SKELETOME: An eResearch Platform for the Skeletal Dysplasia Domain

Tudor Groza1, Craig McNamara2, Andreas Zankl3, Jane Hunter4
1The University of Queensland, St. Lucia, Australia, tudor.groza@uq.edu.au
2The University of Queensland, St. Lucia, Australia, c.mcnamara1@uq.edu.au
3Genetic Health Queensland, Royal Brisbane and Women’s Hospital, Herston, Australia, a.zankl@uq.edu.au
4The University of Queensland, St. Lucia, Australia, j.hunter@uq.edu.au

SKELETAL DYSPHASIAS

Skeletal dysplasias are a heterogeneous group of genetic disorders affecting human skeletal development. Currently, there are over 440 recognized types, categorized into 40 groups. Patients with skeletal dysplasias have complex medical issues, such as short stature, degenerative joint disease, scoliosis or neurological complications. Since most skeletal dysplasias are very rare (<1:10,000 births), data on clinical presentation, natural history and best management is sparse. The lack of data makes existing patient cases a precious resource for biomedical research because they enable scientists to study, among other things, the effects of single genes on human bone and cartilage development and function. The resulting insights may lead to a better understanding of the pathogenesis of more common connective tissue disorders, such as arthritis or osteoporosis.

The International Skeletal Dysplasia Society (ISDS) [1] has attempted to address some of these problems with its Nosology of Genetic Skeletal Disorders [2]. Since 1972, the ISDS Nosology lists all recognised skeletal dysplasias and tries to group them by common clinical-radiographic characteristics and/or molecular disease mechanisms. The ISDS Nosology is revised every 4 years by an expert committee and the updated version is published in a medical journal, being widely accepted as the official nomenclature for skeletal dysplasias within the biomedical community. While the content is invaluable, the format of the Nosology has several short-comings, including: (i) an inflexible classification scheme – each disorder being listed in one group based either on its clinical radiographic appearance or on its underlying molecular genetic mechanism; (ii) limited amount of cross-referenced information – each entry contains only the Online Mendelian Inheritance in Man (OMIM) number [3], the chromosome locus and the gene name, without being linked to widely used semantic data repositories, like the Gene Ontology [4], which would allow users to study further up-to-date relevant information; and most importantly, (iii) the lack of a shorter publishing cycle – the content quickly becomes out-dated, as genes or disorders discovered after the publication date can no longer be included until the next revision (4 years later).

THE SKELETOME PROJECT

Since genetic disorders are typically quite rare, a global network of patients, clinicians and researchers is necessary to accumulate the critical mass of data and knowledge needed to address some of the greatest challenges in medical genetics, i.e., the development of evidence-based clinical management guidelines, the study of genotype-phenotype correlations and the identification of disease modifier genes. Skeletal dysplasias are an ideal topic for a global medical collaboration network as the number of medical conditions is relatively small and well defined and there is an existing, tightly-knit and motivated community of clinicians and scientists willing to contribute, share and exchange case studies, data, diagnoses and clinical information.

Recognition of this opportunity, led to the establishment of the SKELETOME project – a collaboration between information scientists, Semantic Web researchers and clinical geneticists, led by the University of Queensland. In addition to a Web-based platform [5] for enabling and encouraging the international skeletal dysplasia community (researchers, experts, clinicians) to contribute content, the most important objectives for the project (which emerged from direct discussions with the community) are the following:

- **Common terminology.** The diagnosis and management of skeletal dysplasias depends on highly specialised domain knowledge across a number of disciplines (radiography, genetics, orthopaedics), which is not easily comprehensible to individual communities or hospitals. In order to enable the exchange of knowledge between experts (across languages and disciplines), a common terminology is required, hence leading to a shared conceptualisation of the domain.

- **Data integration.** Large datasets containing rich information on molecules (genes, proteins) already exist and the information relevant to skeletal dysplasias needs to be extracted and cross-referenced with the clinical data and knowledge produced by SKELETOME. The data cross-reference requires integration both at conceptual level, as well as, at actual data and instance level.

- **Knowledge transfer and sustainable knowledge evolution.** The knowledge collectively acquired represents a valuable asset from the conceptual perspective of the domain (materialized in the ISDS Nosology). Consequently, a seamless transfer of this knowledge is required to enable the dynamic and continuous evolution of the conceptual domain.
• Capturing provenance and expertise. The contributed content may take several forms, ranging from personal observations to scientific publications. Independently of the form, SKELETOME requires a mechanism to keep track of the provenance of the data and knowledge, in order to ensure proper privacy and access control. It also needs to provide a measure of certainty of derived data and to leverage expertise from the content and to streamline the delivery of the most relevant information to the most appropriate person.

The SKELETOME platform introduces an ontology-driven knowledge engineering cycle that supports the continuous evolution of the knowledge captured in the ISDS Nosology. Its foundation lays in the Bone Dysplasia Ontology (BDO) [6], a comprehensive and formal representation of the many different concepts involved in documenting the full complexity of the skeletal dysplasia domain. BDO defined more than 1200 concepts that describe complex relationships between disorders, phenotype and genotype, by integrating and re-using well known ontologies, such as, the Human Phenotype Ontology [7] or the NCI Thesaurus [8]. The ontology transforms the existing ISDS Nosology into a machine-understandable knowledge base and provides a shared conceptual model that serves as foundational building block for future extensions and as a mechanism for facilitating knowledge extraction and reasoning.

Using this ontology, the SKELETOME platform provides the following services:
- a collaboration environment for experts to exchange knowledge and link disorder descriptions to related evidence and Web resources (e.g., publications, radiographic data, gene databases, etc);
- semantically enhanced content annotation and integration services;
- ontology-driven text processing of publications leading to rich semantic annotations;
- enhanced image search and retrieval via ontology-based annotation;
- inference using domain knowledge.

ACKNOWLEDGEMENTS

The work presented in this paper is supported by the Australian Research Council (ARC) under the Linkage grant SKELETOME - LP100100156.

REFERENCES


ABOUT THE AUTHORS

Dr. Tudor Groza is a Postdoctoral Research Fellow in the eResearch Lab at The University of Queensland. He has published more than 40 peer-reviewed publications on topics related to knowledge representation, information extraction and knowledge discovery in the biomedical domain.

Craig McNamara is a web designer and developer in the eResearch Lab at The University of Queensland. Craig has a Bachelor of Software Engineering (Hons) and a Bachelor of Multimedia Design from the University of Queensland. He is currently working on the HCI side of the SKELETOME project. His interests include development of social platforms, user experience design and user interaction design.

Prof. Andreas Zankl is an internationally renowned expert on skeletal dysplasias, clinical geneticist at RBWH and leader of the Bone Dysplasia Research group at UQCCR. He is the author of 50+ journal articles on the topic, discovered two genes and is the clinical coordinator of the European Skeletal Dysplasia Network.

Prof. Jane Hunter is Director of the eResearch Lab and has published over 100 peer reviewed papers in the areas of eScience and Semantic Web. She is also Deputy Chair of the Academy of Sciences National Committee on Data in Science and on the Scientific Committee of the World Data System.