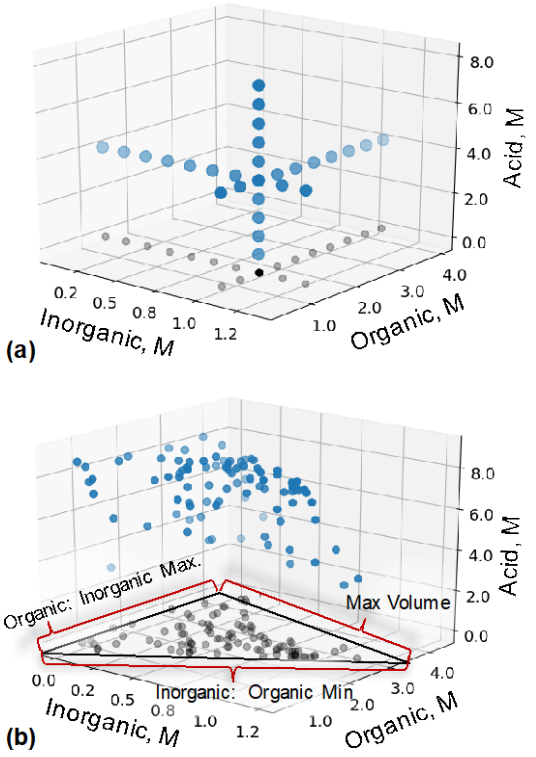




…popularity is uncorrelated to reaction success rates, cost, and technical considerations.

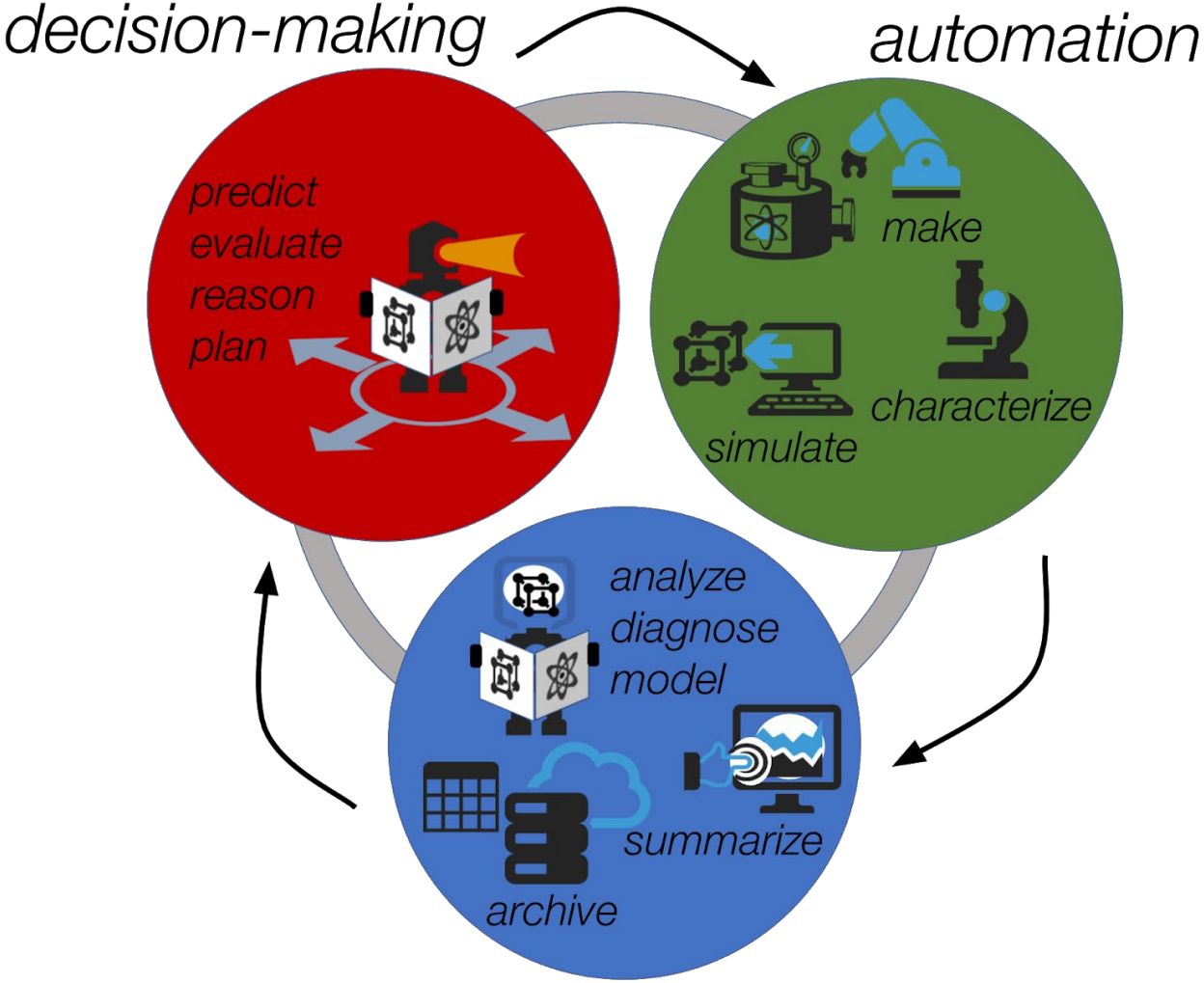


Datasets without anthropogenic bias can be ~20% smaller, yet give better ML models by all metrics.



Nature **573**, 251–255 (2019)
doi:10.1038/s41586-019-1540-5

A Dream: Autonomous research labs



Perspective & Review article: *Matter* (2021) 10.1016/j.matt.2021.06.036

Opinion: Automated experimentation will improve reproducibility and replicability.

Complete disclosure of laboratory process.

Cross-lab replicability.

Complete record of success and failure ("dark reactions") for ML models

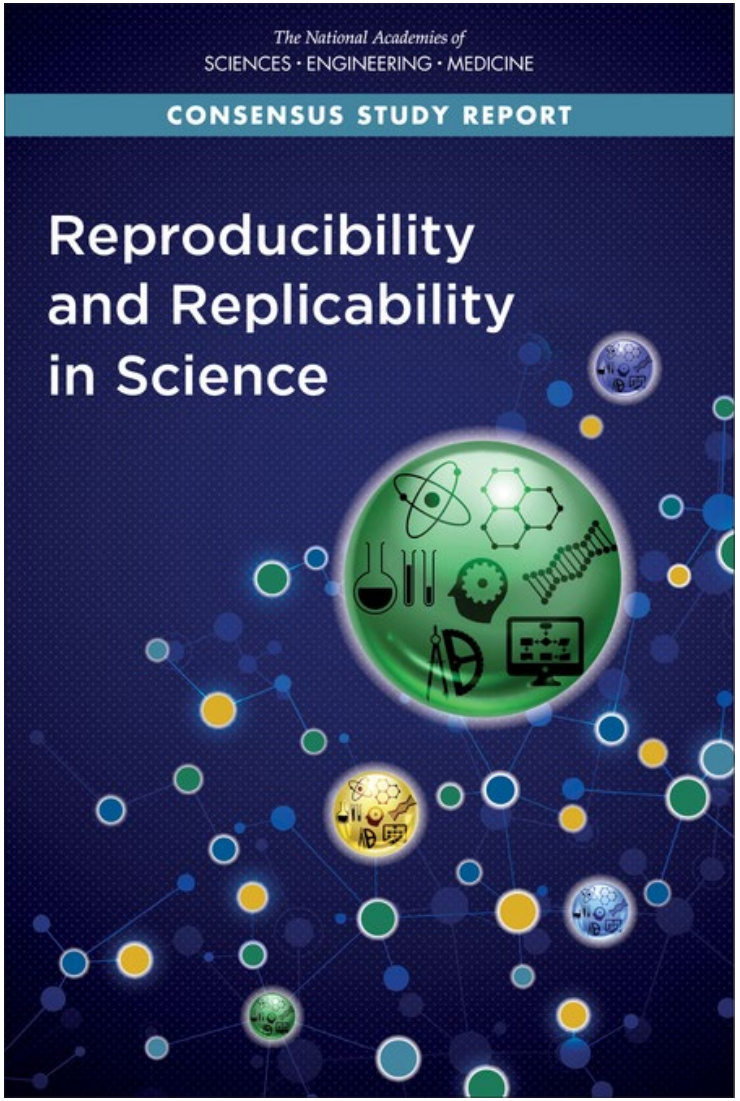
Nature **533**, 73-76 (2016) doi:10.1038/nature17439

Reduce human sampling biases that pollute datasets.

Nature **573**, 251–255 (2019) doi:10.1038/s41586-019-1540-5

Data capture of "unimportant" details enables "automated serendipity"

Appl. Phys. Lett. **119**, (2021) 041903 doi:10.1063/5.0059767



http://sites.nationalacademies.org/sites/reproducibility-in-science/

However: Automation doesn't solve all problems…

Image source: http://www.hallofjusticecomics.com/magnus-robot-fighter-37/

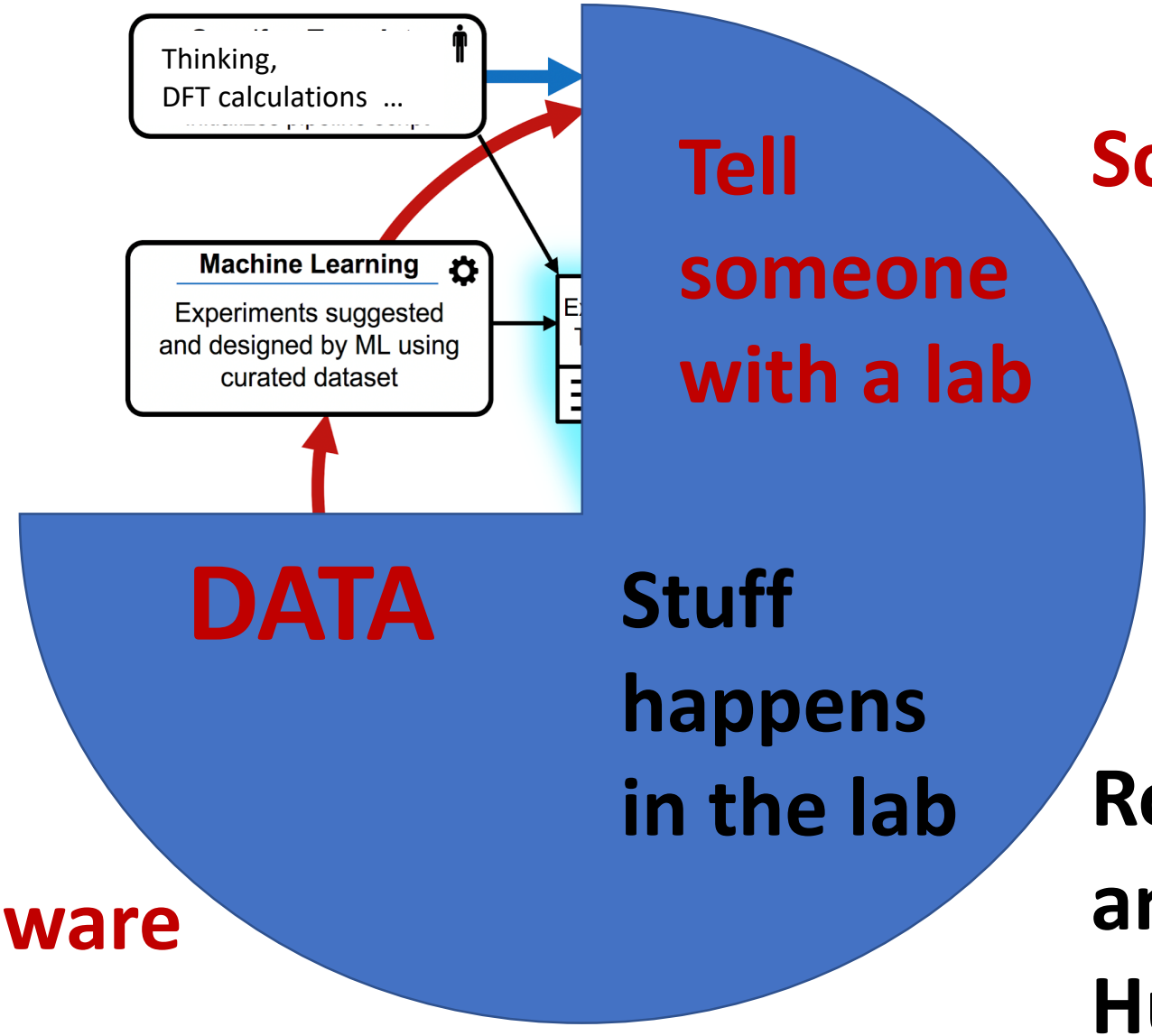
There remains a need to **capture human operator actions** (example: preparing stock solutions, etc.)

Legacy non-automated equipment with unique capabilities—require data import/export.

Over the next decade, expect **"islands of automation"** (rather than completely autonomous systems) will probably the norm.

Opinion: It's easy to get distracted by robots/gadgets—they're cool, but not the whole story.

The need for experimental workflow management



Putting together the pieces of the puzzle…

Automate the Experiments: Robot-Accelerated Perovskite Investigation and Discovery **(RAPID)**

Zhi Li & M. Ani Najeeb, et al. *Chem Mater.* 2020 doi:10.1021/acs.chemmater.0c01153
Zhi Li & M. Ani Najeeb, et al. *Chem Mater.* 2022 doi:10.1021/acs.chemmater.1c03564

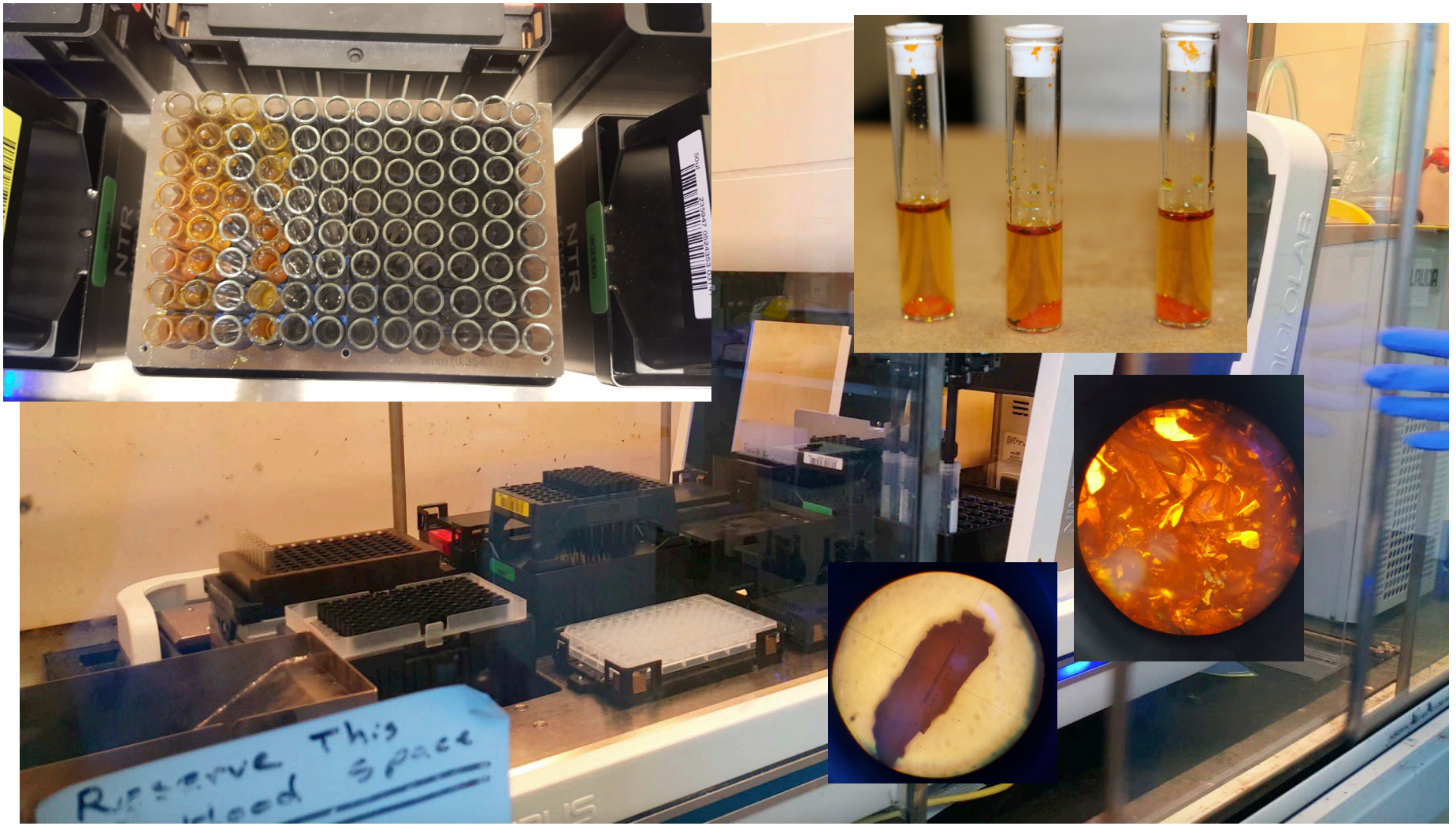
Collect complete data records: Experiment Specification Capture and Lab Automation Technology Environment **(ESCALATE)**

I.Pendleton & G. Cattabriga, et al. *MRS Communications* (2019) - doi:10.1557/mrc.2019.72
http://github.com/darkreactions/ESCALATE

Develop and test machine learning for the Lab!

