

# **ePIC collaboration DB discussion**

**Peter Steinberg, BNL, 16 Nov 2023, ePIC discussion**

# Where are we on this process?

- **ePIC needs a collaboration database**
  - I propose calling it the eCDB for now
- **My understanding is that this is a sub-discussion of the more general issue of ePIC collaboration tools**
  - We are not trying to solve the general problem of the lack of these tools at BNL & Lab
  - I understand that BNL has funding to explore different options, e.g. a local instance of Glance
  - However, this has an unknown timescale, and I believe it could utilize a more limited ePIC “phonebook”
- **Thanks to everyone who has contributed to the requirements document**
  - I will summarize the main points here, and highlight a few cases where a solution may not be trivial

# What are the goals for today?

- **ePIC is an active collaboration with 170+ institutions but no-one quite knows how to keep track**
- **We need a functioning database of**
  - Institutions
  - Individual members
  - CC membership
- **We also have an immediate need to have the means to keep track of attributes pertinent to the ePIC membership committee (eMC), e.g.**
  - Good standing
  - Committed FTE
- **One can also envision an urgent need to keep track of talks, etc at conferences**
- **So while we may not be able to satisfy all of the requirements, we should be trying to do so in the medium term**
  - Improve existing phonebook
  - Provide input into development of a more integrated “glance like” solution

# Summary of requirements doc

## ePIC Executive Summary

The ePIC Collaboration needs a **Collaboration Database** to carry out its mission. Broadly speaking, ePIC is looking for the core functionality that exists in applications already maintained by the BNL SDCC in the form of the “phone book” Web applications used by EICUG, STAR, and sPHENIX. The Collaboration also needs additional/revised functionality as detailed below.

The most basic functionality of the Collaboration Database is to maintain contact information about the Collaboration members, institutions, and member affiliation with institutions. A member of the ePIC Collaboration can be *affiliated with more than one institution* and this affiliation may evolve over time.

For the most part, data entry and modification will be the responsibility of the ePIC Collaboration Council via its designated members.

# eCDB requirements

- **Extensibility of attributes (to any object, institute or member)**
- **Historical information**
- **Protection of PII**
- **CC member(s) per institute**
- **Command line interface (CLI) to allow more complex queries, scripting, etc.**
- **Data import/export**

# Extensibility of attributes

- **For simplicity, consider “institutes” and “members” as the primary “objects” in the eCDB**
  - Each of these could have different attributes that are needed (or not) at different times in ePIC’s development
  - Attributes should be easily added or removed at any time, without the need to modify any eCDB schemas
  - They should also keep a time history, to make it possible to reconstruct a time series (e.g. institutes joining, leaving and joining again, or a series of an individual’s roles in ePIC)
- **Examples of attributes**
  - working group membership
  - special status (early career, etc.)
  - ePIC author
  - institutes and/or members in good standing
  - “foreign keys” to refer to external DBs
    - *e.g. ORCID’s as each person’s unique identifier*

# History of attributes

- **As mentioned before, every attribute should have a “start time” and potentially “end time” associated with it**
  - This allows the DB to display institutional history, ePIC roles, ePIC talks, etc.
- **Attributes should also carry information on who created the attribute and when**
- **Examples in requirements doc**
  - For a member, one can trace institute affiliation over time
  - Authorship and membership qualifications

# Protection of PII

- **Some information is fundamentally private, but can be useful to ePIC, labs, DOE, etc. in an anonymized form**
  - e.g. employment status, gender, etc.
- **Individuals should ideally be able to control who can access certain information, and hide it even from supervisors, DB admins, etc.**
- **Thus, privacy/access should be definable on a per-attribute basis**
- **Will also require some thought as to how to define various roles (CC member, “community manager”, DB admin) and how to assign different access “levels” to them**
  - Need secure authentication/authorization system that enables defining these roles.



# Collaboration council

- **eCDB needs to be able to assign multiple CC members to a single institution**
  - This seems straightforward is “CC member” is an attribute of an individual

# Command line interface

- **With many institutions and members, and many different types of attributes, one cannot always expect a web interface to provide access to all information**
  - Requires possibility of a command-line interface (CLI) to both read/write records to DB, and read/write their attributes
- **Examples in document**
  - List of members per institute
  - History of institutes for a single member
  - A valid author list at a particular or for a particular document
  - Queries not yet on web UI
  - Service tasks and members assigned to them, for use by eMC
    - *Not clear to me if this needs to be in this DB, or in a separate one*

# Data import/export

- **Data import/export functions are needed to allow ePIC to manage the contents of the DB**
  - management, administrators, committee chairs, CC members,...
- **Import functions should include**
  - Providing a list of people (e.g. ORCIDIDs) for inclusion into ePIC, and setting some basic attributes
- **Export functions should include**
  - Providing data in a digestible form (CSV, JSON, etc.)
  - Entire export of the DB for migration to a new system