**Areas of SCAI research and development around OpenBEL**

1. **BEL network resources in the area of neurodegeneration**

All networks will be publically available. The Alheimer Model is already published, the others are in preparation.

1. **Alzheimers disease Model** (part of this model is published [1]),
	* + 1. 9645 Nodes and 10251 edges
			2. 35266 articles, 44437 BEL statements
2. **Parkinson’s Disease Model** [2]
	* + 1. 1424 nodes and 2690 edges
			2. 432 Citations and 2236 BEL statements
3. **Post Traumatic Stress Disorder BEL Model (**Kodamullil et al. in preparation**)**
	* + - 1. 1731 nodes and 5004 edges
				2. 250 citations and 3554 BEL statements
4. **Type 2 Diabetes BEL Model** [3]
	* + 1. 1094 nodes and 2414 edges
			2. 1333 BEL statements
5. **Epilepsy BEL Model** (Filep et al. in preparation)
6. **Neuroinflammation Model (**Kodamullil et al in preparation**)**
	* + 1. 743 nodes and 1385 edges
			2. 990 BEL statements and 228 articles
7. **Extension of BEL syntax**
8. **Epigenetic Models** [3]

Alzheimer, Parkinson and Multiple Sclerosis Model

Example: g(dbSNP:rs429358)

1. **Genetic Variants Models** [4]

Alzheimer and Parkinson

Example: g(HGNC:SNCA, gmod (M,926))

1. BEL syntax extensions in the context of misfolded protein spreading

Example: p(HGNC:SDC2) -> tloc(a("Amyloid beta-peptides"), BRCO:Striatum, BRCO:"Superior\_temporal\_gyrus")

1. BEL (plus): integrating quantitative data in BEL :

First ideas/examples:

-Correlation with high/low expression:

 a(BRCO: “substantia nigra”) - - HGNC:SNCA, pmod(P), var (High)

 -inclustion of time point annotations for longitutional data

 Set Time Point= “Month 0”

1. **BEL Graph Database**

BEL networks are stored by the Graph Management and Analyze Software BiNE (Biological Network Explorer) in 3 different formats depending on the intended use :

1. Graph database Neo4J with the possibility to use the Graph Query Language Cypher

2. Relational database MySQL which outperforms graph databases in serveral issues and used to store the data- and user management

3. NetworkX Object with an awesome number of graph algorithms and a very high performance

The BiNE compiler/parser loads the BEL networks, after checking syntax/semantic and mapping, into the database systems. Several modules are extending the node and edge attributes with additional information from InterPro, UniProt, CTD, PubMed, ClinicalTrial.gov, DrugBank and other databases. BEL networks can be visualized in the BiNe graph explorer providing standard graph search algorithms like shortest path and the possibility to filter for annotations and entities. Currently we are developing a sophisticated search engine which allows the user to extract revelant subgraphs depending on uploaded or precompiled large experimental datasets using various graph mining algorithms. Depending on the requirements we will provide workflow wizards supporting the user to find the needle in a haystack. Because of the relevant drug/diseased-based database links (e.g. DrugBank, CTD and ClinivalTrials.gov), the engine can support the user to identify hubs and hypothetical chmeicals/drugs changing the disturbed subnetwork of interest. A restful API provides macine accessible interfaces to the serach engine and network databases.

The BEL graph model database is based on Neo4J, a freely available graph database. It also consists of a compiler/parser to upload the BEL documents and visualize the resulting networks in a BEL graph explorer. This explore has search functions for annotations and entities. We are currently working on various graph mining algorithms: path and sub-graph searches linking treatments and biomarkers.

1. **Using BEL to instruct agent-based (dynamic) models**

A method has been developed to automatically convert a BEL document into an Agent-Based Model (ABM). This method makes use both of the BEL statements, BEL annotations, information about BEL nodes contained in an ontology and partly, information coming from (an) external file(s). The Java program loads the BEL KAM (OpenBEL Framework 2.0.0 Dragon Fruit; migration TODO) and converts it into NetLogo ABM code. The user can directly load the code into NetLogo, change default settings, adjust paramenters where needed and run the simulation. This way, the user can add dynamics to his BEL code. In brief, the tool converts BEL essentials to an ABM that displays an abstract level of the biology modelled in BEL.

