



Abstract Book

Europe Biobank Week 2016

August 22th 2016



Summary

Oral Communications	p03
P – Plenary sessions	p04
A – Innovation sessions	p15
B – Patient and ELSI sessions	p50
C – Data sessions	p86
D – Biomarker sessions	p115
E – Quality management sessions	p145
F – LPC Forum	p 156
G – Biobanking for environmental and biodiversity sessions	p168
Poster Communications	p178
Biospecimen Research	p179
Education & Standards	p201
ELSI	p215
Informatics	p227
Management	p240
New technology	p259
Organisation profiles	p265



Oral Communications



Oral Communications

C – Data sessions



Session: Biobanks as data and knowledge infrastructures

C3-5: Semi-automated free-text data annotation in a biobank registry using (bio-) medical ontology concepts from a RDF-database

P. Hofer-Picout*(1), S.Neururer(1), G.Goebel(1)

(1)Department of Medical Statistics, Informatics and Health Economics, Medical University of Innsbruck, Austria

Free-text data entered in biobank information systems has a negative impact on integration and reusability of clinical and sample data for research purposes. Standard medical terminologies facilitate readability and comparability of biomedical data across heterogeneous databases. However, single terminologies frequently do not provide enough concepts to cover a specific domain. For instance, ICD-10 provides concepts for most common diseases, but only a small subset of definitions for the rare diseases domain. Many (bio-) medical ontologies provide comprehensive semantic background knowledge, such as synonyms, taxonomies and cross-references to other terminologies. Using concepts of multiple, cross-referenced ontologies to annotate free-text in biobank information systems or registries will improve semantic retrieval and machine-readability of biobank resources. In our approach, (bio-) medical ontology concepts in a graph database are automatically recommended to data administrators, entering free-text diagnosis definitions in a hospital-wide biobank registry, based on lexical matching and SPARQL. We assume that accurate concept recommendations encourage data administrators to use coding standards rather than custom definitions. In a semi-automated step, concepts recommended by our prototype are manually confirmed to ensure correctness of annotations. Selected ontology concepts are stored together with free-text entered, providing additional background information which can be used for semantic queries and data integration across heterogeneous systems. Moreover, annotated biobank registry data are mapped with corresponding concepts in the RDF-graph database. Finally, we illustrate the practical usability of semantic annotation and graph-based representation of biobank data by exemplar semantic query scenarios using a prototype and the open source libraries Elasticsearch and BlazeGraph.

Keynotes: semantic, annotations, registry