



Comparison of Digital Vita and Vivo Data Models

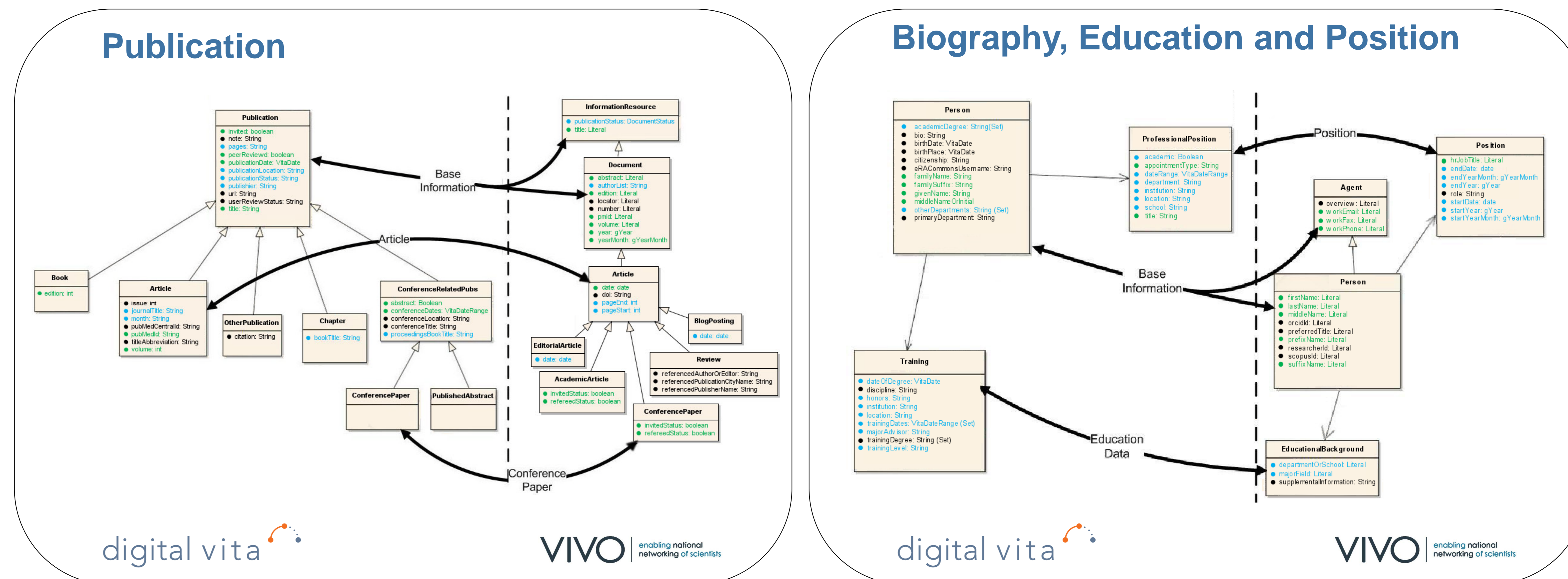
Titus Schleyer, DMD, PhD¹, William Shirey, MS², Charles Borromeo, BS² (¹Center for Dental Informatics; ²Department of Biomedical Informatics}, University of Pittsburgh, Pittsburgh, PA)



BACKGROUND

- Development and implementation of research networking systems a strong trend in biomedicine
- Future landscape most likely to consist of a variety of systems of different provenance
- Interoperability at the data layer important first consideration
- NCCR-funded VIVOWeb project primarily intended as a data platform
- Digital Vita, a CV-based research networking application developed at the University of Pittsburgh, as potential VIVOWeb end user application

COMPARISON OF SELECTED CLASSES AND ATTRIBUTES



Legend: ● Exact Match ● Similar ● No Match

CONCLUSION

- One to one data mapping between parts of VIVO and Digital Vita possible
- Overlap in person's bibliographic, biographic, education and activity information
- Preferred approach: standardized service calls that encapsulate data structure

OBJECTIVE

- Perform an initial comparison of the respective data architectures and models to assess feasibility of implementing Digital Vita as an application on top of VIVOWeb