1. **BEL Information Extraction workFlow (BELIEF)**

BELIEF offers a text mining web service to extract BEL statements automatically [5]. The user can upload full text documents or fetch abstracts from PubMed. The uploaded documents are annotated with various BEL namespaces and annotations and the extracted results are presented in a curation interface. In this interface, the recognized entities, extracted BEL statements, and context annotations are visualized. The user can edit or add statements and save them in a database. Furthermore, for the saved statements a syntax check is performed. Resulting BEL documents can be downloaded from the service.

1. **Concept Search**

A web-based entity search on all BEL namespaces and annotation definitions. For all BEL resources where we have terminologies, the the user can search not only the preferred names but the synonyms as well. This search component can easily be integrated in other services through an API.

1. **BioCreative task and evaluation platform for text mining**

BioCreative is an assessment platform that offers training and test environment for text mining tasks. In 2015, Fraunhofer SCAI (together with ETH Zürich and PMI) organized a BEL task to extract BEL statements from given sentences and to provide supporting text excerpts form publications [6]. In course of the BioCreative task organisation, we created

* a BEL gold standard corpus [7, 8] to develop BEL information extraction tools and
* a web-based BEL Evaluation Platform [9] to evaluate BEL information extraction tools.

**References**

[1] Kodamullil et al. 2015: Kodamullil AT, Younesi E, Naz M, Bagewadi S, Hofmann-Apitius M. Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis. Alzheimers Dement. 2015

Nov;11(11):1329-39.

[2] Karki et al. Comorbidity analysis between Alzheimer’s disease and Type 2 Diabetes Mellitus based on disease network models and the role T2DM drugs. Submitted to JAD

[3] Khanam Irin A et al. Modelling Approaches on Epigenetic Factors in Neurodegenerative and Autoimmune Diseases and Their Mechanistic Analysis.J Immunol Res. 2015;2015:737168.

[4] Naz M et al 2015 : Naz M, Kodamullil AT, Hofmann-Apitius M.Reasoning over genetic variance information in cause-and-effect models of neurodegenerative diseases. Brief Bioinform. 2016 May;17(3):505-16.

[5] Madan et al. The BEL Information Extraction Workflow (BELIEF): Evaluation in the BioCreative V BEL and IAT track. Database (Oxford). 2016 pii: baw136.

[6] Rinaldi et al. BioCreative V track 4: a shared task for the extraction of causal network information using the Biological Expression Language. Database (Oxford). 2016 Jul 9;2016. pii: baw067.

[7] Corpus: <https://wiki.openbel.org/display/BIOC/Datasets>

[8] Fluck et al. Training and evaluation corpora for the
extraction of causal relationships encoded in biological expression
language (BEL). Database (Oxford). 2016 Aug 23;2016. pii: baw113.

[9]: BEL Evaluation Platform:
<http://bio-eval.scai.fraunhofer.de/cgi-bin/General_server.rc>

**Developments needed for the OpenBEL Platform**

1. Platform for name spaces and annotations
	1. Regular updates of the current resources
	2. Upload of customer name spaces and annotation files
	3. Update of equivalence files for namespaces and annotations
2. BEL editor (BEL manager???)

Necessary/useful functions:

* 1. Syntax check of statements
	2. Autocompletion of entity names and functions
	3. Storing of BEL documents/nanopubs
1. Data Store for BEL documents/nanopubs

Necessary/useful functions (BEL manager further development???)

* 1. Upload/download
	2. Search functions
1. Compilation process
	1. Only available in old framework
	2. Only available for BEL 1.0
	3. Compilation process for BEL 2.0 is still missing
	4. Updates for new BEL syntax
	5. Compilation configuration interface (e.g. including equivalences or orthologues, expanding families….)
2. BEL network database

Storage of BEL networks

1. BEL visualisation

Currently in Cytoscape, for OpenBEL platform Cytoscape export and plugin has to be updated