I. Pendleton et al. *J. Phys. Chem C*. (2020) doi:10.1021/acs.jpcc.0c01726
Y. Tang et al. *J. Chem. Inf. Model.* (2021) doi:10.1021/acs.jcim.0c01307
P. W. Nega et al., *J. Appl. Phys*. (2021) doi:10.1063/5.0059767
V. Shekar et al. *J. Chem. Phys*. (2022) doi:10.1063/5.0076636
V. Shekar et al., ChemRXiv (2022) doi:10.26434/chemrxiv-2022-l1wpf

Automating the Experiments with "Robot-ready" reactions



Liana Alves '18
Alyssa Sherman '18
Peter Cruz Parilla '20

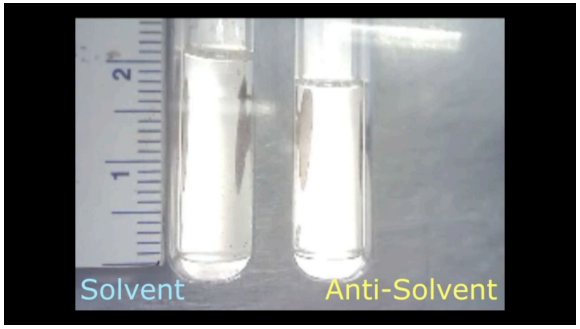
Emily Brown '19
Mansoor Ali Najeeb Nellikkal

Zhi Li (LBL/Molecular Foundry)
Emory Chan (LBL/Molecular Foundry)

Robot-Accelerated Perovskite Investigation & Discovery (RAPID)

Inverse Temperature Crystallization (ITC/WF1/"RAPID 1")

use a retrograde solubility effect to grow large high-quality crystals without strong acids
Saidaminov *et al*. *Nat. Commun.* (2015).
Zhi Li & M. Ani Najeeb et al. Chem. Mater. (2020) doi:10.1021/acs.chemmater.0c01153



Antisolvent Vapor Diffusion (WF3/"RAPID 2")

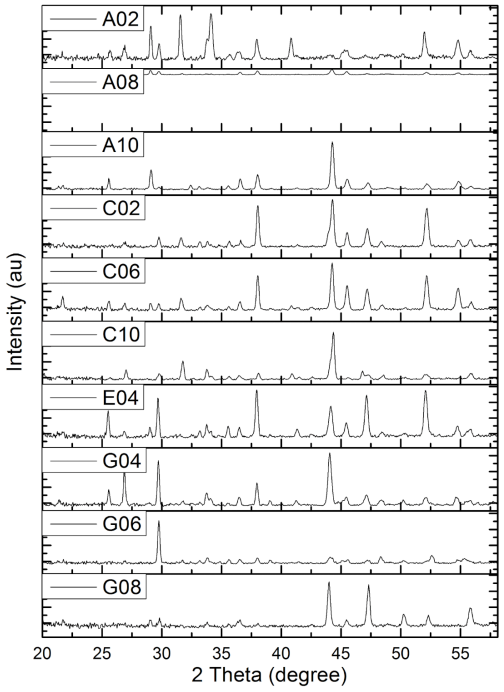
use slow vapor **transport** of an antisolvent to gradually reduce solubility of the precursor solution
Z. Li & M. Ani Najeeb et al. *ChemRxiv (2021) doi:10.33774/chemrxiv-2021-w2c7b*

Chemical diversity:

- Lead halides (single organic cation)
- Ruddlesden-Popper phases (multiple organic cations)
- Copper halides

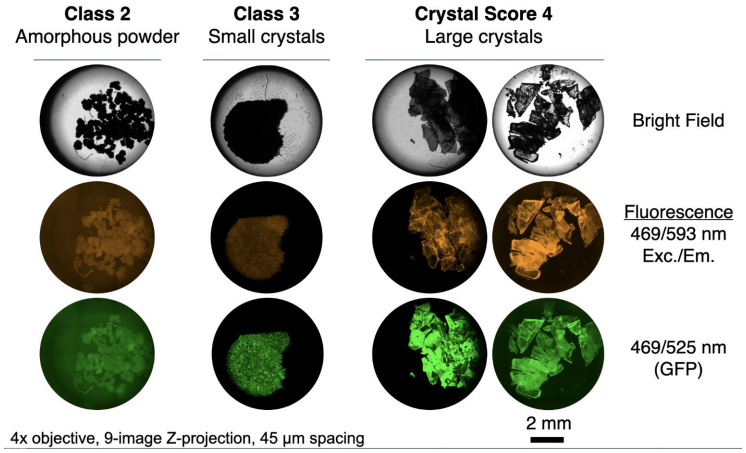
High-throughput characterization capabilities

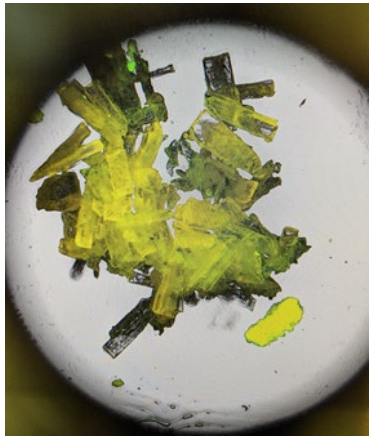
ITC - 96 parallel reactions, ~ 5 hr time scale
ASVC - 24 parallel reactions, ~18 hr time scale



Standard characterization

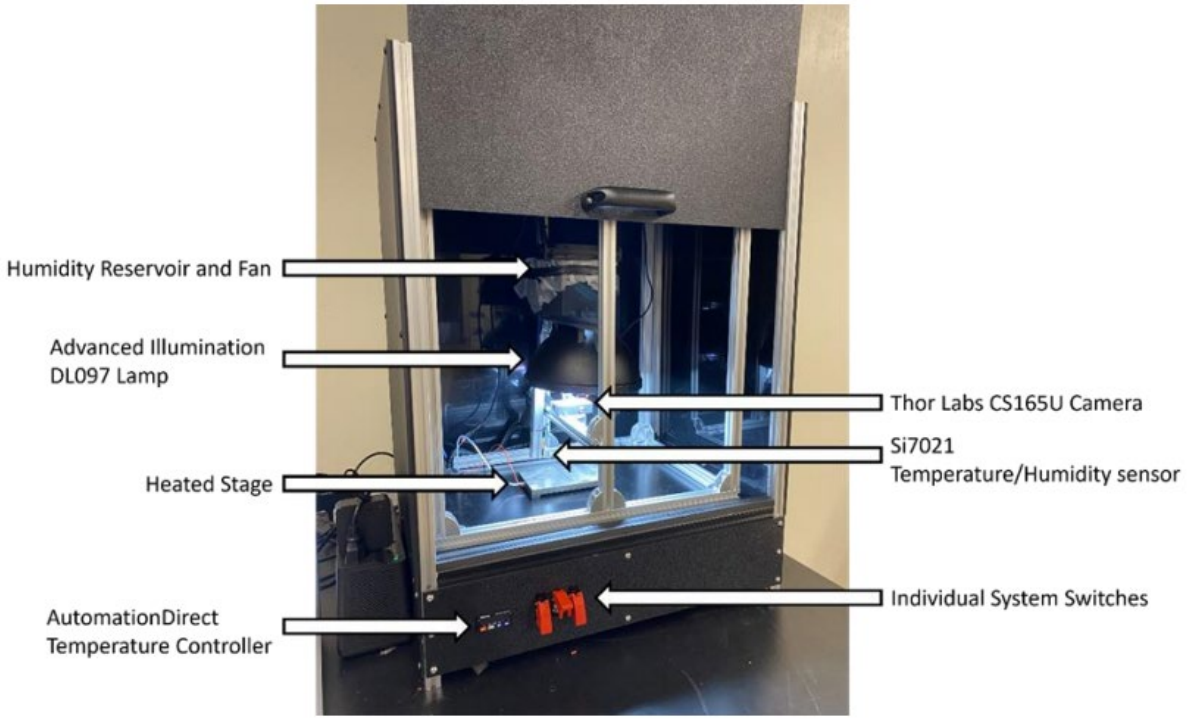
- X-ray diffraction
- Optical imaging microplate reader
- Microplate-based absorption & fluorescence spectroscopy





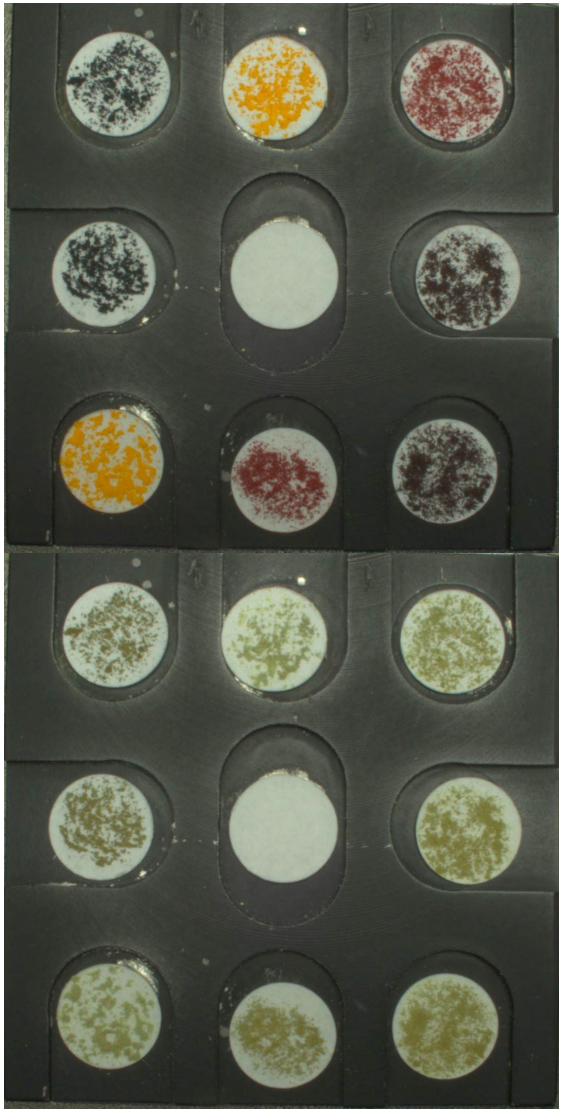
$[bdaH_2][PbI_3]$

Accelerated Sample Aging Chamber



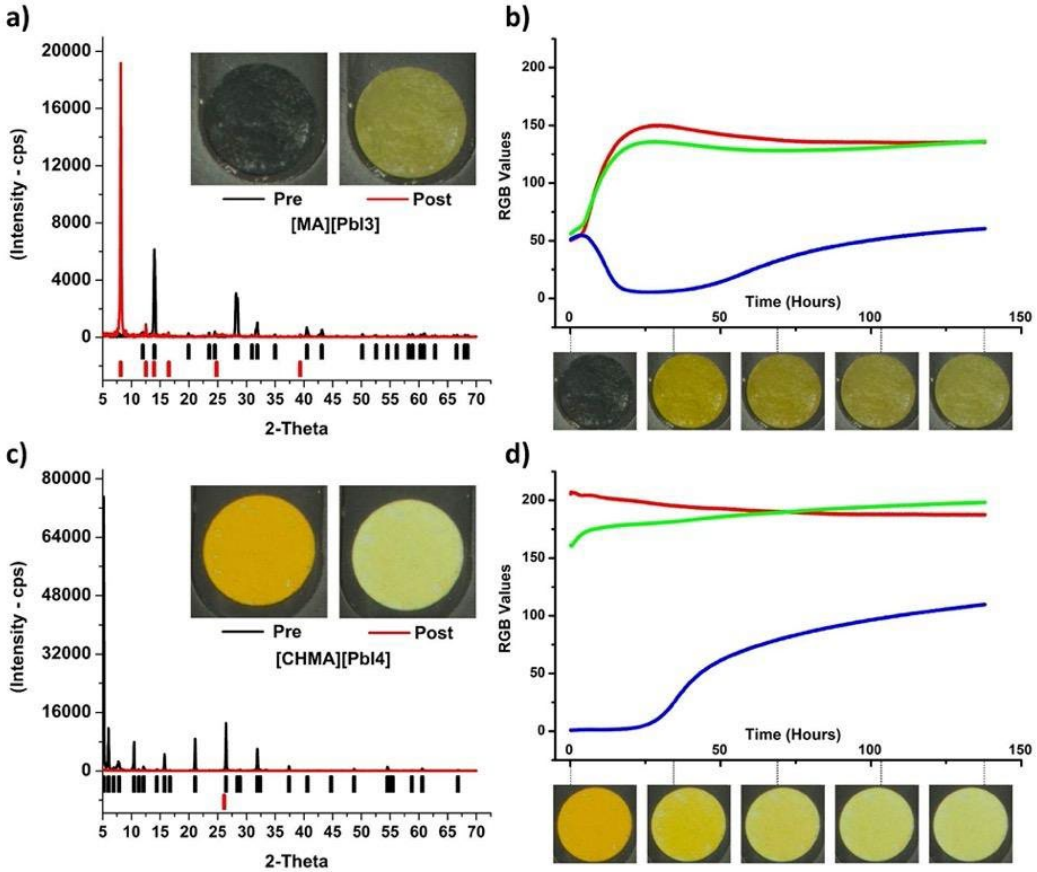
- Up to nine samples at the same time
- Illumination to 1.5 Sol
- Humidity to 86 % relative humidity
- Temperature to 85 ºC
- Images are collected at regular intervals
- Color calibration is achieved using RGB values from color standards
- Computer-vision with unsupervised clustering to measure degradation kinetics