METHODS

- Analysis of Digital Vita 2.0 and VIVO 1.0 models
- Class by class comparison of the two models
- Comparison at the attribute level of intersecting classes

Results

- Digital Vita and VIVO domain models intersect on the following information:
 - Bibliographic
 - Education
 - Collaboration
 - Biographic
 - Activity
- The models diverge with regard to:
 - Geographic location [VIVO]
 - Organization [VIVO]
 - Colleague status [DV]

- VIVO's domain model based on several standard ontologies (e.g. FOAF, Dublin Core and Bibontology), Digital Vita's on a custom information model
- Certain overlap at the field level, less so at the child/parent class levels
- Data fields, types and structure to be reconciled during mapping
- Mapping complicated by internal and external evolution of domain models

Contact:

Titus Schleyer, Center for Dental Informatics, University of Pittsburgh, titus@pitt.edu, ph. 412-648-8885

Abstract

Digital Vita (DV) is a research networking system developed by the Clinical and Translational Science Institute of the University of Pittsburgh. It integrates CV management functions with academic social networking and basic research team management and collaboration functions. DV allows researchers to manage their complete academic CVs, output CVs in a variety of formats, such as online profiles and NIH biosketches, build their social network through collaborations on publications and grant announcements, as well as explicit "colleague requests" to "attend requests" in Facebook, group their colleagues into research teams of their choosing and request biosketches, build biosketchs, requests, and maintain and selectively update multiple versions of NIH biosketches.

VIVO is a national research networking platform funded by the National Center for Research Resources designed to enable the discovery of research and scholarship across disciplines and institutions. An open source semantic web application originally developed and implemented at Cornell University, it contains researcher interests, activities, and accomplishments. VIVO supports browsing and a search function which returns faceted results for rapid access of desired information. Content in and from VIVO institutions may be maintained manually, through the VIVO in-browser web interface, or from systems of record, such as human resources, grants, courses, and faculty activity databases, or from data base providers such as publication aggregators and funding agencies.

In order to begin integrating DV with the VIVO platform, we performed a comparative analysis of the data models of the two systems. Digital Vita's data model is based on the standard format for the full academic CV of the University of Pittsburgh. Its information categories include biographical data, education and training, academic appointments, grants and contracts, awards and honors, publications, presentations, and service activities. It also comprises other information, such as personal statements (for NIH biosketches) and search team lists (for searching and the online profile). Colleague names such as publications and presentations are derived from a base class which allows the system to track collaborations associated with these items. The association allows DV to link collaborators and their records automatically.

VIVO uses an ontology to represent researchers and their information. Ontologies such as Friend of a Friend (FOAF) and Bibontology provide the foundation for the VIVO ontology. VIVO data models (classes) derived from these base ontologies to represent people, documents, events and organizations. An example in the sub-classification of a FOAF person which includes classes for emeritus professor, faculty member, student and others. The VIVO ontology also adds several top level concepts, such as categories to model biographic information and academic activities, for instance presentations, teaching and educational background, among many others.

In comparing their respective data models, we noted several areas of congruence and divergence between DV and VIVO. Both systems model biographical, bibliographic and educational information related to academics, and allow for the association of people as collaborators on items such as publications and presentations. The VIVO ontology models geographic location, organization and generic event related information which DV does not include. DV includes support for people to request and track "colleague" status with others and while VIVO appears not to. The similarities listed here will be crucial to successful system integration when data is shared between the two systems. Through key data elements are present in both systems to make a likely data system integration at the data level will not pose significant problems.

The results of this project are important for application developers, policy decision makers and others. Application developers for the two platforms must compare the information models of their applications to that of VIVO in order to determine the degree of "data compatibility" between the systems. Policy decision makers must understand which type of data the respective systems manage in order to determine which applications or combinations thereof can best fulfill their needs.

References

- Schleyer T, Spatnik H, Butler BS, Subramanian S, Wren D, Proffers ML, Karamallakis P, Maffei C. Feedback for scientific management and services for optimizing low scientific collaboration an established Journal of medical literature research. 2008; 10(1):24.
- Govis V. Networking in VIVO. Nature. 2009; 462(7):123.

Acknowledgment: Supported in part by NCCR grant 1CL18R031413-01.

Please take one: