What Is Citrine?

venture- and grant-funded Si Valley startup w/revenue
our platform extracts insights from materials data
Who Is Citrine?

three founders, five full-time employees

backgrounds: DFT, GaN growth, Red Hat, Snapchat
Citrine’s Vision

unite all materials data & algorithms in one platform

provide *Moneyball* analytics for your research
Things Citrine Does Not Do

DFT or atomistic modeling
consulting
charge academics $
“Software is eating the world”
-Marc Andreessen, cofounder of Netscape & Si Valley VC

...so why not materials?
lack of structured data
Where We Are
1. The materials data landscape is highly fragmented
Materials Data Pipeline

Data generation

Laboratory equipment

Transport/store

USB stick, email, Dropbox, PC

Wrangling

Excel, PS, MATLAB, Python, R

Search/Aggreg.

Google, WoS, ICSD

Report Authoring

Word / LaTeX

Distribution

Elsevier, ArXiv, Researchgate

Bibliography Tools

Mendeley, EndNote, Zotero, Papers
Materials Data Landscape

Inception
- Institutions
  - Universities
  - National Labs
  - Industry
- Synthesis/mfg
  - Applied Materials
  - Oerlikon
  - Veeco
  - Homemade tools

Creation
- Characterization
  - Bruker
  - PANalytical
  - Sigma
  - Agilent
  - KLA-Tencor
- Testing
  - Newport
  - Weibull
  - Intertek
- Software
  - ICME
  - Accelrys
  - AFLLOW
  - ASE

Distribution
- Publishers
  - Elsevier
  - APS
  - ACS
  - Wiley
  - Springer
- Societies
  - MRS
  - TMS
  - ASM
  - APS
  - ACS

General Databases
- SpringerMaterials ($)
- NIMS
- MatWeb

Domain Databases
- ICSD ($)
- Powder Diffraction File ($)
- Nanohub
- ASM Phase Diagrams ($)
- Granta

Computational Databases
- Materials Project
- AFLLOWLIB
- Harvard Clean Energy Project
- NoMaD
- CatApp
2. Not only that, materials data are incredibly diverse.
materials have a wide variety of data formats and sources
materials have many degrees of freedom besides composition
some metadata examples (there are many, many more):
...crystalline

amorphous

grain structure

nanoparticles

nanotubes

nanowires

composites
all of these layers tell us about how materials behave, so...

any solution to the mishmash must include all of them
Times Are Changing
Harnessing the power of supercomputing and state-of-the-art electronic structure methods, the Materials Project provides open web-based access to computed information on known and predicted materials as well as powerful analysis tools to inspire and design novel materials.
February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
      Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research
Publisher Open Data Efforts
(rise of numerous smaller databases creates need for meta-aggregation)
Citrine’s Work
Machine Learning for Thermo

B. Meredig & C. Wolverton, PRB 89, 094104 (2014)
Auto-Discovery of Correlations

Citrination Platform

“all materials data”
semantic search, APIs
data extraction from pdf’s
single data standard
data-driven apps & models
<table>
<thead>
<tr>
<th>Material</th>
<th>Property</th>
<th>Value</th>
<th>Conditions</th>
<th>Reference</th>
</tr>
</thead>
</table>
Materials Data Standard

JSON-based definition of arbitrary materials objects & processes

Able to accommodate wide variety of materials data
Compound positions determined by weighted composition
(e.g., SiGe would be halfway between Si and Ge; Mg$_2$Si is 1/3 of the way from Mg to Si.)