R. Keesey ‘20 et al. *Digital Discovery* **2023** https://doi.org/10.1039/D2DD00089J



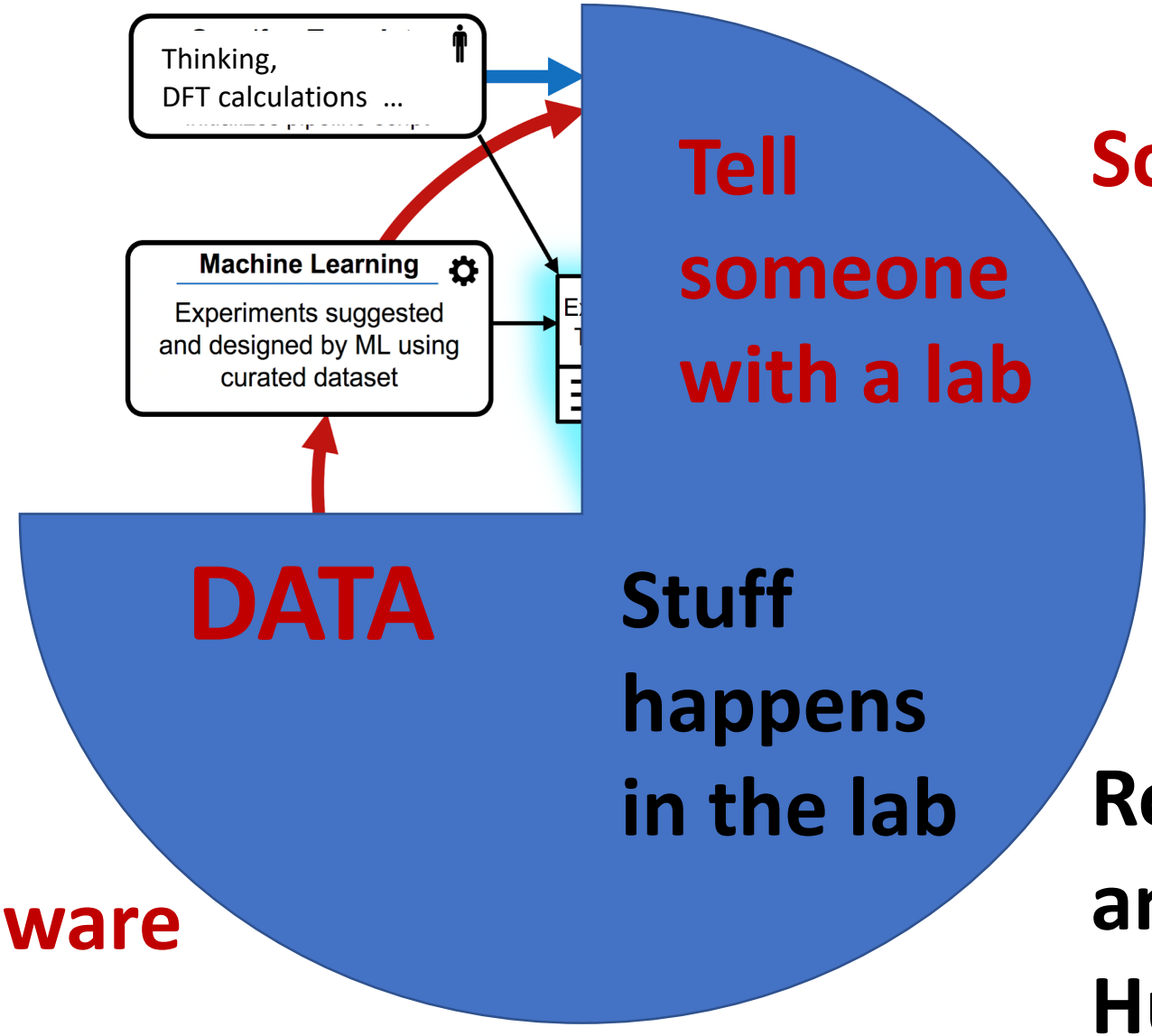
t=0 hours

t=160 hours

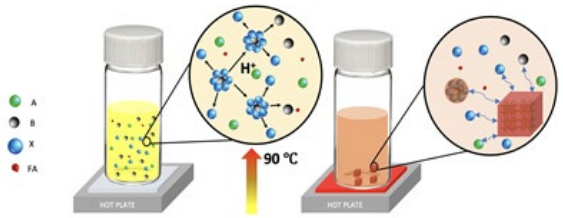


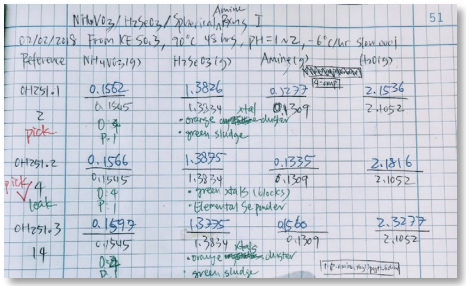
R. Keesey ‘20 et al. *Digital Discovery* **2023** https://doi.org/10.1039/D2DD00089J

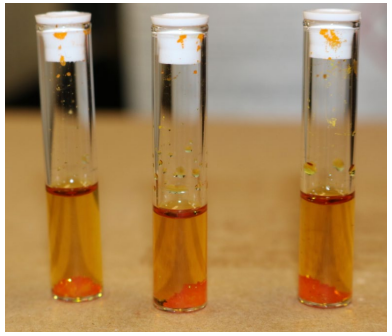
The need for experimental workflow management

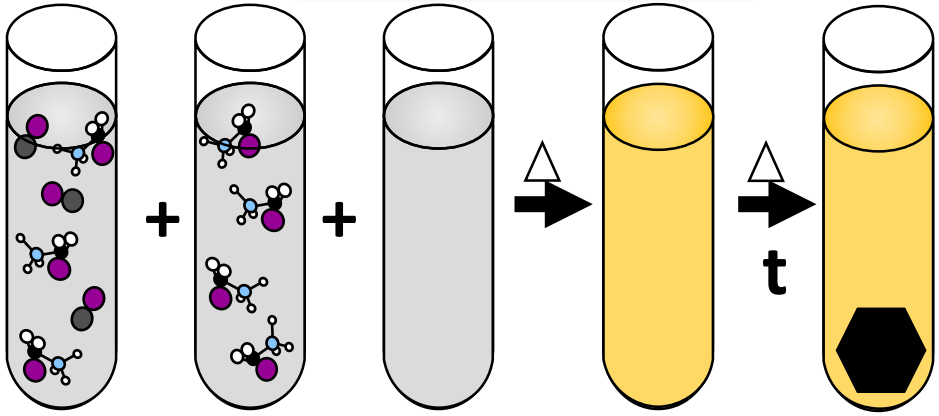


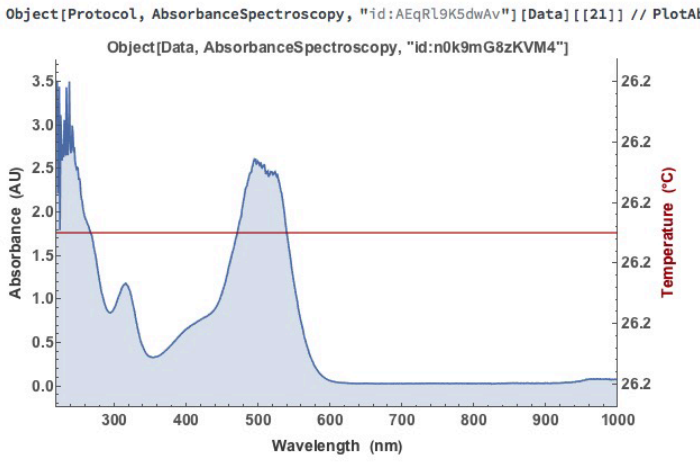
Challenge: Managing experiment plans and comprehensive data capture

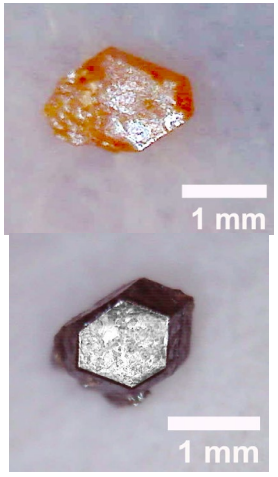


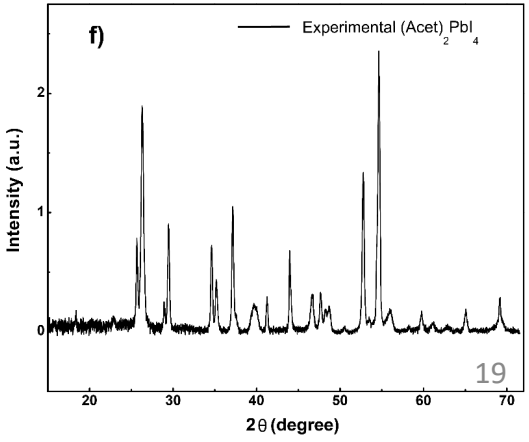




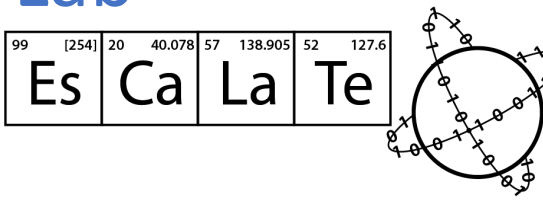








Experiment Specification Capture and Lab Automation Technology Environment



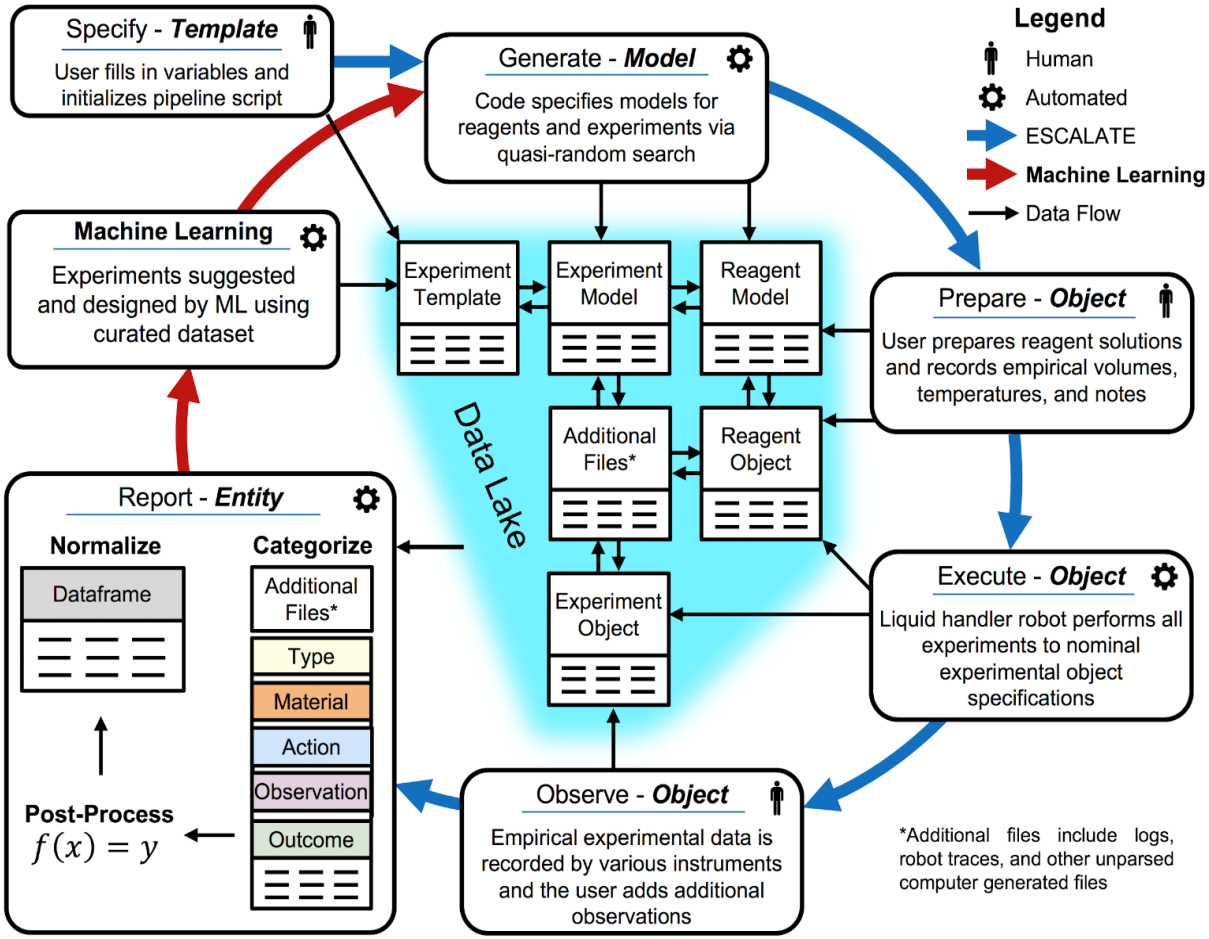
Provide an Application Programming Interface (API) for humans or algorithms to **specify** new experiments

Generate **instructions** for human operators and robots to conduct experiments.

Archive experimental data and metadata

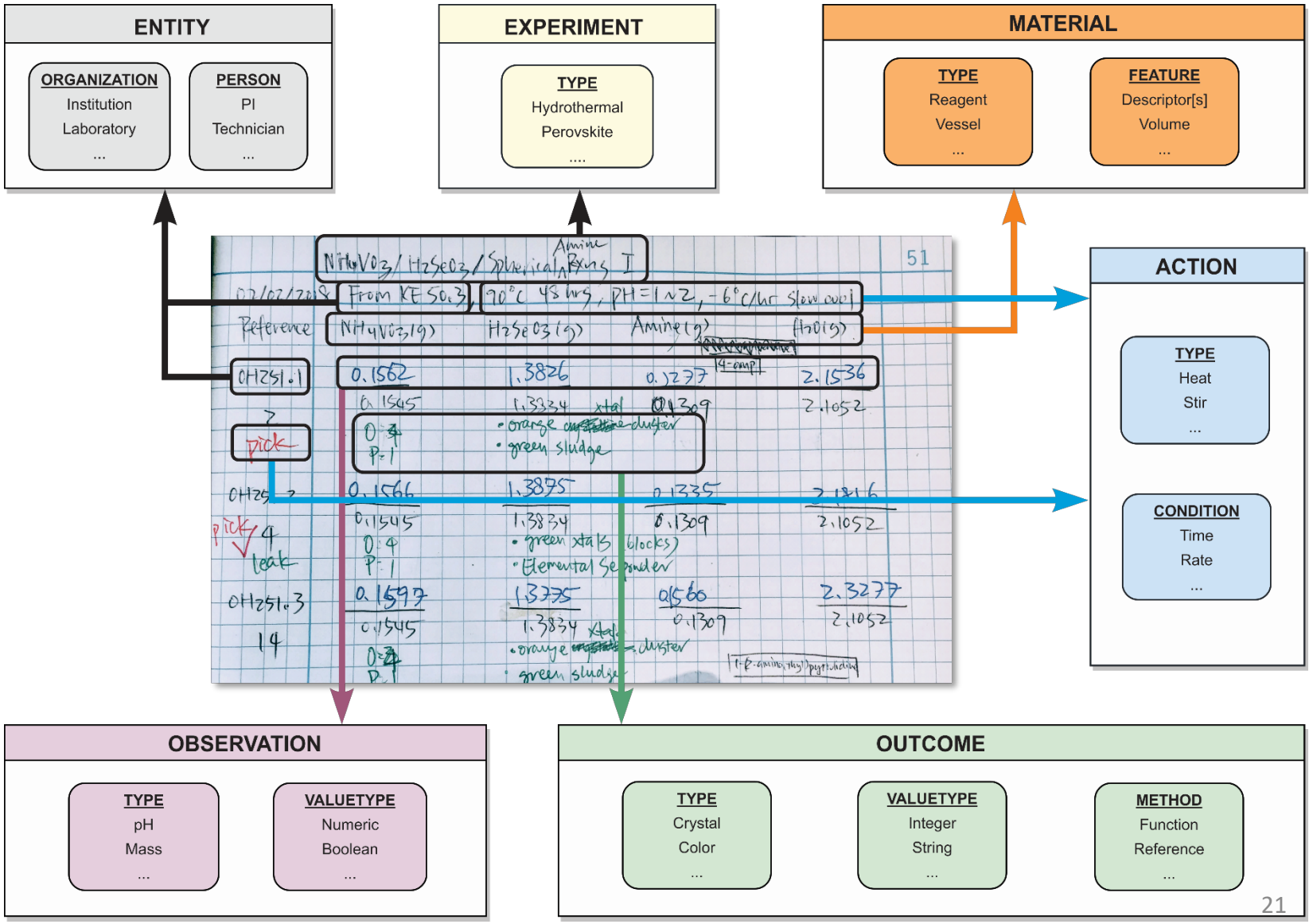
Add **interpretive layer** (cheminformatics, stoichiometric calculations, etc.) to collected data

Facilitate **data reporting and export**

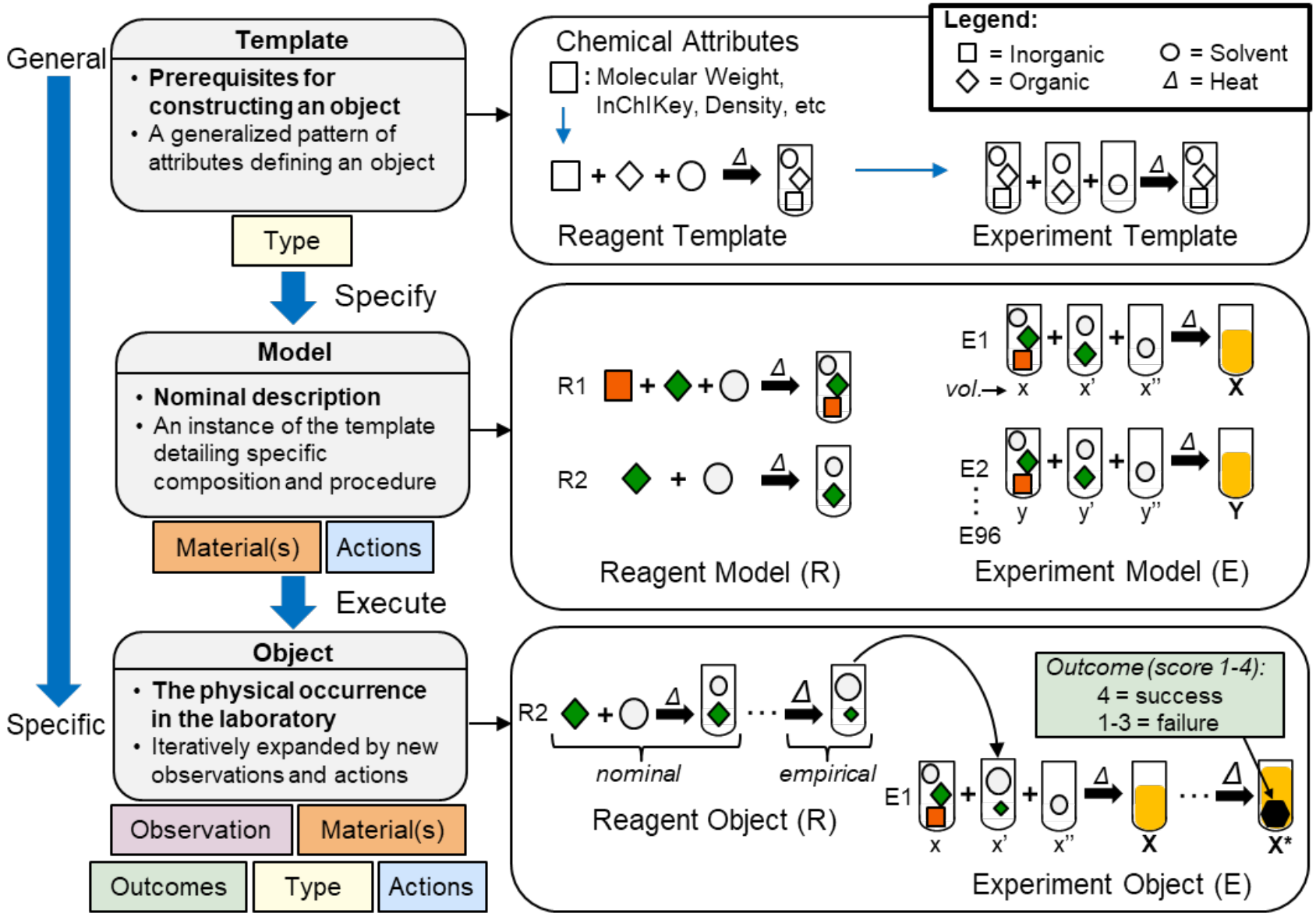


I.Pendleton & G. Cattabriga, et al. ***MRS Communications*** (2019) - doi:10.1557/mrc.2019.72
https://github.com/darkreactions/ESCALATE/

How do you describe an experiment?

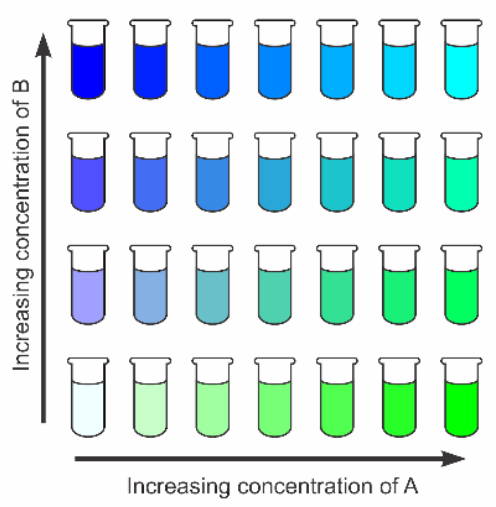


Experiment descriptions span levels of specificity.

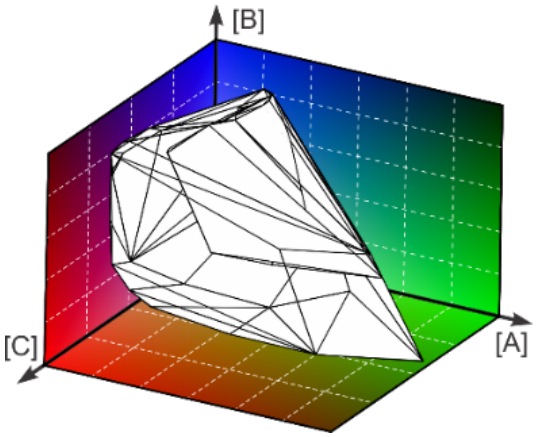


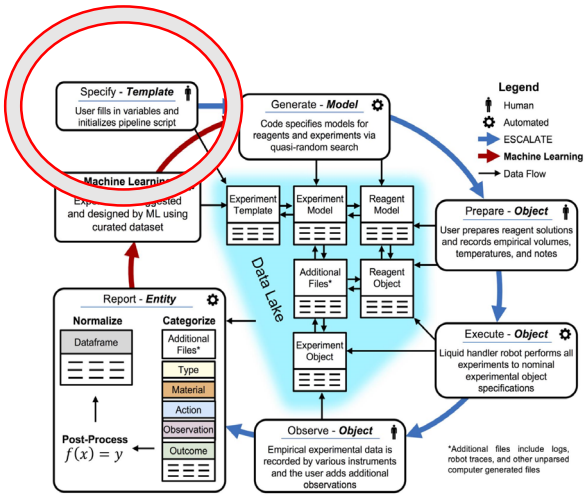
I. Pendleton & G. Cattabriga, et al. ***MRS Communications*** (2019) - doi:10.1557/mrc.2019.72 - https://github.com/darkreactions/

Templates constrain feasible experiments



What experiments are feasible, given the stock solution concentrations?

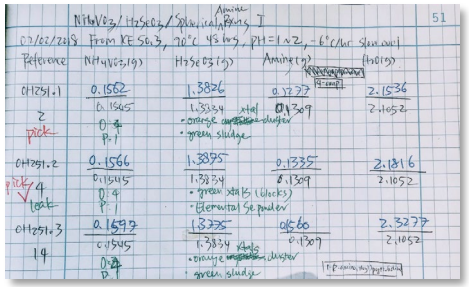




What are instrument limitations?

I. Pendleton & G. Cattabriga, et al. ***MRS Communications*** (2019) - doi:10.1557/mrc.2019.72 - https://github.com/darkreactions/

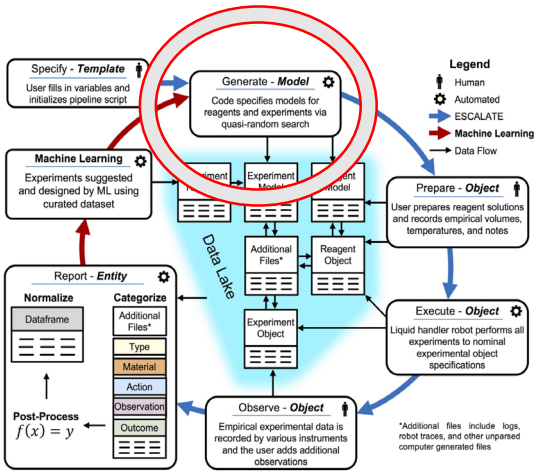
Models specify notional experiment plan

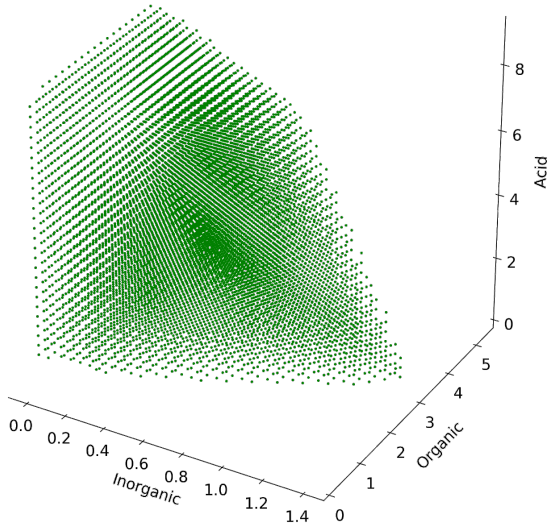


General specification of:

- *Materials* to be used
- *Actions* to be performed
- *Observations* to be collected

for the desired experiment





"Wildcard" parameters can be:

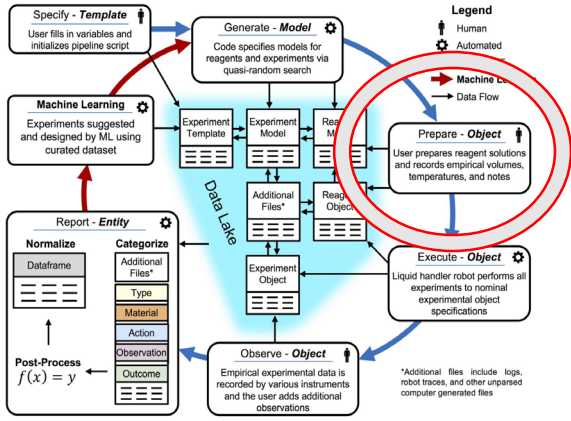
- *Sampled* within template constraints
- *Enumerated* to give a complete state set
- *Specified* explicitly by user or algorithm

I. Pendleton & G. Cattabriga, et al. ***MRS Communications*** (2019) - doi:10.1557/mrc.2019.72 - https://github.com/darkreactions/

(V2) Web-based forms to instruct and collect observations from human operators…

	Run Data	Reagent Preparation Information		
Date Created	2018-09-13	Reagent	Temp (C)	Stir (RPM)
Time Created _UTC	14_00_50	1	***null***	***null***
Laboratory	LBL	2	***75***	***450***
Operator Name	***Zhi Li***	3	***75***	***450***
Exp Workflow Ver	1.10			
Generator Workflow Ver	1.20			
Notes	Experiments performed based challenge problem recommendations			
Experimental Summary:	No modifications to workflow 1 protocol			

	Chemical Abbreviation	Nominal (Ideal) Amount	Actual (Empirical) Amount	Measurement Unit
Reagent 1	GBL	20.6	***30.0***	milliliter
Reagent 2	Final Volume =	27.2	***37.5***	milliliter
Chemical 1	PbI2	18.8	***18.7***	gram
Chemical 2	EtNH3I	14.1	***14.2***	gram
Chemical 3	GBL	27.2	***27.5***	milliliter
Reagent 3	Final Volume =	18.7	***17.0***	milliliter
Chemical 1	EtNH3I	9.5	***9.6***	gram
Chemical 2	GBL	9.2	***9.2***	milliliter



Google sheets front-end…
…but a JSON data object!

I. Pendleton & G. Cattabriga, et al. ***MRS Communications*** (2019) - doi:10.1557/mrc.2019.72 - https://github.com/darkreactions/

Querying Available Materials

REST API & GUI allows creation and retrieval of

- **Materials**
- Properties
- Action Definitions
- Experiment Specifications
- Results

```
r = get_data(endpoint='material',
             data={'fields': 'description'})
r
```

```
<Response [200]>
GET: OK
Found 294 resources, returning list of dicts)

[{'description': 'Gamma-Butyrolactone'},
 {'description': 'Dimethyl sulfoxide'},
 {'description': 'Formic Acid'},
 {'description': 'Lead Diiodide'},
 {'description': 'Ethylammonium Iodide'},
 {'description': 'Phenethylammonium iodide '},
 {'description': 'Acetamidinium iodide'},
 {'description': 'n-Butylammonium iodide'},
```

Full tutorial: https://github.com/darkreactions/ESCALATE/blob/master/demonstrations/REST_API_DEMO.ipynb
https://github.com/darkreactions/ESCALATE/blob/master/UI%20User%20Guide.docx

Additional material properties can be defined by user

REST API allows creation and retrieval of

- Materials
- **Properties**
- Action Definitions
- Experiment Specifications
- Results

```
r = get_data('propertydef',
             {'fields': ['description']}
            )
r
```

```
<Response [200]>
GET: OK
Found 15 resources, returning list of dicts)

[{'description': 'particle-size {min, max}'},
 {'description': 'mesh {min, max}'},
 {'description': 'capacity'},
 {'description': 'cross-linkage %'},
 {'description': 'moisture % {min, max}'},
```

Full tutorial: https://github.com/darkreactions/ESCALATE/blob/master/demonstrations/REST_API_DEMO.ipynb
https://github.com/darkreactions/ESCALATE/blob/master/UI%20User%20Guide.docx

Querying material properties

```
r = post_data('materia
              {'materia
               'propert
               'value':
              }
          )
```

```
<Response [201]>
POST: OK, returning ne
```

Search

ADD MATERIAL

Export to csv

Chemical Name	Identifiers
1,3-dimethylpyridinium iodide	Chemical_Name: 1,3-dimethylpyridinium iodide Abbreviation: 13dimethylpyridinium InChI: InChI=1S/C7H10N.HI/c1-7-4-3-5-8(2)6-7;/h3-6H,1-2H3;1H/q+1; InChIKey: BVUPLMFUPRCOEH-UHFFFAOYSA-N SMILES: CC1=C[N+](=CC=C1)C.I Molecular_Formula: C7H11IN+
1,4-Benzene diammonium iodide	Chemical_Name: 1,4-Benzene diammonium iodide Abbreviation: Benzenediaminedihydroiodide InChIKey: RYYSZNVPBLKLRS-UHFFFAOYSA-N SMILES: C1=CC(=CC=C1[NH3+])[NH3+].[I-].[I-] Molecular_Formula: C6H10I2N2 InChI: InChI=1S/C6H8N2.2HI/c7-5-1-2-6(8)4-3-5;;/h1-4H,7-8H2;2*1H
1,4-Benzenediammonium bromide	Chemical_Name: 1,4-Benzenediammonium bromide Abbreviation: BenzenediammoniumBromide InChI: InChI=1S/C6H8N2.2BrH/c7-5-1-2-6(8)4-3-5;;/h1-4H,7-8H2;2*1H InChIKey: KTHOOQOVVOZMIG-UHFFFAOYSA-N SMILES: [NH3+]c1ccc([NH3+])cc1.[Br-].[Br-] Molecular_Formula: C6H10Br2N2

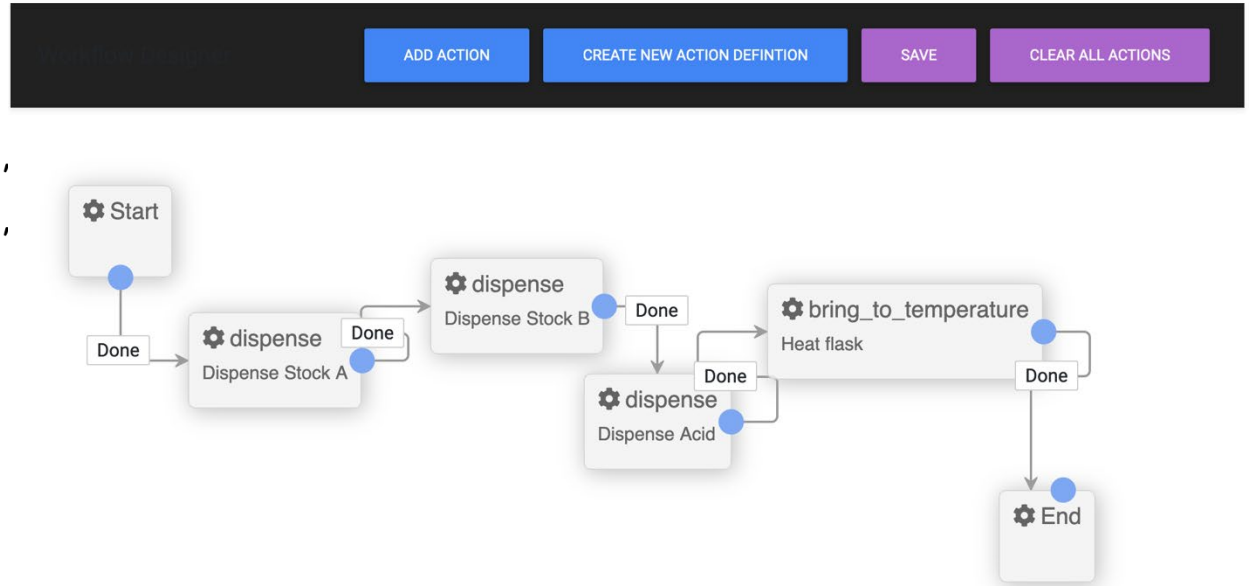
Full tutorial: https://github.com/darkreactions/ESCALATE/blob/master/demonstrations/REST_API_DEMO.ipynb
https://github.com/darkreactions/ESCALATE/blob/master/UI%20User%20Guide.docx

