BioKC: a platform for quality controlled curation and annotation of systems biology models

Luxembourg Centre for Systems Biomedicine

Carlos Vega[1], Valentin Grouès[1], Marek Ostaszewski[1], Venkata Satagopam[1], Reinhard Schneider[1]
[1] Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Luxembourg

Abstract

Standardisation of biomedical knowledge into systems biology models is essential for the study of the biological function. However, biomedical knowledge curation is a laborious manual process aggravated by the ever-increasing growth of biomedical literature. High quality curation currently relies on pathway databases where outsider participation is minimal.

The increasing demand of systems biology knowledge [1] presents new challenges regarding curation calling for new collaborative functionalities to improve quality control of the review process. These features are missing in the current systems biology environment, whose tools are not well suited for an open community-based model curation workflow. On one hand, diagram editors such as CellDesigner or Newt provide limited annotation features. On the other hand, most popular text annotations tools are not aimed for biomedical text annotation or model curation. Detaching the model curation and annotation tasks from diagram editing improves model iteration and centralizes the annotation of such models with supporting evidence. In this vain, we present BioKC (Biological Knowledge Curation), a web-based platform for systematic quality-controlled collaborative curation and annotation of biomedical knowledge following the standard data model from Systems Biology Markup Language (SBML).
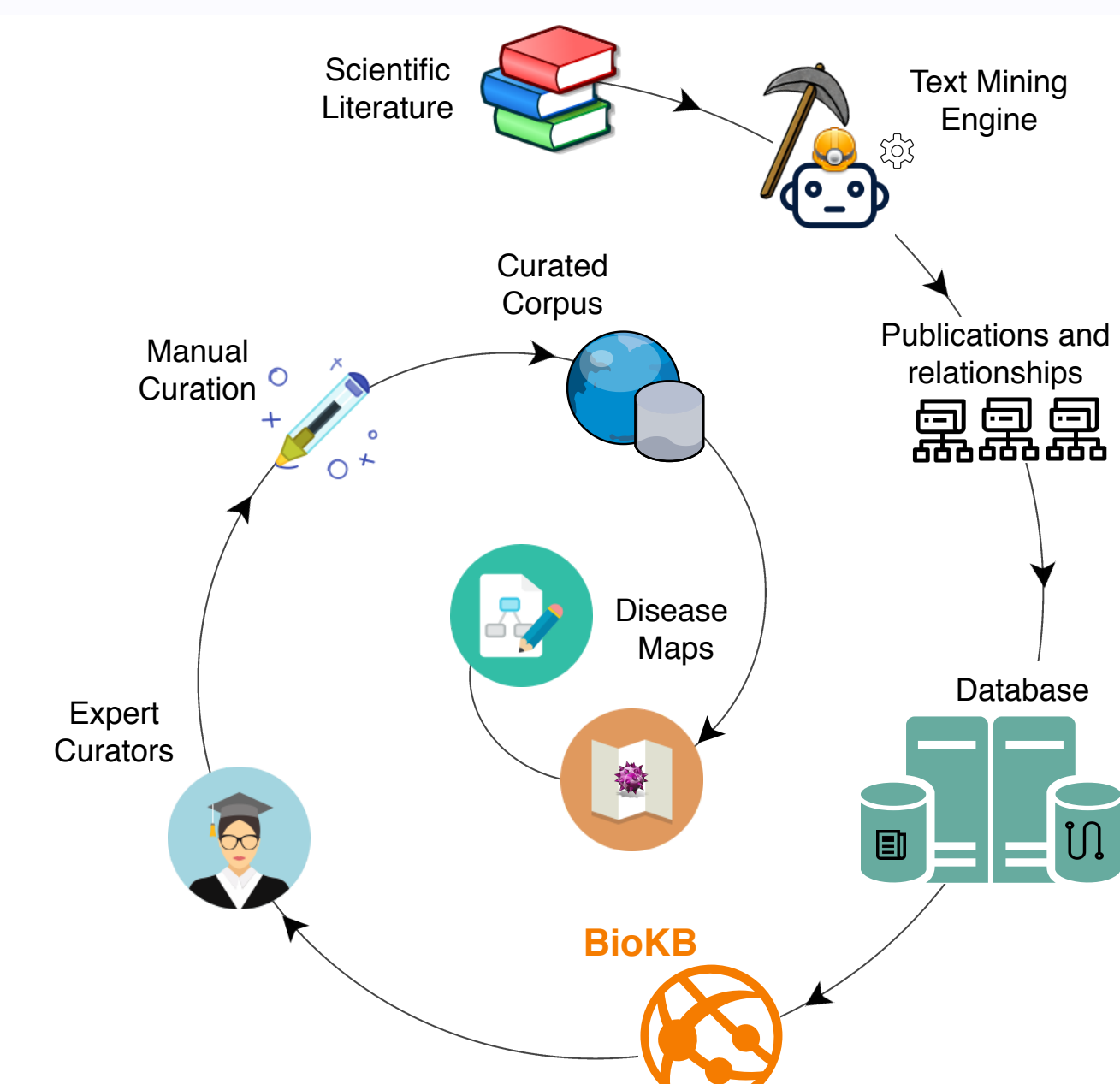
A Fact as a Knowledge Unit

- Fact: Minimal piece of representative knowledge that can be cited, referenced and attributed.
- Nano Publications employs three basic elements: assertion, provenance and publication information [3].
- BioNotate snippets confirm or rule-out a relationship between two known entities [4].
- **BioKC** SBML-based facts provide flexibility on the fact granularity and interoperability with other tools.

