

eduCourse revisited and extended - psuCourse, psuCourseOffering, psuPlaces - a Description, a link to slides, and some thoughts...

What this is:

We've done some preliminary test of concept work here at Penn State to build out eduCourse: in our case, as a schema called psuCourse and a sub-class, psuCourseOffering.

These schema permit representation in LDAP of course offerings and information related to those offerings for purposes of lookup. The entries can be granular to the section level.

We establish that there "is such a thing as English 101" in psuCourse, and then instantiate it as an offering in psuCourseOffering. The offering carries such attributes as semester, campus, section, etc.

We also have introduced psuPlaces into the picture, a schema for the representation of locations, granular down to the room level. The schema includes an attribute for the GPS location of the building in which a room resides (and could, with the inclusion of a single additional attribute, take GPS location to the room level).

These schema tie together with Person entries from our Central Person Registry (CPR), OrgUnit entries from our eduORG directory, as well as interacting with each other.

Taken to the extent permitted by such schema, virtually all associations between persons, courses, and locations, and orgUnits, could be represented as DNs pointing to the appropriate entries in the appropriate trees. Please consider the thoughts behind this effort, outlined below.

My thanks to Jim Vuccolo and to Derek Morr for generously giving of their time as we gave formal expression to these ideas. Without their assistance, this document would be merely the ephemeral ramblings of one putatively versed in information representation, without regard to the constraints confronting any pure idea through encounter with reality.

The PowerPoint slide deck may be found at <https://iam.psu.edu/sites/default/files/psucourse.ppt>

About the thinking behind this:

This approach has emerged in our thinking as just one possibility means to "de-silo-ize" the bewilderingly difficult, entirely intractable information landscape that has grown up over decades of accretion at our university. The reader should regard this work, I suppose, as theoretical, that is, theoretically possible, even if not immediately (or perhaps ever) practicable. Even those among our Penn State Identity and Access Management Technical Architects Group who are intrigued enough by this approach to investigate it don't believe this is something we can simply "up and do" right out of the gate. That said, the obstacles to adoption of these ideas probably have more to do with the world into which they would be introduced than to their merit in abstract.

We are in the midst of building out a Central Person Registry - it is designed to structure and store information relating to persons and their identity. We have come to see the potential desirability of a *limited number of purpose-designed registries*, whose entries are related to one another, but which do not duplicate one another: in effect, a total deconstruction of the massively duplicated information currently spread across our information landscape. In place of the vast field of information silos we're now confronted with, we can envision "*a neighborhood of registries*", with each registry holding cardinal entries for a "type" of information.

In many respects, this represents a massive *simplification* in the representation of core information. Of course, such simplification at the center probably comes at costs elsewhere. At the present time, though, we are paying those costs already; our present landscape appears to achieve simplicity only to the extent we can muster by storing needed, but redundant, data at the local level, without consideration to the complexity accrued through maintenance, modification, even analysis of the "as-built" overall.

In truth, the degree of complication we're already living with is unsustainable. The overwhelming complexity is felt now both the center and at the localities.

Toward the goal of that simplification, the schema discussed here are envisioned primarily as targets for lookup from the perspective of local needs. The data in these directory structures are intended to be "authoritative enough" to permit routine business decisions and processes to rely upon them. The registry-based cardinal entries upon which these data views themselves would depend would be created and/or updated only by agents that have been given authority to perform such updates under a regimen of controls.

Such an architecture is, in many respects, a natural progression from the architecture required to reach the level of control we're trying to achieve with person identity information.

To sum up, in this investigation we're applying, by analogy, a kind of fractal approach, iterating a pattern and growing out the coherent structure of the CPR to see what would happen if such coherency were permitted to spread.

Indeed, failure to bring similar levels of control, and similar simplicities of approach, to ancillary data related to persons and their complex interactions with our organizational units, programmatic offerings, business services, and multiple locational contexts might doom the control we're hoping to bring to identity with the CPR.

In the present landscape even the Central Person Registry, however elegantly conceived and designed, may not be the sufficiently irresistible force we would hope to wield against the immovable object of our present inertia: such stasis born as it is out of legacy thrown upon legacy, kluge duct taped to prior jury rig, reactionary crisis avoidance heaped upon emergency intervention, along with a general disinclination to veer too far from an established trajectory.

Thanks for your time and consideration,

--Michael Pelikan