Workflows organize action/material relationships

```
perovskite_demo = get_data('experimenttemplate',
         {'description': 'perovskite_demo',
          'expand': 'workflow'})
[wf['description'] for wf in perovskite_demo['workflow']]
```

```
<Response [200]>
GET: OK
Found one resource, returning dict

['Perovskite Demo: Preheat Plate',
 'Perovskite Demo: Prepare Stock A',
 'Perovskite Demo: Prepare Stock B',
 'Perovskite Demo: Dispense Solvent',
 'Perovskite Demo: Dispense Stock A',
 'Perovskite Demo: Dispense Stock B',
 'Perovskite Demo: Dispense Acid Vol 1',
 'Perovskite Demo: Heat Stir 1',
 'Perovskite Demo: Dispense Acid Vol 2',
 'Perovskite Demo: Heat Stir 2',
 'Perovskite Demo: Heat']
```



Full tutorial: https://github.com/darkreactions/ESCALATE/blob/master/demonstrations/REST_API_DEMO.ipynb
https://github.com/darkreactions/ESCALATE/blob/master/UI%20User%20Guide.docx

Default action definitions & parameters

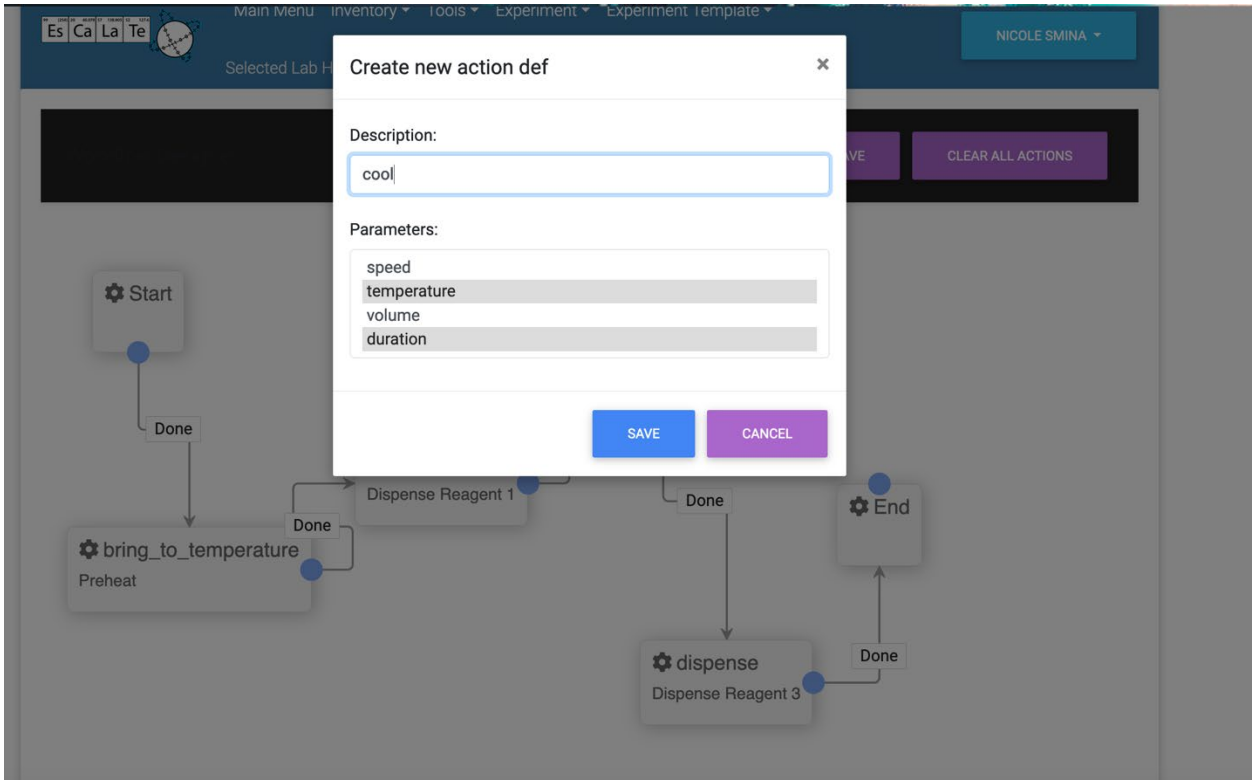
```
get_data('actiondef',
        {'expand': 'parameter_def',
         'fields': ['description',
                    'parameter_def.description',
                    'parameter_def.unit_type']
        })
```

```
<Response [200]>
GET: OK
Found 7 resources, returning list of dicts)

[{'description': 'dispense',
  'parameter_def': [{'description': 'volume', 'unit_type': 'volume'
 {'description': 'heat_stir',
  'parameter_def': [{'description': 'speed', 'unit_type': 'rate'},
   {'description': 'temperature', 'unit_type': 'temperature'},
   {'description': 'duration', 'unit_type': 'time'}]},
```

Full tutorial: https://github.com/darkreactions/ESCALATE/blob/master/demonstrations/REST_API_DEMO.ipynb
https://github.com/darkreactions/ESCALATE/blob/master/UI%20User%20Guide.docx

Creating new action definitions



Full tutorial: https://github.com/darkreactions/ESCALATE/blob/master/demonstrations/REST_API_DEMO.ipynb
https://github.com/darkreactions/ESCALATE/blob/master/UI%20User%20Guide.docx

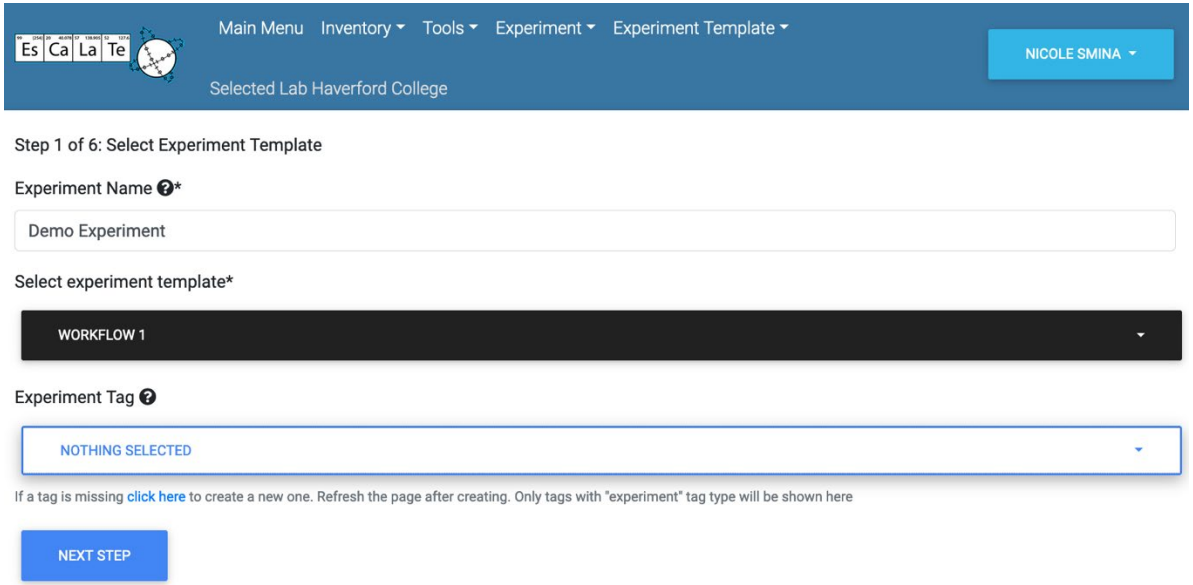
New experiments are instances of workflows

```
resp = post_data(
    f'experimenttemplate/{perovskite_demo["uuid"]}/create',
    editable_template
)
```

```
<Response [200]>
POST: OK, returning new resource dict
```

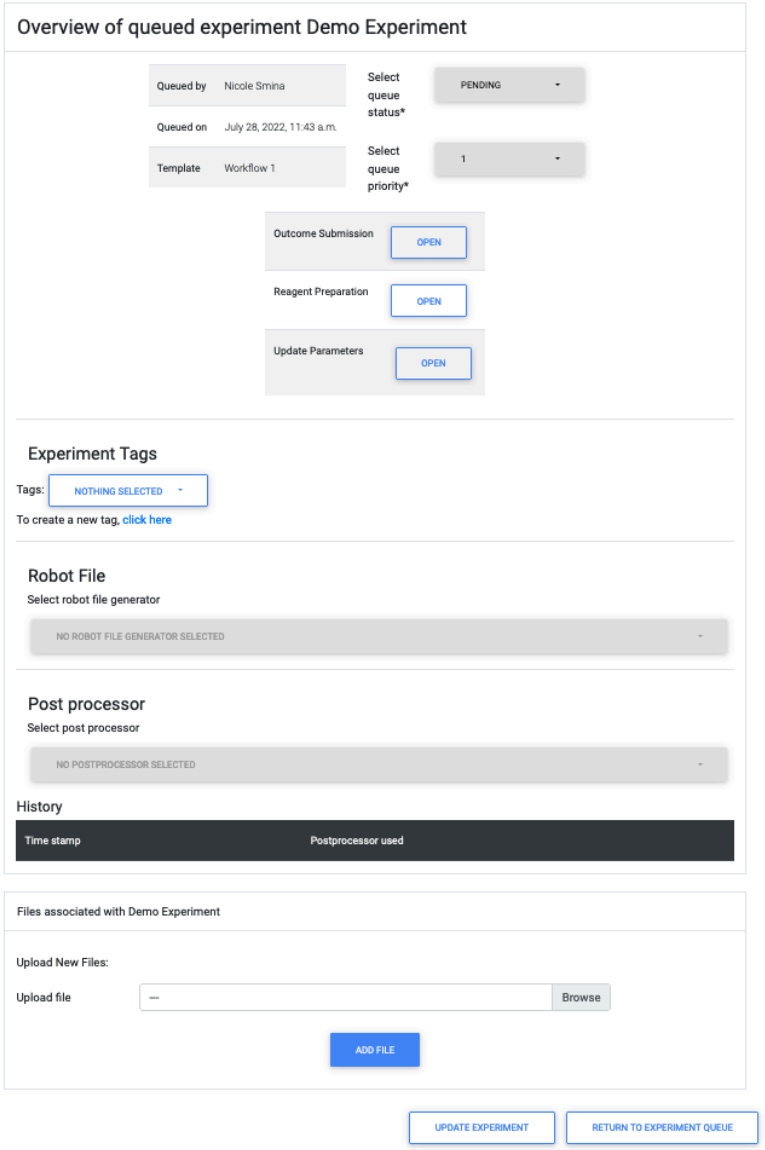
```
resp
```

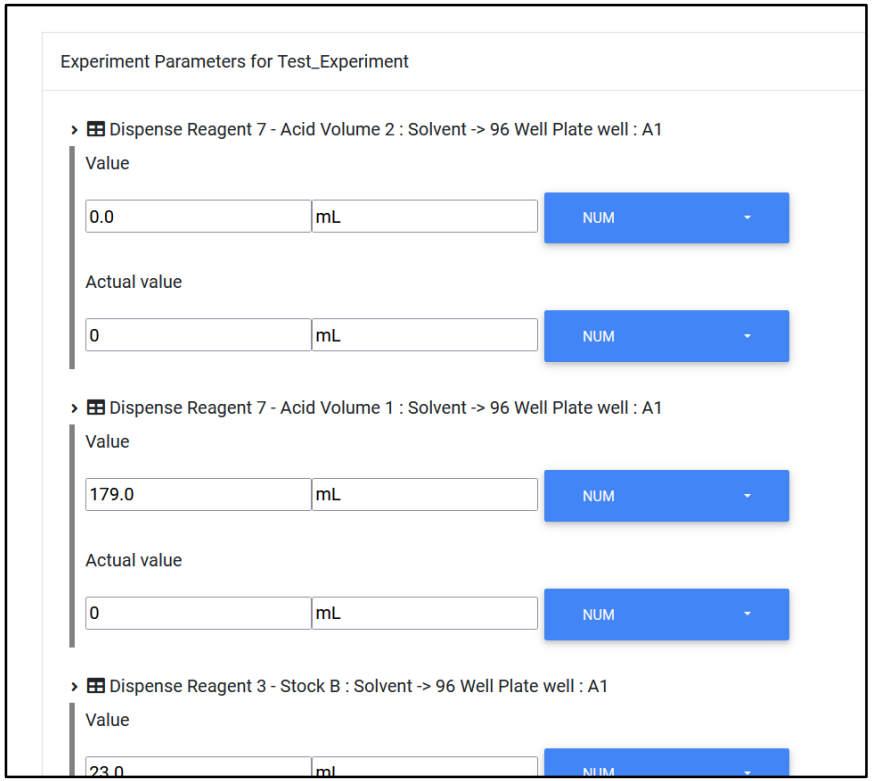
```
{'new_experiment_created': 'http://localhost:8000/api/experime
```



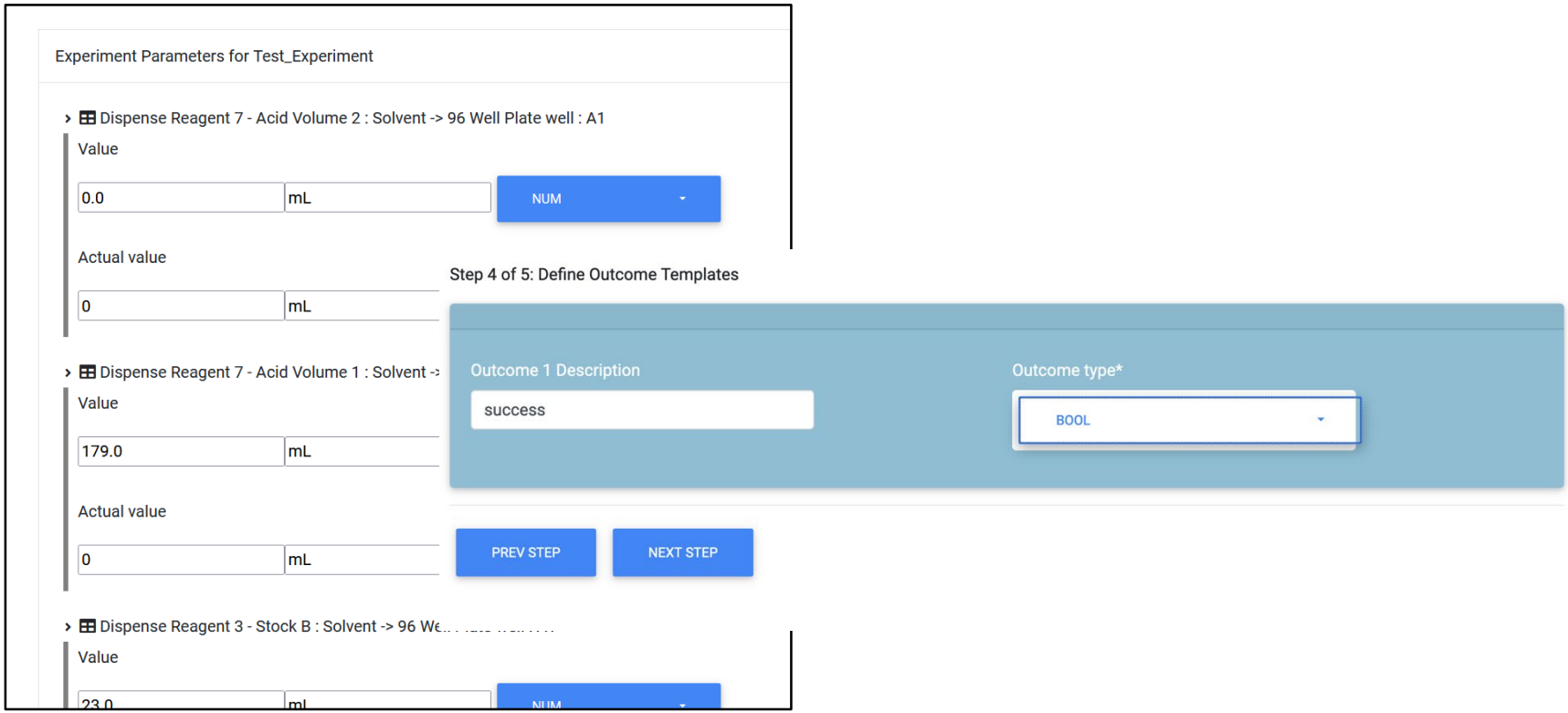
Full tutorial: https://github.com/darkreactions/ESCALATE/blob/master/demonstrations/REST_API_DEMO.ipynb
https://github.com/darkreactions/ESCALATE/blob/master/UI%20User%20Guide.docx