BioKC Proposal

- Traditionally, curation requires layout-aware tools that lack collaborative features.
- There is an interoperability gap between visual tools and annotation/curation tools (see Table 1). *
- BioKC is built on top of BioKB [2], a web-based interface designed to browse the text-mining results of almost 30 million publications (see Figure 1).
- BioKC systematic workflow for the curation and annotation of facts ensures high quality control.
- To assist this process, BioKC includes review mechanisms such as tasks, comments, change log, etc. (Fig 2).
- Facts can be annotated with supporting evidence from BioKB or third-party sources
- Annotations also support ontology terms from identifiers.org.

Available at http://biokb.lcsb.uni.lu/

Figure 1. BioKB and BioKC pipeline for systems biology curation.

Purpose	Tool	Web-based	Collaborative	Linkable	Annotation	Layout-Aware	SBML / SBGN
General Text Annotation	tagtog [5]	✓	✓	✓	✓	X	X
	brat [6]	✓	✓	X	✓	X	X
	WebAnno [7]	✓	✓	X	✓	X	X
	FLAT [8]	✓	✓	X	✓	X	X
Biomedical Text Annotation	MyMiner [9]	✓	X	X	✓	X	X
	BioQRator [10]	✓	X	✓	✓	X	X
	ezTag [11]	✓	X	✓	✓	X	X
Diagram Editor	CellDesigner [12]	X	X	X	✓	✓	✓
	Newt [13]	✓	X	X	✓	✓	✓
Visual Repository	WikiPathWays [14]	✓	✓	✓	X	✓	✓
	KEGG [15]	✓	X	✓	X	✓	X
	Reactome [16]	✓	X	✓	X	✓	✓
Visualisation Platform	Cytoscape [17]	X	X	X	✓	✓	X
	NaviCell [18]	✓	X	✓	✓	✓	✓
	MINERVA [19]	✓	X	✓	✓	✓	✓
	BioUML [20]	✓	✓	X	✓	✓	✓

Table 1. Purpose, type and capabilities of different tools. Collaborative column states which tools allow multi-user simultaneous operation. Linkable criterion refers to the ability to share and use the tool output as annotatable content via URI-like links. Conversely, the annotation criterion indicates if a tool is able to produce annotations on the content.