Distant, novel class of thermoelectrics

Universe of known TE compounds

MW Gaultois, AO Oliynyk, A Mar, TD Sparks, GJ Mulholland, & B Meredig, “A Recommendation Engine for Suggesting Unexpected Thermoelectric Chemistries: Initial Experimental Validation.”
# CITRINE MATERIALS RECOMMENDATION ENGINE

## Thermoelectrics

<table>
<thead>
<tr>
<th>COMPOUND NAME</th>
<th>Fe(_2)MnSn</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fe(_2)MnSn</strong></td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COMPOUND</th>
<th>SEEBECK COEFFICIENT</th>
<th>RESISTIVITY</th>
<th>THERMAL CONDUCTIVITY</th>
<th>BAND GAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tl(_9)BiTe(_6)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>TaVO(_5)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>Bi(_2)Te(_3)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>Tl(_9)SbTe(_6)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>GdTaO(_4)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>TaPO(_5)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>Bi(_2)Te(_5)Pb(_2)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>CrPbO(_4)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>MoPbO(_4)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>CaWO(_4)</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>
TE Search: All Ternary Systems
General Materials Properties

![Graph showing the relationship between formation enthalpy examples in the training set and cross-validation mean absolute error (kJ/atom).]
Where We’re Going
# Trapped Non-Digital Data

## Energy band gaps in elemental and binary compound semiconductors and insulators

<table>
<thead>
<tr>
<th>Entry No.</th>
<th>Compound</th>
<th>E(g) (eV)</th>
<th>Temp. (Kelvin)</th>
<th>dE/dT (eV/Deg)</th>
<th>R</th>
<th>By</th>
<th>On</th>
<th>Tr</th>
<th>Effect</th>
<th>Color</th>
<th>Ref</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Li</td>
<td>1</td>
<td>13.6</td>
<td>300.0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>D</td>
<td>D</td>
<td></td>
<td>532</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Li</td>
<td>1</td>
<td>12.61</td>
<td>300.0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>E</td>
<td></td>
<td></td>
<td>532</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Li</td>
<td>1</td>
<td>12.6</td>
<td>.</td>
<td>2</td>
<td>9</td>
<td>3</td>
<td>E</td>
<td></td>
<td></td>
<td>384</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Li</td>
<td>1</td>
<td>12.1</td>
<td>77.0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>U</td>
<td></td>
<td></td>
<td>313</td>
<td>ABSORPTION EDGE.</td>
</tr>
<tr>
<td>5</td>
<td>Li</td>
<td>1</td>
<td>12</td>
<td>200.0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>U</td>
<td></td>
<td></td>
<td>313</td>
<td>ABSORPTION EDGE.</td>
</tr>
<tr>
<td>6</td>
<td>Li</td>
<td>1</td>
<td>11.7</td>
<td>300.0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>U</td>
<td></td>
<td></td>
<td>313</td>
<td>ABSORPTION EDGE.</td>
</tr>
<tr>
<td>7</td>
<td>Li</td>
<td>1</td>
<td>9.33</td>
<td>55.0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>U</td>
<td></td>
<td></td>
<td>55</td>
<td>G EDGE.</td>
</tr>
<tr>
<td>8</td>
<td>Li</td>
<td>1</td>
<td>7.95</td>
<td>80.0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>D</td>
<td>D</td>
<td></td>
<td>501</td>
<td>TRANSITION G15 -&gt; G1.</td>
</tr>
<tr>
<td>9</td>
<td>Li</td>
<td>1</td>
<td>7.5</td>
<td>55.0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>U</td>
<td></td>
<td></td>
<td>55</td>
<td>G EDGE.</td>
</tr>
<tr>
<td>10</td>
<td>Li</td>
<td>3</td>
<td>Sb</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>U</td>
<td></td>
<td></td>
<td>246</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Li</td>
<td>1</td>
<td>I</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>U</td>
<td></td>
<td></td>
<td>55</td>
<td>L EDGE.</td>
</tr>
</tbody>
</table>
User Data Upload

User Model Development

“Github for materials models”
Challenges

incentives
existing workflows
need for buy-in
Stakeholders

- universities
- government labs (DOE labs, NIST...) in the US, EU, Japan, China...
- funding agencies
- journal publishers
- scholarship search engines
- professional societies
- database providers
- equipment makers
- materials industry (Dow, DuPont, Alcoa, Corning…)
- industries that rely on matls (aerospace, electronics, energy...)

and YOU.
WE WANT YOU!
Ways to Get Involved

e-mail bryce@citrine.io to join mailing list.

Try citrination.com and give us feedback.

Contribute data.

Contribute models to platform (alpha).

Grant proposals – drive our dev.