Experiment queue and instructions performance in the laboratory





Collect data as defined in the experiment template



Full tutorial: https://github.com/darkreactions/ESCALATE/blob/master/demonstrations/REST_API_DEMO.ipynb
https://github.com/darkreactions/ESCALATE/blob/master/UI%20User%20Guide.docx

Stored data can be retrieved on demand

REST API allows creation and retrieval of

- Materials
- Properties
- Action Definitions
- Experiment Specifications
- **Results**

```
crystal_scores = get_data('measure',
                          {'measuredef':
                                (get_data('measuredef',
                                          {'description': 'crystal_score'})['url']
                                )
                          })
results['crystal_score'] = crystal_scores['measure_value']['value']
```

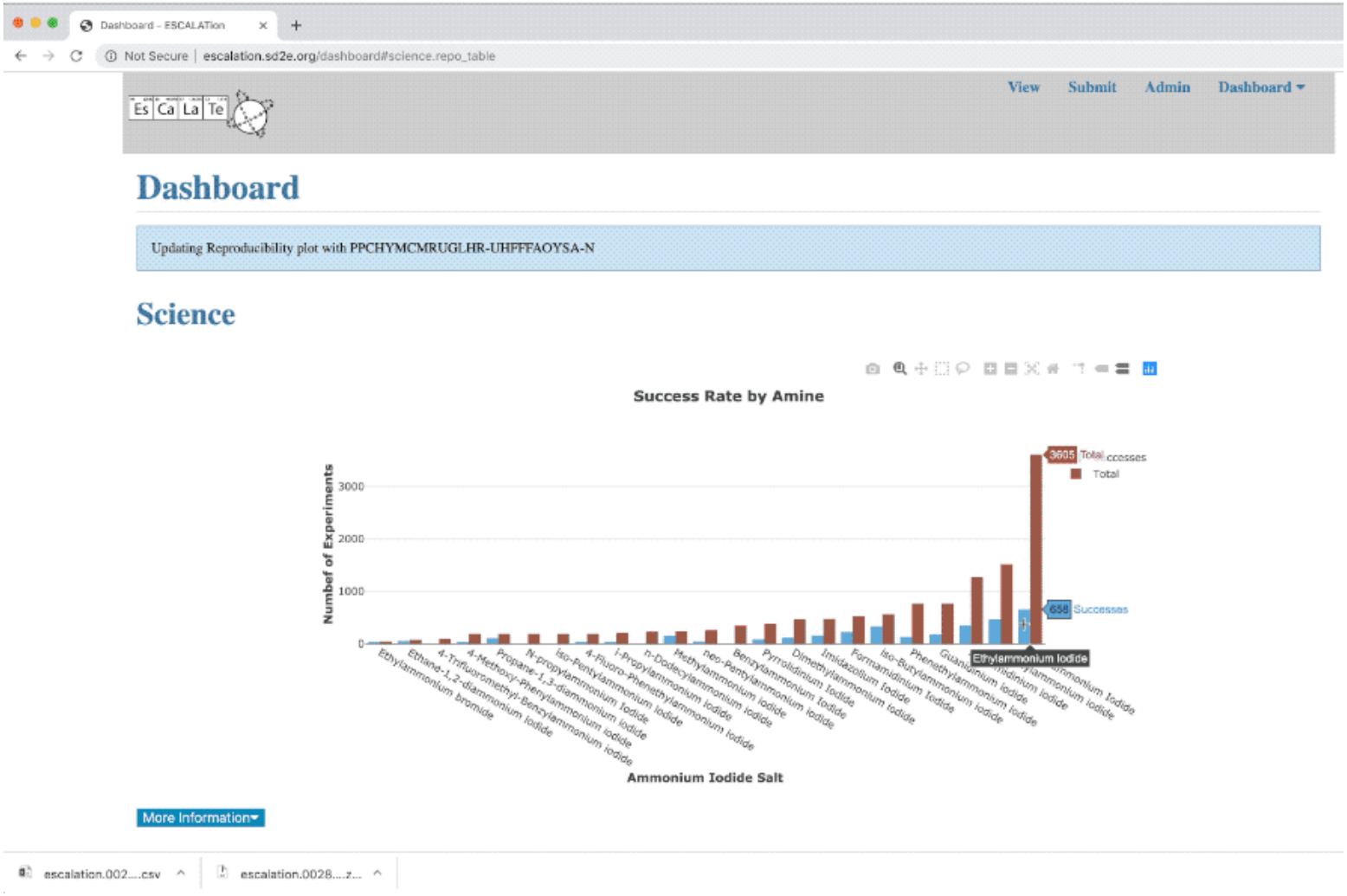
```
<Response [200]>
GET: OK
Found one resource, returning dict
<Response [200]>
GET: OK
Found one resource, returning dict
```

```
results
```

		action_source	Acid	Solvent	Stock A	Stock B	crystal_score
experiment_id	action_dest						
59c19e0e-fafd-4562-a6af-73b24d7778d4	Plate: 96 Well Plate well#: A1	0.0	0.0	0.0	0.0	1
	Plate: 96 Well Plate well#: A10	0.0	0.0	0.0	0.0	1

Full tutorial: https://github.com/darkreactions/ESCALATE/blob/master/demonstrations/REST_API_DEMO.ipynb
https://github.com/darkreactions/ESCALATE/blob/master/UI%20User%20Guide.docx

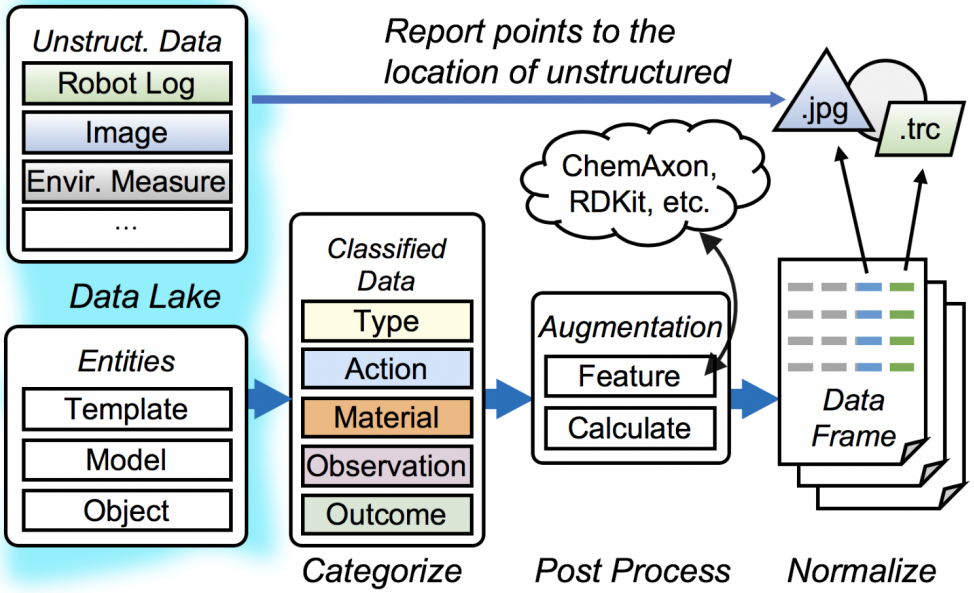
escalation live data dashboard

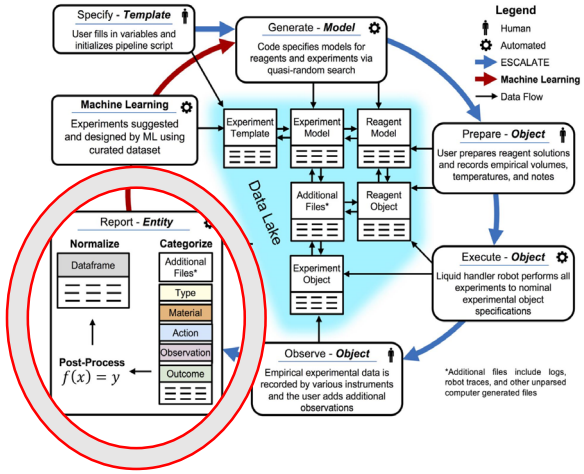


~~http://escalation.sd2e.org/~~

https://github.com/twosixlabs/escalation

Export tabular *reports* from the unstructured data



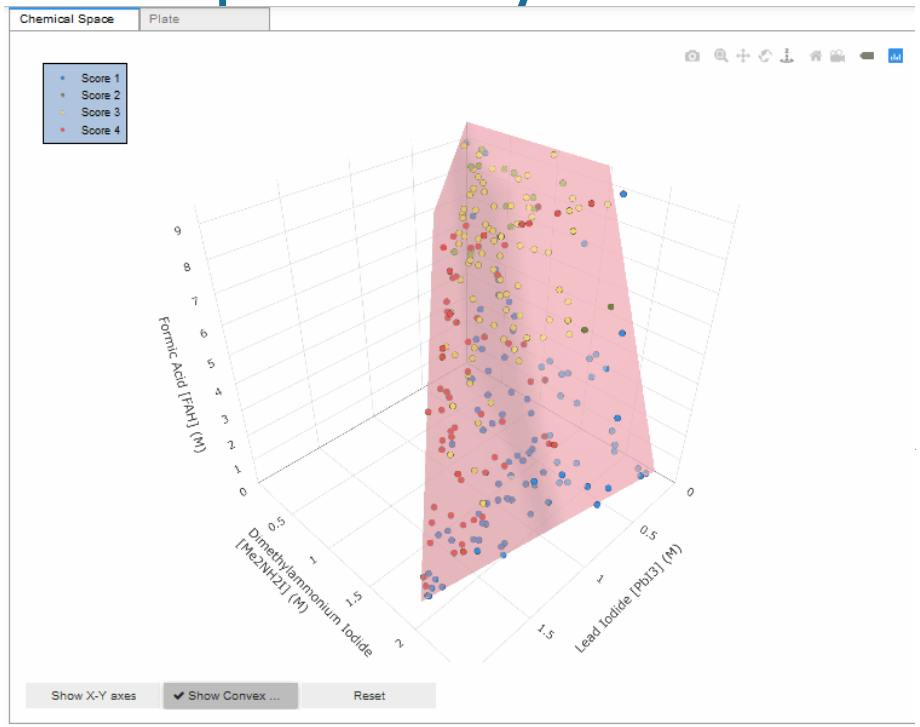


Automate the featurization with two python modules:

(1) chemdescriptor: interfaces with RDKit, ChemAxon, and Mordred to automate featurization of organic molecule components https://github.com/darkreactions/chemdescriptor

(2) calculator: provide safe evaluation of arbitrary numerical calculations involving observed features/descriptors

Comprehensive data capture enables new types of *data-enabled publications* enabling reproducibility and replicability.



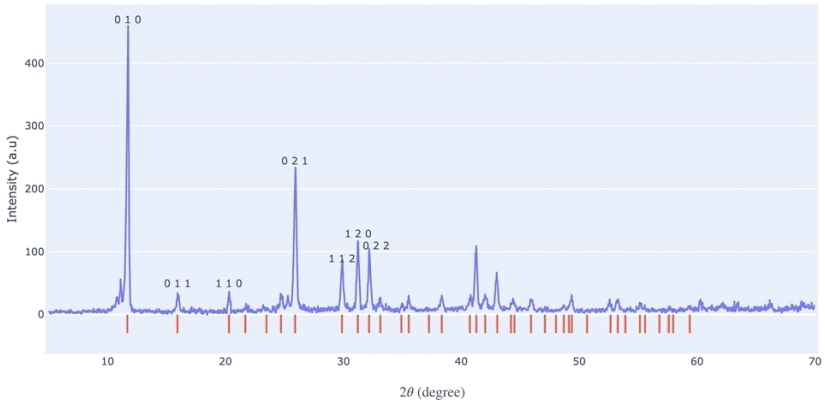
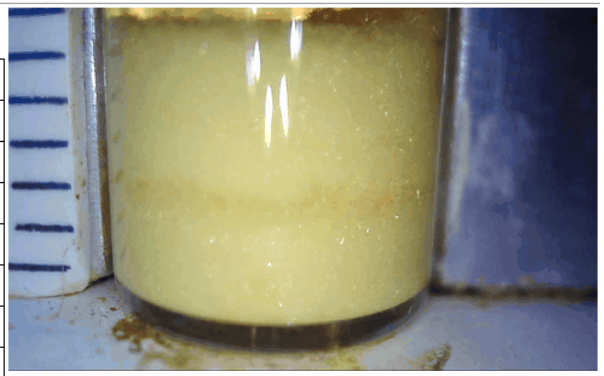


Plate ID:
2019-06-10T21_14_00.866379+00_00_LBL

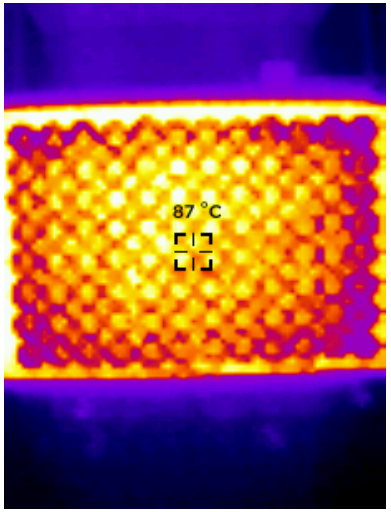
Well ID	B9
Formic Acid [FAH]	5.619
Lead Iodide [PbI2]	0.9
Dimethylammonium Iodide [Me2NH2I]	1.602
Mixing Time Stage 1 (s)	900.0
Mixing Time Stage 2 (s)	1200.0
Reaction Time (s)	12600.0
Stir Rate (RPM)	750.0
Temperature (C)	95.0



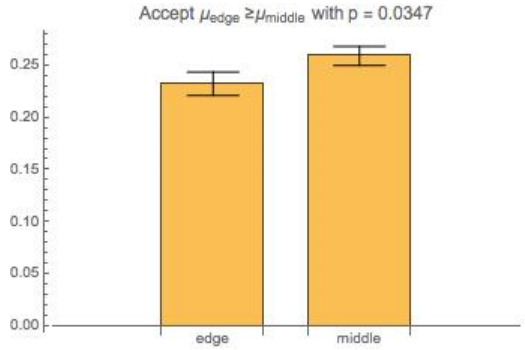
jupyter binder

Zhi Li & M. Ani Najeeb et al. Chem. Mater. (2020) doi:10.1021/acs.chemmater.0c01153

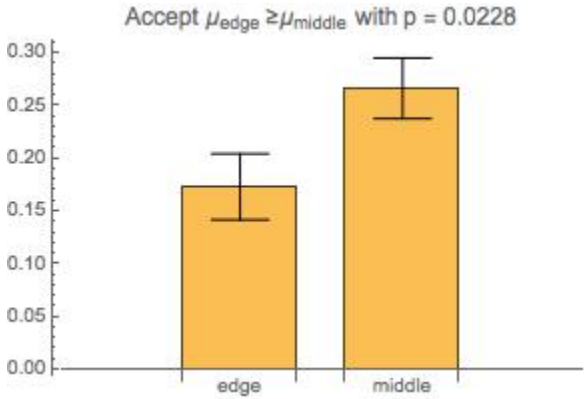
What can you do with comprehensive data capture?



Answer: Crystallization success is 2% less likely at edges than middle over all reactions.



