

EIC Coll Tools - Intro

6/26/2024

- News about the project
- Follow up from previous meetings

News about the project

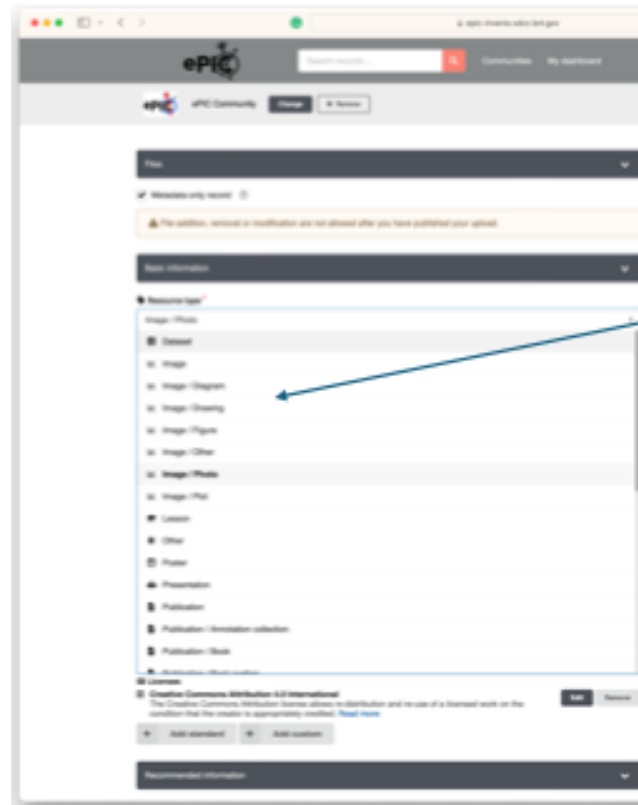
- Project status and plans for FY25 were submitted as part of the yearly Program Development (PD) reporting.
- For FY25-26, the project leads will be Peter and Ofer (co-PIs)
- Very valuable meeting with ePIC spokesperson John LaJoie on 6/21: <https://indico.bnl.gov/event/23988/>. Peter will summarize the meeting today.

From Previous meetings

- Uma's Notebook (Llama2 example) attached to previous meeting (6/12) indico : <https://indico.bnl.gov/event/23897/>
- Simplification of InvenioRDM upload page
- Inputs for various categories of documents and roles

