An Integrated Collaborative Environment for Materials Research

Matthew Jacobsen
Materials & Manufacturing Directorate
Presentation Roadmap

• Introduce ICE
• Review integration case
• Present a vision for the future of ICE and like systems
Acknowledgements

• Dr. Charles Ward
• Bryon Foster
• The rest of the team
Materials and Manufacturing
Research Infrastructure

• 700+ scientists and engineers
• 108,000 sq ft lab space, 200 lab modules
• 750+ computers associated with research equipment
• 1000+ computers on desks: 2 separate networks
• 80+ scientific and engineering software packages
• Local computational clusters & remote HPC

And no supporting collaborative research environment
An Integrated Collaborative Environment

- The Materials Genome Initiative (MGI) calls for a Materials Innovation Infrastructure, in agreement with the goals of ICMSE
- ICE is a highly tailored, federated infrastructure built for the R&D community
- ICE represents a joint effort between software and materials engineers to deliver game-changing functionality
ICE-Enabled Capabilities

- The coordination and management of research activities
- The collection of research data (structured and unstructured)
- Complete traceability of material evolution
- Legacy data sources to continue to exist in many cases, but with connections to ICE API
- Growth of the RX ICMSE culture
The Federated Architecture allows for self-governance of connected systems. Systems may be COTS tools, in-house developed applications, or any hybrid thereof. Systems do not talk directly to each other - ICE “brokers” all transactions between connected systems.
Architectural Solution

- ICE Core - Collaboration platform (Hub), Common Service Bus and Apps (Django), advanced visualization (Plotly)
- ICE Extended - Material properties database (Granta), MTS Echo, Dream.3D
- Persistent identification, triple-based metadata, data type registration and SSO
- Graphical workflow design tools, item management, file management, advanced search tools
**Case 1: PID Stored Locally**

- **Step 1:** Create Record
- **Step 2:** API Call to Notify ICE
- **Step 3:** Metadata Registered
- **Step 4:** PID Linked to Local ID

**Case 2: PID Linked to Local ID**

- **Step 1:** Search Terms Entered
- **Step 2:** API Call for PID Search
- **Step 3:** Metadata, Local ID, and Location Registered
- **Step 4:** Endpoints Called to Return Data

**Case 3: Searching/Querying Data**

- **Step 1:** Search Terms Entered
- **Step 2:** API Call for PID Search
- **Step 3:** Endpoints Determined for PIDs with Metadata Matching Terms
- **Step 4:** Endpoints Called to Return Data

**Detailed Design & Behaviors**

- **Step 1:** File Upload
- **Step 2:** API Call for PID Issuance
- **Step 3:** Metadata and Location Registered
- **Step 4:** PID Issued/File Saved Locally
Data Creation via Workflow
## Data Retrieval via Search

### Search

<table>
<thead>
<tr>
<th>Object ID</th>
<th>Object Type</th>
<th>Date Created</th>
<th>Creator</th>
<th>Source</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>Casting</td>
<td>11/18/2015, 5:25:05 PM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>Casting</td>
<td>11/10/2015, 8:31:48 AM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>JRD</td>
<td>11/20/2015, 3:16:54 PM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>Sectioning</td>
<td>11/24/2015, 10:18:02 AM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Torsile Test</td>
<td>11/24/2015, 11:46:22 AM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>Torsile Test</td>
<td>12/4/2015, 11:49:32 AM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Casting Thermocouples</td>
<td>12/10/2015, 9:21:15 AM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>Casting</td>
<td>1/8/2016, 9:14:36 AM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>Torsile Test</td>
<td>1/8/2016, 9:14:38 AM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Casting Attributes</td>
<td>1/8/2016, 12:47:24 PM</td>
<td>Workflow</td>
<td>ICE forms</td>
<td></td>
</tr>
<tr>
<td>13202</td>
<td>GRANTA</td>
<td>Invalid Date</td>
<td>unknown</td>
<td>GRANTA</td>
<td></td>
</tr>
<tr>
<td>13201</td>
<td>GRANTA</td>
<td>Invalid Date</td>
<td>unknown</td>
<td>GRANTA</td>
<td></td>
</tr>
<tr>
<td>26678</td>
<td>GRANTA</td>
<td>Invalid Date</td>
<td>unknown</td>
<td>GRANTA</td>
<td></td>
</tr>
<tr>
<td>16221</td>
<td>GRANTA</td>
<td>Invalid Date</td>
<td>unknown</td>
<td>GRANTA</td>
<td></td>
</tr>
<tr>
<td>26493</td>
<td>GRANTA</td>
<td>Invalid Date</td>
<td>unknown</td>
<td>GRANTA</td>
<td></td>
</tr>
<tr>
<td>26694</td>
<td>GRANTA</td>
<td>Invalid Date</td>
<td>unknown</td>
<td>GRANTA</td>
<td></td>
</tr>
<tr>
<td>13249</td>
<td>GRANTA</td>
<td>Invalid Date</td>
<td>unknown</td>
<td>GRANTA</td>
<td></td>
</tr>
<tr>
<td>13250</td>
<td>GRANTA</td>
<td>Invalid Date</td>
<td>unknown</td>
<td>GRANTA</td>
<td></td>
</tr>
</tbody>
</table>
System Connection

• Test case – U of M’s Materials Commons
• Add Materials Commons API to ICE.Search
  – ICE delegates search mechanism to Materials Commons
  – Materials Commons relies on Elasticsearch (full text) vs object search (ICE.Search)
• Connection established after 4 hours of collaboration
  – RESTful call with authentication token and search string
  – JSON returned, shaped into search result format
Search Extended to Materials Commons

Primary Results

<table>
<thead>
<tr>
<th>Object ID</th>
<th>Object Type</th>
<th>Date Created</th>
<th>Creator</th>
<th>Source</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>529156e57-6d8b-4fbc-a858-0edd1a319e69</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:52:41 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>816a01c3-35a7-45a1-a503-3459b9b01e1</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:53:02 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>b93e700a-3575-41e1-ac51-678330362ea</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:53:34 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>6d40813-62b9-bc8b-62713465394</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:52:36 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>c6077468-4ab9-b21b-cc2670a1d8d2</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:53:40 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>cf3a33d4-1c00-4350-92ac-129b2053a54</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:53:06 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>816e66cc-b7f8-f8c0-b7c6-69445c4eb3ff</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:53:25 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>7b559d5c-e818-40f2-a298-7ca15db2ad8c</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:53:26 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>4df08721-b553-4b69-bca0-68bce1a9d9743</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:53:50 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>54a347ca-77a2-4e0b-b5f7-6221f7b2191</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:52:46 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
<tr>
<td>d6c78f4c-82d4-4b8e-b8c8-3b157998471</td>
<td>Materials Commons - files</td>
<td>4/25/2014, 1:52:54 PM</td>
<td><a href="mailto:miao@sasich.edu">miao@sasich.edu</a></td>
<td>Materials Commons</td>
<td></td>
</tr>
</tbody>
</table>
Object Instantiation

• Persistent Problem – how to treat workflow processes, participants, and items (physical and digital) as first class objects?
• Begin to register various data types – object “classes”
• Ex. Tension test, titanium specimen, etc.
• Invoke registered data types wherever possible
• Index all metadata assignments based on object type
New Functionality

• Data Model Builder – open up the DTR to certain users
• Graphical interface for defining data models and linkages/nesting
• DTR is implemented with OO principles of inheritance
• Use a NoSQL structure to define “parent” classes (casting) and child classes (investment casting)
• Restrict instantiation of new objects (even metadata) to those entries in the DTR.
Example 1 – Data Model Builder

Data Model

Data Model Name: Text
Select Domain(s): Option 1
Extend Existing Data Model: Option 1
Field Name
Type
Text
Option 1

Cancel Save Create & Design Form
Example 2 – Form Builder
An Improvement, but...

- Still not “semantic” – how do we relate our classes?
- We need a simple way (baby steps) to start building vocabularies, taxonomies, and domain-specific ontologies
- Our users are overwhelmed at the utterance of “ontology”
- Enter the Basic Formal Ontology
Basic Formal Ontology

- Created by Dr. Barry Smith and others circa 2000
- Establishes a high level framework for building out domain ontologies
- Successfully used in biomedicine, human genome project, Army, etc.
- Extended by “common core” ontologies, and further in domain specific ontologies
• Try to abstract objects from processes (test frame from the test for example) and use “occurrents” only as needed
• Most things can and should be described as continuants
• Separate objects from qualities/properties
Approach

• Whiteboard a concept
• Build a taxonomy
• Define relationships
• Construct domain ontology from taxonomy and relational elements
• Continuously refine the ontology
• Propagate into other domains
Example – Tension Test

- First stab – not perfect, but gives plenty of elements to start fitting into a taxonomy
- Key point – the SME must be involved and be comfortable with the flow
Taxonomy and Relationships

- Materials
  - Metals
    - Stainless Steel
  - Non-Metals
  - .....
- Quality
  - Porosity
  - Density
  - Transmittance
  - .....
- Relationships
  - Participates in
  - Contains

- Systems like Granta do this pretty well already
- Downside is that the qualities are dependent

- Object instances pull from all tiers:
  - Ex: Sample of Stainless Steel has qualities X, Y, Z, and was part of Test A
  - Qualities are only invoked in the instance, not the class
Value Proposition

- System integration is greatly enhanced by using common schema/vocabulary/ontology
- Eases total ecosystem burden with standard models/classes
- Existing schema/ontology momentum in many S&T communities
Next steps

• Engage SMEs and flesh out the mechanical test domain
• Build into BFO domain ontology in Protégé
• Flatten out the taxonomy and ontology
• Build an inferencing engine for determining identities based solely on qualities, similar to a graph-based templating search
• Implement common domain elements in partnering systems

• We need to collaborate!!