Effect is larger in some amines (e.g., guanidinium)



Data set: TACC/0026.perovskitedata.csv
Method: Error bars estimated by bootstrapping; One sided p-values calculated by shuffling
Analysis: https://www.wolframcloud.com/obj/jschrier0/Published/2019.07.09_perovskite_edge_success.nb

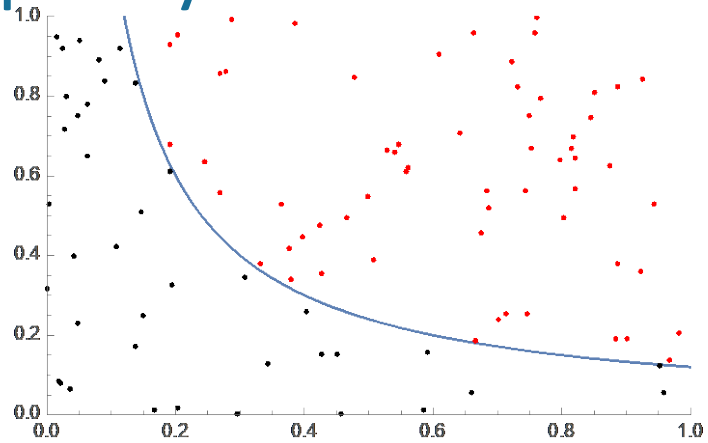
Crystallization success by amine

Mean success probabilities

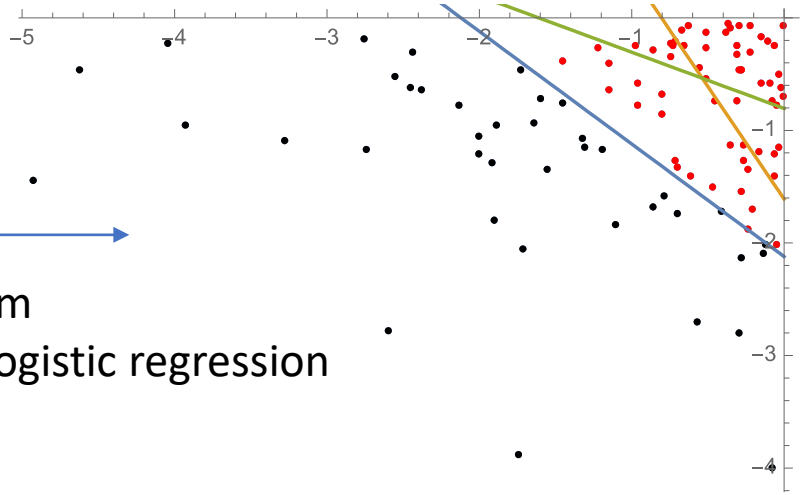
		N_{expt}	edge: 4	interior: 4	edge: 3+	interior: 3+
	Phenethylammonium iodide	96	0	0	0.0556	0.0333
	n-Butylammonium iodide	112	0.0652	0.0909	0.326	0.273
	Ethylammonium Iodide	1461	0.183	0.190	0.272	0.295
p= 0.004	Methylammonium iodide	112	0.478	0.742	0.761	0.833
p= 0.024	Guanidinium iodide	384	0.174	0.267	0.285	0.358
	Acetamidinium iodide	336	0.230	0.257	0.452	0.429
	Formamidinium Iodide	216	0.422	0.460	0.578	0.802
	Imidazolium Iodide	192	0.347	0.342	0.514	0.617
	Benzylammonium Iodide	132	0	0	0.0175	0.0133
	neo-Pentylammonium iodide	24	0.0833	0.417	0.167	0.500
	i-Propylammonium iodide	24	0.250	0.0833	0.333	0.0833
	iso-Butylammonium iodide	240	0.633	0.553	0.633	0.553
	Dimethylammonium iodide	96	0.194	0.317	0.861	0.850
	n-Dodecylammonium iodide	96	0	0	0.0278	0
p= 0.035	Pyrrolidinium Iodide	192	0.153	0.275	0.167	0.292
	Ethane-1,2-diammonium iodide	40	0.813	0.667	0.813	0.667

Data set: TACC/0026.perovskitedata.csv
Method: Error bars estimated by bootstrapping; One sided p-values calculated by shuffling
Analysis: https://www.wolframcloud.com/obj/jschrier0/Published/2019.07.09_perovskite_edge_success.nb

Use machine learning models for batch-to-batch quality control



Log transform
Becomes a logistic regression



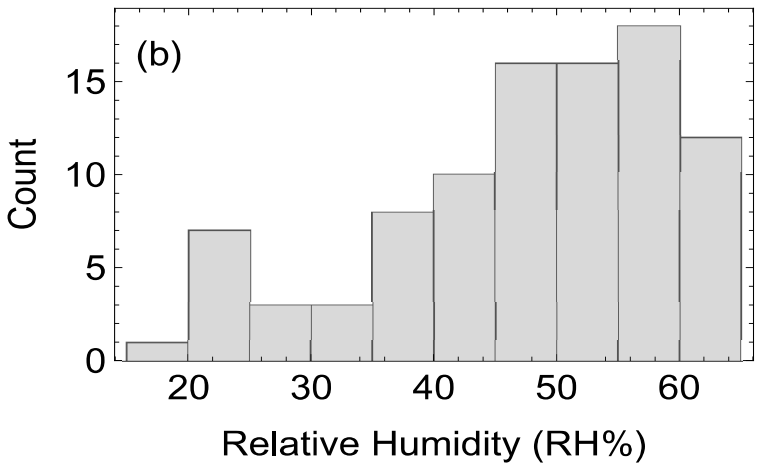
$K_{sp} = [A]^a [B]^b$

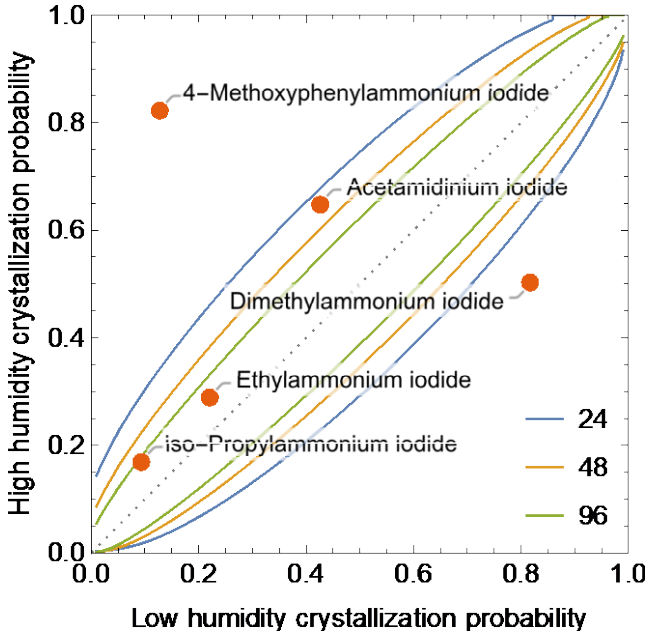
$\ln K_{sp} - a \ln [A] + b [B] > 0$

amine	MCC	precision	recall	empiricalCrystallizationRate	highQualityCrystallizationRate	nExpts	solvent	parameters
N,N-Dimethylethane- 1,2-diammonium iodide	1.0	1.0	1.0	0.0208333	0.0208333	96	{Dimethylformamide}	{-281.967, 89.1619, 133
4-Fluoro-Benzylammonium iodide	0.941816	0.956522	0.956522	0.252747	0.197802	91	{Dimethylformamide}	{-8.9111, 6.19455, 2.281
Cyclohexylammonium iodide	0.938867	0.952381	0.952381	0.221053	0.105263	95	{Dimethyl sulfoxide}	{-41.6045, 18.7744, 0.72
t-Butylammonium Iodide	0.934211	0.947368	0.947368	0.2	0.2	95	{Gamma-Butyrolactone}	{-60.5375, 35.232, 19.43
i-Propylammonium iodide	0.925451	0.954545	0.913043	0.11165	0.0631068	206	{Gamma-Butyrolactone}	{-218.29, 117.815, 81.93
4-Methoxy-Phenylammonium iodide	0.92168	0.94	0.959184	0.352518	0.0863309	139	{Gamma-Butyrolactone}	{-46.7178, 23.457, 22.85
N-propylammonium Iodide	0.880414	0.888889	0.888889	0.0708661	0.0	127	{Gamma-Butyrolactone}	{-16.9666, 8.03303, 10.6
4-Trifluoromethyl-Benzylammonium iodide	0.861254	0.75	1.0	0.0319149	0.0	94	{Gamma-Butyrolactone}	{-1590.88, 579.759, 498
iso-Butylammonium iodide	0.833309	0.908537	0.949045	0.560714	0.560714	280	{Gamma-Butyrolactone}	{-10.1541, 4.32014, 0.91
Propane-1,3-diammonium iodide	0.832774	0.911765	0.885714	0.4	0.342857	175	{Dimethylformamide}	{-23.7089, 12.8737, 4.62
N,N-dimethylpropane- 1,3-diammonium iodide	0.831411	0.918367	0.918367	0.515789	0.168421	95	{Dimethylformamide}	{-4.85746, 4.98338, 0.28
Phenylammonium Iodide	0.810443	0.9	0.84375	0.333333	0.302083	96	{Gamma-Butyrolactone}	{-4.12565, 2.27743, 1.69

Use extreme (meta)data collection as a substitute for experimental control

- Lab humidity is hard to control, but we can measure it.
- Consider these fluctuations as natural experiments.





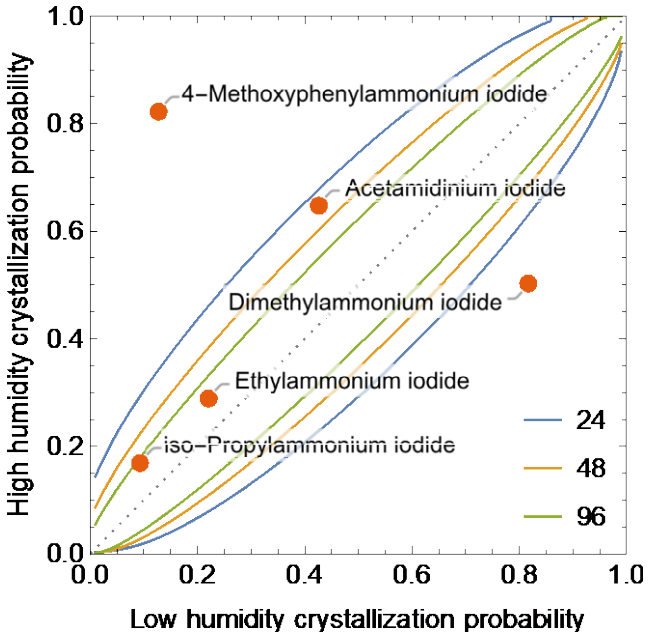
Analyze 8000+ historical experiments, find cases where similar reactions were performed under both *high* and *low* humidity conditions.

Look for discrepancies between the reaction outcomes at high and low humidity

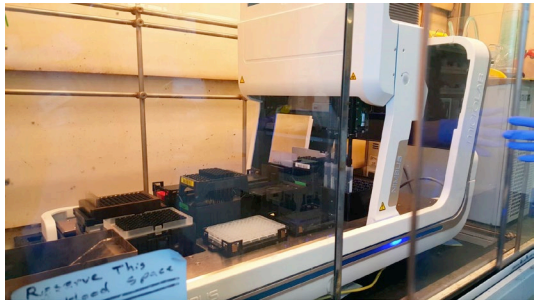
Use a statistical model to prioritize the most interesting experiments to explore.

P Nega et al. *Appl. Phys. Lett*. **119**, (2021) 041903 doi:10.1063/5.0059767

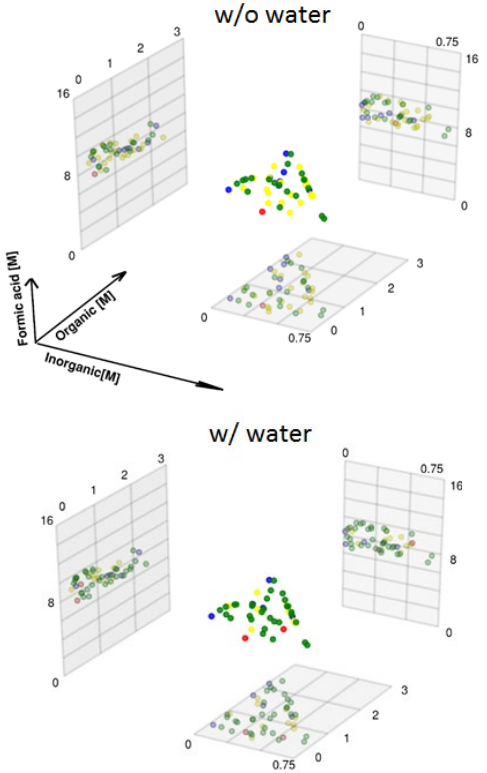
Limited Sloppiness + Data Analysis = ”Automated serendipity”



Find discrepant batches and prioritize by the sample size needed to confirm the observed effect



Use the robot to do >1200+ deliberate paired experiments (with and without added water)



Acquire statistically significant confirmation of the effect

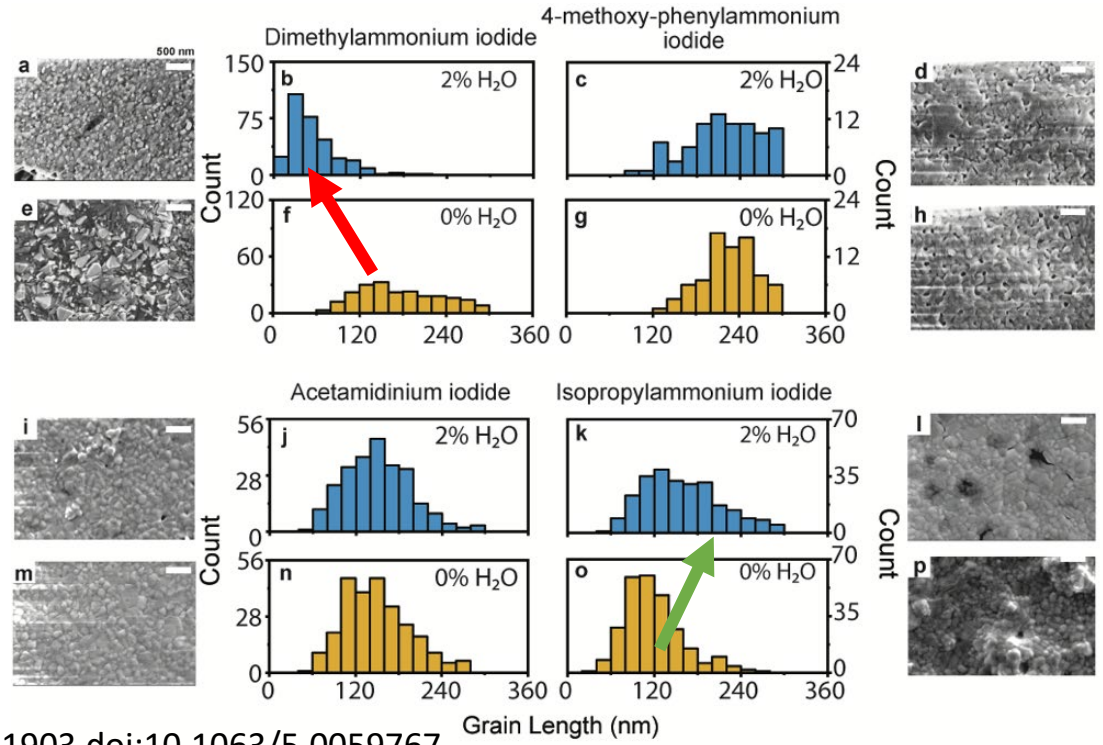
P Nega et al. *Appl. Phys. Lett*. **119**, (2021) 041903 doi:10.1063/5.0059767

Water can promote or inhibit crystal formation in ITC

Organoammonium	N_{pairs}	N_{dry+}	N_{wet+}	N_{++}	N_{+-}	N_{-+}	N_{--}	McNemar p
Dimethylammonium	192	77	81	71	10	26	85	0.0080
4-Methoxyphenylammonium	144	2	11	0	11	2	131	0.015
Acetamidinium	144	60	33	38	5	22	79	0.0010
iso-Propylammonium	144	3	10	3	7	0	134	0.0083

Similar trends observed in thin-film growth

Use SolTrain to grow thin-films…quantify grain length distribution by SEM



P Nega et al. Appl. Phys. Lett. 119, (2021) 041903 doi:10.1063/5.0059767

Demonstrating ML in the laboratory

Build interpolative surrogate models from limited initial reaction data. *Chem. Mater. (2020)* doi:10.1021/acs.chemmater.0c01153

Benchmark extrapolative models to predict reaction outcomes for novel reagents using physicochemical features *J. Phys. Chem C.* (2020) doi:10.1021/acs.jpcc.0c01726

Model-fusion methods to combine ML predictions and identify anomalies *J. Chem. Inf. Model.* (2021) doi:10.1021/acs.jcim.0c01307

Active learning experiment selection for phase boundary mapping & control *Chem. Mater.* (2022) doi:10.1021/acs.chemmater.1c03564

Benchmark meta-learning and active-meta learning experiment planning in the lab *J. Chem. Phys. (2022)* doi:10.1063/5.0076636

Combine computer vision with simulation to infer virtual experiments from time-series observations *Chem Mater.* (2022) doi:10.1021/acs.chemmater.2c00247

"Bakeoff competition" of experiment selection algorithms in the lab ChemRXiv (2022) doi:10.26434/chemrxiv-2022-l1wpf

Identifying crystal growth additives with iterative machine learning + feature selection. *Cryst Growth & Design (2022)* doi:10.1021/acs.cgd.2c00522

*** Complete data and code published for every study**

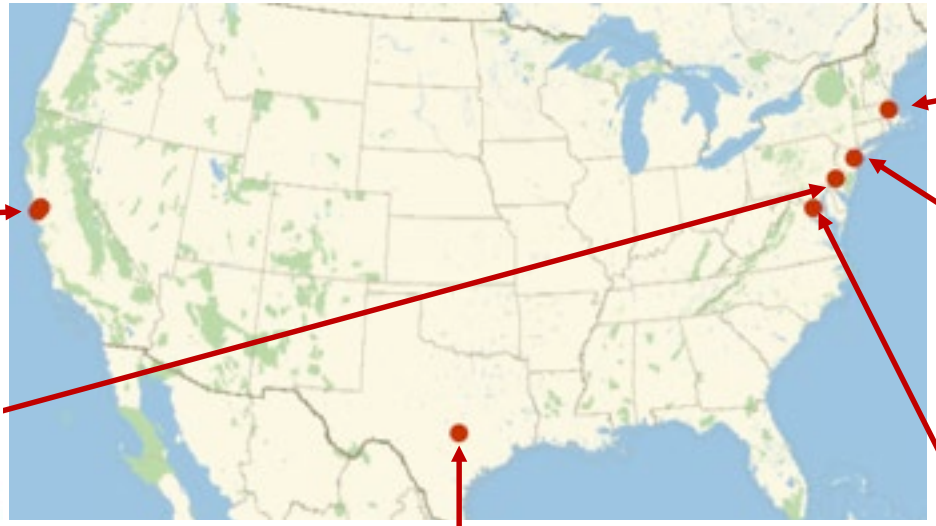
The Perovskite Challenge Competition

Invite external teams to submit algorithms to (remotely) play this game

Evaluate **11 Bayesian/active/meta-learning algorithms** in a standard process

(experiment plans/ML)

(automated experimentation)



BERKELEY LAB MOLECULAR FOUNDRY

HAVERFORD COLLEGE

FORDHAM UNIVERSITY
THE JESUIT UNIVERSITY OF NEW YORK

COLUMBIA UNIVERSITY

Lausanne

ATINARY
ENABLING SELF-DRIVING LABORATORIES

TACC

(compute infrastructure)

netrias

twosix LABS

V. Shekar *et al.* ChemRxiv (2022) doi:10.26434/chemrxiv-2022-l1wpf-v2

The Perovskite Challenge Competition

Repeated competition: 4 different organic reagents x two trials each

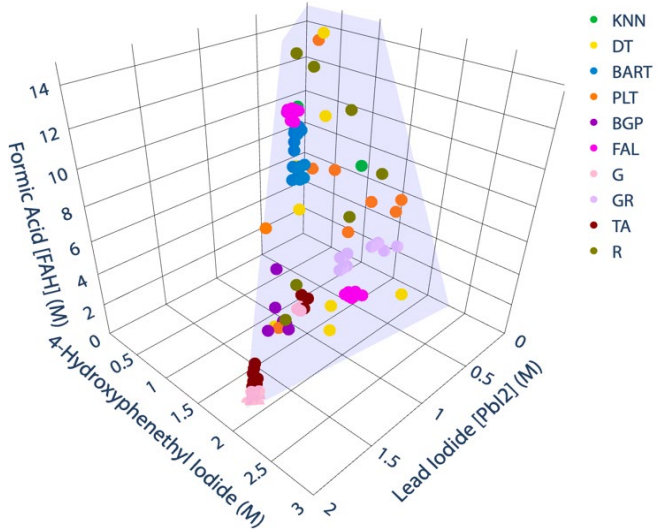
- Everyone starts with the same 10 randomly selected experiments
- Request 10 additional experiments
- Make 10 predictions of best outcome

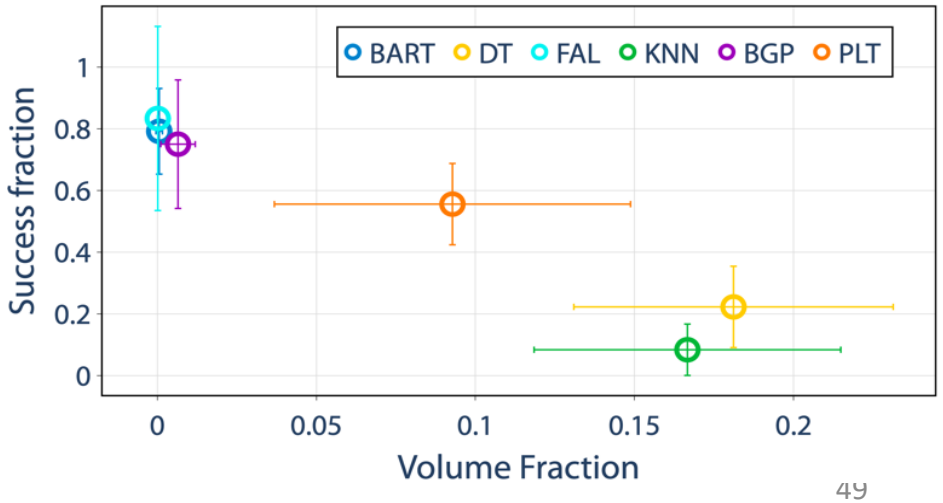
Best performance (most reliable across amines & cold-starts, highest success) by Gaussian Process-type methods & Bayesian Additive Regression Trees.

Side information about chemical properties & historical data is useful

Many active algorithms get trapped in local minima depending on the cold-start

All algorithms tend to “clump” in regions of initial success—propose some solutions to force recommendations away in a model agonstic way





V. Shekar *et al.* ChemRxiv (2022) doi:10.26434/chemrxiv-2022-l1wpf-v2

Development timeline

Software

v 1 (2018)—start observing experimental teams, built a pre-prototype to collect data as we go along

V 2 (2019-2020)—build a prototype to show functionality, test data model

v3 (2021-2022)—full-fledged SQL database backend, GUI & REST API

Experiments:

ITC(2017*-2019) automate the synthesis experiments

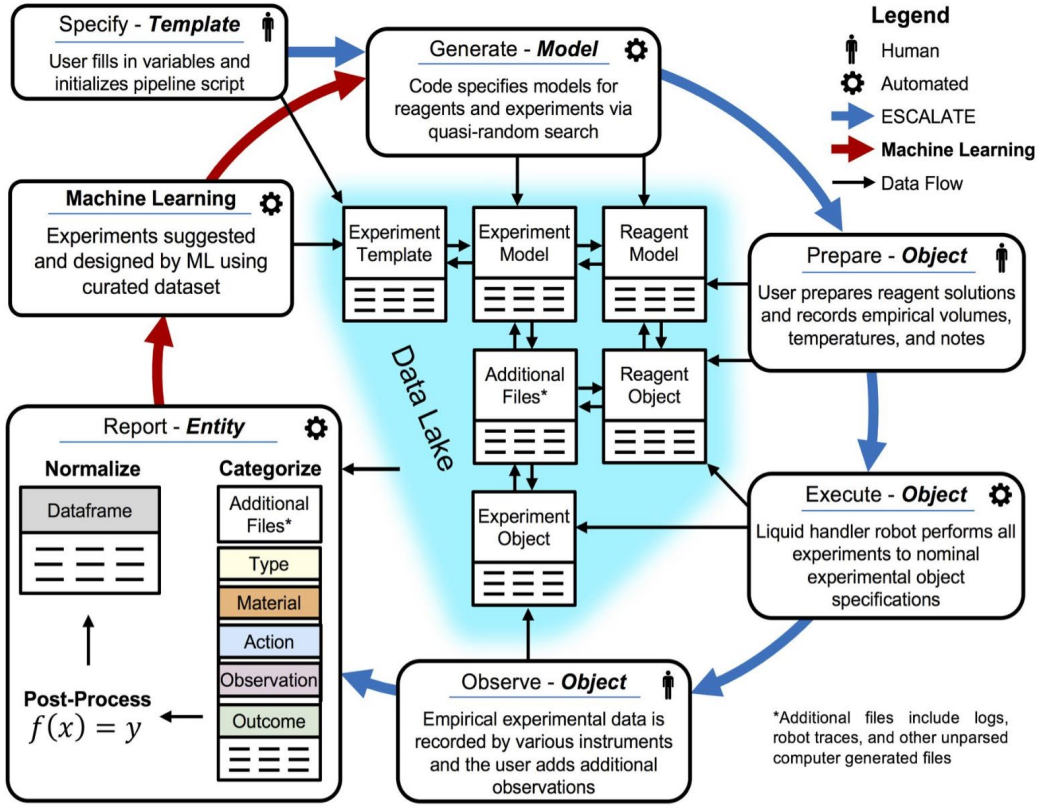
ASCV (2019-2020) automate another synthesis approach

Characterization complexity (2020-2022): more modalities and data types

Conclusion

- We've developed open-source software for dealing with hybrid automated+manual experimentation
- We've demonstrated distributed experimentation across different labs, with instructions coming from remote specifiers
- We've used this capability to benchmark ML methods in the lab, on a real physical problem (discovering new perovskites)
- P.S. We've discovered and characterized lots of new materials…

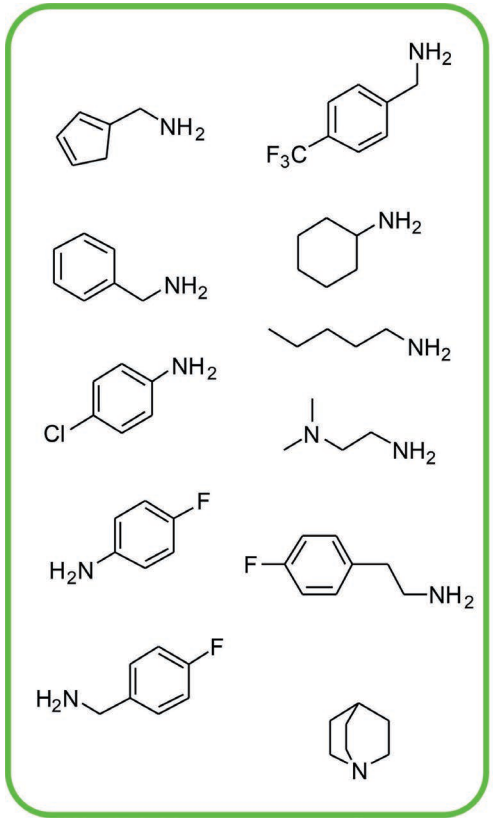
Questions? jschrier@fordham.edu

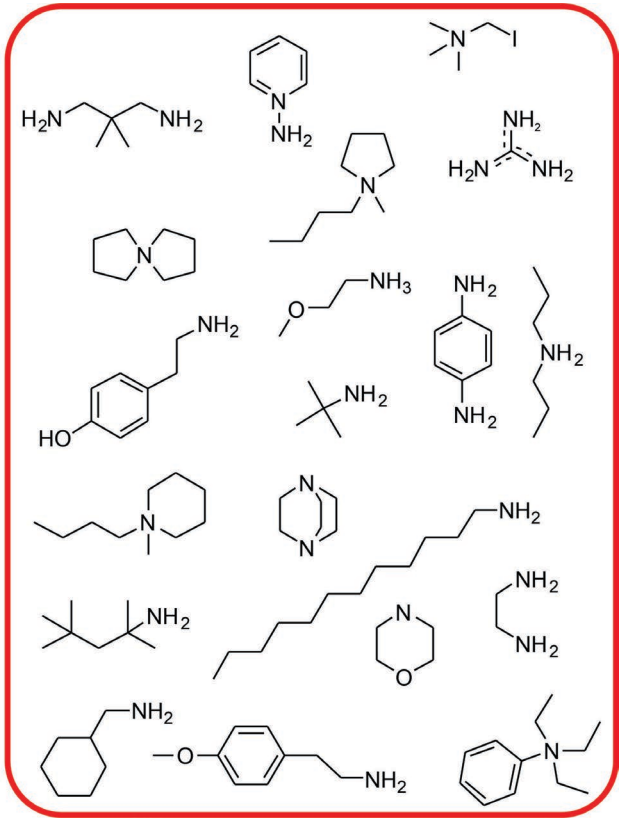


Current project: Copper Halides Perovskitoids

We've modified the RAPID2 antisolvent vapor diffusion approach to be able to grow copper halide compounds using the iodide salts

During Spring 2023, undergraduate Becca McAlper'23 has run a campaign spanning **31 organoammonium components**, and structural characterization of **11 new compounds***





* When I made this slide a month ago…

Conversation starters: Sustainability

It was an unusual opportunity to develop this software:

- Software engineer
- Plus a chemistry postdoc
- Plus some really excellent students

As an academic scientist, it's unlikely I'll do more than maintenance…

- Staffing turnover was a challenge—industry is calling…
- We can hire people without the skills (but then when they learn...)
- I'm a scientist, not a software engineer (level of interest)
- Typical NSF grants don't fund software engineers
- A part-time postdoc might keep it alive, but…there are other science/career goals

Conversation starters: Automation

Automation is great!

- Experiment plans and results are “born digital”
- Easier to capture results, easier to show value

Automation is over-rated!

- Prediction: We are entering a transitional decade, with increasing “islands of automation”
- Lots of valuable legacy equipment will remain in use…need a way to use it
- Early-stage experimentation is still “fiddly” and may not be suited to automation or cloud labs unless all unit operations are well supported. Sample transfer is a limiting step
- Make sure that data-enabled tools help in a non-automated world.

Conversation starters: Adoption

We must demonstrate value for rank-and-file experimental chemists:

- **Carrot:** Case studies showing increase rate of scientific discovery/productivity
- **Stick:** Requirements by funders/journals/user facilities

Let's not be afraid of 80% solutions:

- Technology activation barrier is a major challenge.
- What can a single user adopt that creates value? (as opposed to requiring an entire community of adopters)
- Lightweight overlays on top of existing tools (e.g., spreadsheets) might be enough
- Making onerous tasks easier might be a way to attract users

Conversation starters: Education

It is easier to teach a chemist SQL than to teach a programmer chemistry…

- We already teach lab notebook skills…
- Opinion: The best way to do this is to incorporate these skills into *existing* pedagogical lab experiences (throughout undergrad curriculum)
- Software Carpentry-type models
- Even small things can boost general awareness
- Maybe we just need lab scientists to care about this, so they can have intelligent conversations with software engineers and user facility scientists

Next Scientific focus areas

- **Rare earth element separation** (joint with Los Alamos NL)
  - Can we combine molecular simulations + automated experiments to develop new separations for f-elements (used in radioisotope medicine, critical energy materials)?

- **Liposomes** (joint with Central Conn. State Univ.)
  - Can we uncover physicochemical principles for the formation of small lipid structures related to the origins of life?
  - Can we develop new liposome formulations as a biotechnology platform (drug delivery, synthetic biology, etc.?)
  - **POSTDOCTORAL RESEARCHER POSITIONS AVAILABLE**—Let's chat!

Other things I like

- ESAMP / Materials Knowledge Graph (John Gregoire & Toyota Research Institute)—create a rich record of materials/action graphs, provenance, properties
- Cheminfo / Kadi4mat —smart data processing components for building ELNs with data analysis workflows
- XDL (Lee Cronin)—for chemical experiment description and reduction to practice
- PyIron — model of code-based workflow specification
- Aquarium (Jeff Klavins)—specifiying ways to tell human workers to do things in the lab and getting results back
- Cloud lab model (e.g., Strateos, Emerald Cloud Lab, Carnegie Mellon, increasingly National Lab User Facilities…)—as an abstraction language (with their own workflow management behind the scenes)