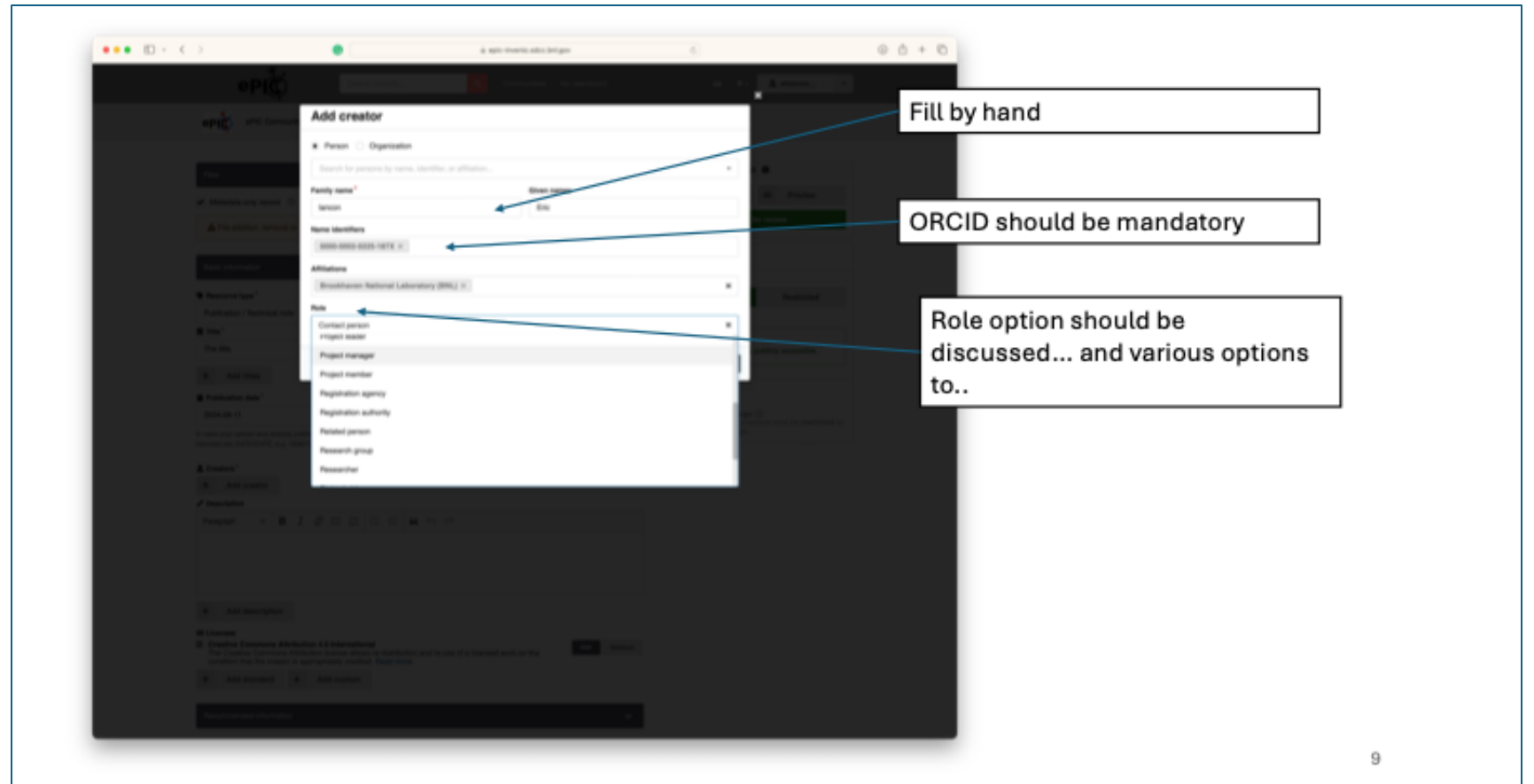
Categories of documents

Basic information about record



Too many options
Should agree on a reduced list

Uploader roles



The screenshot shows the 'Add creator' form in the ePIC system. The form includes fields for 'Family name', 'Given name', 'New identifiers', 'Affiliations', and 'Role'. Three callout boxes point to specific elements:

- Fill by hand**: Points to the 'Given name' field.
- ORCID should be mandatory**: Points to the 'New identifiers' field.
- Role option should be discussed... and various options to..**: Points to the 'Role' dropdown menu, which lists options such as 'Contact person', 'Project leader', 'Project manager', 'Project member', 'Registration agency', 'Registration authority', 'Funded person', 'Research group', and 'Researcher'.

Llama3 command line

>>> do you know Glance the software from CERN?
You're referring to Glance, a fascinating project from CERN!



>>> what is the IRI
The IRI!



IRI stands for International Research Institute (formerly known as the Information Retrieval Institute). It's a research institute based at CERN, the European Organization for Nuclear Research.

Integrated Research Infrastructure Architecture Blueprint Activity

FINAL REPORT 2023



Ingest IRI report

```
[10]: #local_path = "WEF_The_Global_Cooperation_Barometer_2024.pdf"
local_path = "IRI-ABA-report.pdf"

# Local PDF file uploads
if local_path:
    loader = UnstructuredPDFLoader(file_path=local_path)
    data = loader.load()
else:
    print("Upload a PDF file")
```

```
[11]: # Preview first page
data[0].page_content
```

```
[11]: 'THE DOE OFFICE OF SCIENCE\n\nIntegrated Research Infrastructure Architecture Blueprint Activity\n\nFINAL REPORT 2023\n\nTHE DOE OFFICE OF SCIENCE\n\nIntegrated Research Infrastructure Architecture Blueprint Activity\n\nFINAL REPORT 2023\n\nCover Image: Using data transported via ESnet and analyzed in near-real time at NERSC from experiments performed\n\nat SLAC, this image represents small-molecule serial femtosecond crystallography, where a free-electron X-ray laser\n\ncaptures the structure of a nanocrystal just before the sample is destroyed. This collaboration is an example of a\n\nsuperfacility implementation that is revolutionizing scientific research.\n\nImage Credit: J. Nathan Hohman. Illustration: Ella Maru Studio.\n\nCreative: Samantha Trieu, Creative Services, Lawrence Berkeley National Laboratory\n\nEditorial: Carol Pott, Computing Sciences Area, Lawrence Berkeley National Laboratory\n\nDOI: 10.2172/1984466\n\n23-CSA0-9059\n\nTable of Contents\n\nExecutive Summary\n\nThe DOE Office of Science IRI Vision and Framework\n\n1 3\n\n1. Goals and Philosophy of the ABA\n\nGoals and Objectives\n\nFoundational Precursor Activities\n\nParticipants, Organization, and Leadership\n\nPhilosophy of Conduct of the ABA\n\n4 4 5 6\n\n2. Structured Steps Towards the IRI\n\nCulminating Insights\n\nTimeline and Approach\n\nGathering Insight\n\nPhase Design\n\nPhase\n\n7 7 7 10\n\n3. Culminating Insights and Artifacts\n\nIRI Patterns and Their Architecture Blueprints\n\nTime-Sensitive Patterns\n\nData Integration-Intensive Patterns\n\nLong-Term Campaign Patterns\n\nIRI Practice Areas\n\nArtifacts\n\nResource\n\nCo-Operations\n\nCybersecurity and Federated Access\n\nUser Experience\n\nWorkflows, Interfaces and Automation\n\nScientific Data Lifecycle\n\nOverarching IRI Principles\n\nGovernance and Steering Considerations\n\nPattern Blueprint Compare and Contrast\n\n11 11 11 12 13 14 14 14 14 15 15 15\n\nConclusions\n\n15\n\nReferences\n\nSummary Presentations\n\n15 16\n\nAppendix A: Leadership and Participants\n\nRoster of Leaders and Participants\n\nLeadership Group\n\nParticipants\n\nHeadquarters\n\n17 18 18 19\n\nAppendix B: IRI ABA Participant Kickoff Webinar Presentation\n\n22\n\nAppendix C: Synthesis of Sprint 1 insights\n\n22
```

```
[12]: ollama pull nomic-embed-text
```

```
[25]: #chain.invoke("What are the 5 pillars of global cooperation?")
chain.invoke("what is the IRI")
```

```
OllamaEmbeddings: 100%|██████████| 1/1 [00:00<00:00, 1.07it/s]
OllamaEmbeddings: 100%|██████████| 1/1 [00:00<00:00, 33.15it/s]
OllamaEmbeddings: 100%|██████████| 1/1 [00:00<00:00, 30.49it/s]
OllamaEmbeddings: 100%|██████████| 1/1 [00:00<00:00, 31.11it/s]
OllamaEmbeddings: 100%|██████████| 1/1 [00:00<00:00, 25.37it/s]
```

```
[25]: " The Integrative Research and Innovation (IRI) is a vision and framework by the Department of Energy (DOE) Office of Science, aiming to empower researchers to seamlessly integrate DOE's world-class research tools, infrastructure, and user facilities for accelerating discovery and innovation. It seeks to create a robust integrative scientific ecosystem for DOE researchers and the broader scientific community through the use of Science Patterns and Practice Areas. The key organizing elements of the IRI Framework are Science Patterns and Practice Areas, which address areas such as user experience, resource collaborations, cybersecurity, workflows, interfaces, automation, data life cycles, portable/scalable solutions, and more."
```

Same question about IRI

Answer is OK



Acronyms for the workflow management system

TRACE <ul style="list-style-type: none">•Tracking•Resource•Analysis•Collaboration•Engine	VISIO: <ul style="list-style-type: none">•Versatile•Integrated•System for•Information and•Organization	STREAM: <ul style="list-style-type: none">•Scientific•Tool for•Research and•Experiment•Access and•Management	MIRROR: <ul style="list-style-type: none">•Management•Interface for•Research•Resource•Organization and•Reporting	SPARK: <ul style="list-style-type: none">•System for•Project•Analysis,•Research, and•Knowledge	FLARE: <ul style="list-style-type: none">•Framework for•Logistical•Analysis and•Research•Environment	SCRIBE: <ul style="list-style-type: none">•Scientific•Collaboration•Resource•Integration•Based•Environment
---	---	--	--	---	---	--

Suggestions please!