Proposal for the process of exchanging variant data

The system and its features

Sharing of data about variation and the associated phenotypes is a critical need, yet variant information can be arbitrarily complex, making a single standard vocabulary elusive and re-formatting difficult. Complex standards have proven too time-consuming to implement.

The GEN2PHEN project has addressed these difficulties by developing a comprehensive data model for capturing biomedical observations, Observ-OM, and building the VarioML format around it. VarioML pairs a simplified open specification for describing variants, with a toolkit for adapting the specification into one's own research workflow. Straightforward variant data can be captured, federated, and exchanged with no overhead; more complex data can be described, without loss of compatibility. The open specification enables push-button submission to gene variant databases (LSDBs) e.g., the Leiden Open Variation Database, using the Cafe Variome data publishing service, while VarioML bidirectionally transforms data between XML and web-application code formats, opening up new possibilities for open source web applications building on shared data. A Java implementation toolkit makes VarioML easily integrated into biomedical applications. VarioML is designed primarily for LSDB data submission and transfer scenarios, but can also be used as a standard variation data format for JSON and XML document databases and user interface components.

The extent to which the system is currently being utilised

VarioML is the native data format for the Café Variome system for variant data discovery (http://www.cafevariome.org/). A plug-in has been developed for the import of data in VarioML format into LOVD v.2.0.

The HVP Standards and Guidelines with which the system is compliant

The VarioML format is compliant with the HGVS variant nomenclature system and is fully extensible to accommodate any future enhancements to that nomenclature.

Access and license terms and conditions under which the system is made available

There are no restrictions placed on the use of VarioML.

Summary

VarioML (Byrne *et al.*, 2012) is a set of tools and practices improving the availability, quality, and comprehensibility of human variation information. It enables researchers, diagnostic laboratories, and clinics to share that information with ease, clarity, and without ambiguity.

Byrne M, Fokkema IFAC, Lancaster O, Adamusiak T, Ahonen-Bishopp A, Atlan D, Béroud C, Cornell M, Dalgleish R, Devereau A, Patrinos GP, Swertz MA, Taschner PEM, Thorisson GA, Vihinen M, Brookes AJ, Muilu J (2012) VarioML framework for comprehensive variation data representation and exchange. BMC Bioinformatics, 13: 254

Proposed by Raymond Dalgleish, March 2013