References

[1] Biomodels https://www.ebi.ac.uk/biomodels/content/news/biomodels-release-26th-june-2017
[2] Biryukov, M., Grouès, V., and Satagopam, V. P. (2017). BioKB - Text mining and semantic technologies for the biomedical content discovery. SWAT4LS
[3] Groth, P, Gibson, A., & Velterop, J. (2010). The anatomy of a nanopublication.
[4] Cano, C., Labarga, A., Blanco, A., & Peshkin, L. Collaborative semi-automatic annotation of the biomedical literature. 11th International Conference Intelligent Systems Design and Applications.
[5] Cejuela, J. M., McQuilton, P., Ponting, L., Marygold, S. J., ... (2014). tagtog: interactive and text-mining-assisted annotation of gene mentions in PLOS full-text articles. Database, 2014.
[6] Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., & Tsujii, J. I. (2012, April). BRAT: a web-based tool for NLP-assisted text annotation. In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics.
[7] Yimam, S. M., Gurevych, I., de Castilho, R. E., & Biemann, C. (2013, August). Webanno: A flexible, web-based and visually supported system for distributed annotations. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations.
[8] Van Gompel, M. and Reynaert, M. (2013). Flat - folia linguistic annotation tool.
[9] Salgado, D., Krallinger, M., Depaule, M., Drula, E., ... & Marcelle, C. (2012). MyMiner: a web application for computer-assisted biocuration and text annotation. Bioinformatics.
[10] Kwon, D., Kim, S., Wei, C. H., & Leaman, R., & Lu, Z. (2014). Assisting manual literature curation for protein–protein interactions using BioQRator. Database, 2014.
[11] Kwon, D., Kim, S., Wei, C. H., Leaman, R., & Lu, Z. (2018). ezTag: tagging biomedical concepts via interactive learning. Nucleic acids research.
[12] Funahashi, A., Morohashi, M., Kitano, H., & Tanimura, N. (2003). CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. Biosilico.
[13] Sari, M., Bahceci, I., Dogrusoz, ... (2015). SBGNViz: a tool for visualization and complexity management of SBGN process description maps. PloS one.
[14] Kutmon, M., Riutta, A., Nunes, N., Hanspers, K., Willighagen, E. L., Bohler, A., ... & Coort, S. L. (2016). WikiPathways: capturing the full diversity of pathway knowledge. Nucleic acids research.
[15] Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research.
[16] Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., ... & Jassal, B. (2014). The Reactome pathway knowledgebase. Nucleic acids research.
[17] Shannon, P., Markiel, A., Ozier, O ... & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research.
[18] Kuperstein, I., Cohen, D. P., Pook, S., Viara, E., Calzone, L., Barillot, E., & Zinovyev, A. (2013). NaviCell: a web-based environment for navigation, curation and maintenance of large molecular interaction maps. BMC systems biology.
[19] Gawron, P., Ostaszewski, M., Satagopam, ... & Schneider, R. (2016). MINERVA—a platform for visualization and curation of molecular interaction networks. NPJ systems biology and applications.
[20] Kolpakov, F., Puzanov, M., & Koshukov, A. (2006, July). BioUML: visual modeling, automated code generation and simulation of biological systems. In Proceedings of The Fifth International Conference on Bioinformatics of Genome Regulation and Structure.
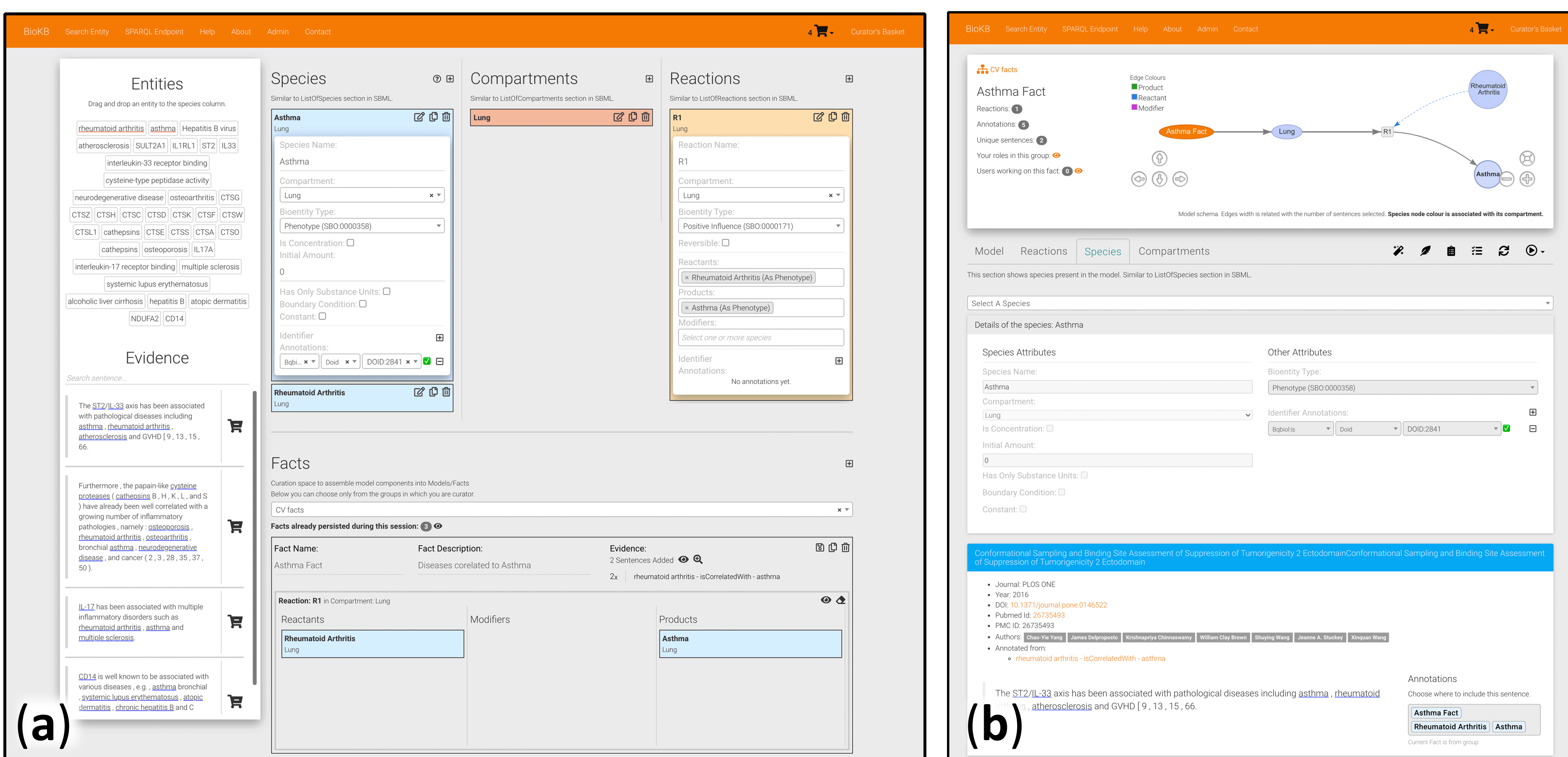
(a)

(b)

Figure 2: BioKC interface and functionalities. **(a)** In the basket view, facts and their elements can be composed from sentences retrieved in BioKB or from TSV files. **(b)** The fact view is where facts can be edited, either from scratch or after being persisted in the basket view.