Ref. Ares(2016)2447614 - 26/05/2016

D3.2 Set up of the BioLinX Intelligence Platform

BioLinX WP 3

BioLinX Partnering and Intelligence platform

Description and screenshots of the BioLinX Intelligence Platform

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 652692

Document description

Relates to Deliverable	D3.2
Title	Set up of the BioLinX Intelligence Platform
Date	May 2016
Dissemination level	Public
Work Package	3
Author(s)	Dr. Andreas Scriba
Version	1.0

Table of Content

1 Introduction

WP3 - BioLinX Partnering and Intelligence Platform

Work package leader:	DECHEMA e.V.
Partners of work package 3:	INNOVATION ENGINEERING SRL (INNEN)
	NV REWIN WEST BRABANT
	PNO CONSULTANTS GMBH
	REWIN PROJECTEN BV
	SP SVERIGES TEKNISKA FORSKNINGSINSTITUT AB

The aim of the workpackage 3 is

1. Set-up of a dedicated BioLinX Partnering Platform
2. Adapt the partnering platform to the needs of the BioLinX project
3. **Set-up and integration of the Intelligence platform provided by Innovation Engineering SRL**
4. Administration the BioLinX Partnering and Intelligence Platform

Milestone nr	Title	Lead	Due Date	Means of verification
M 3.1	Set-up partnering platform	DECHEMA e.V.	February 2016	
M 3.2	Set up of the BioLinX Intelligence Platform	INNEN	April 2016	

Tabel 1: Overview of Milestones

Project role	Name	Beneficiary	email
Coordinator	Dennis van der Pas	REWIN BV	d.vanderpas@rewin.nl
WP lead	Andreas Scriba	DECHEMA e.V.	scriba@dechema.de
D3.1 lead	Andreas Scriba	DECHEMA e.V.	scriba@dechema.de
D3.2 lead	Marco Bonfigli	INNEN	m.bonfigli@innovationengineering.eu

Tabel 2: Contacts

2 Integration of Intelligence Platform

The amount of information and stakeholders in the European biotechnology landscape is overwhelming. In order to identify, contact and meet potential cooperation and business partners quickly and to easily get reliable information on patents, public funded projects and open source literature, the BioLinX project has developed a "one-stop-shop" for relevant business and scientific information that helps to save time, to find partners and to reduce the time-to-market duration. It consists of the BioLinX Partnering Platform and the BioLinX Intelligence Platform.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 652692

2.1 Search functions

The services by the BioLinX Intelligence Platform, provided by INNEN, have been successfully integrated into the BioLinX Partnering Platform. This integration was the major task of work package 3, task 3.2. To get access to the services of both platforms the users have to register only once. This saves time and prevents mistakes because users have to remember only one password.

Basis of the BioLinX Intelligence Platform is the Discover-IT platform that was developed by Innovation Engineering within a former EU-funded project. The content and services from the BioLinX Intelligence Platform have been integrated in the BioLinX Partnering Platform by a technical interface. The integration is built upon an interface exposed by Innovation Engineering and called as a web service by the BioLinX Partnering Platform. DECHEMA has provided the technical interface to query the services/content from the Intelligence Platform by the Partnering Platform and display results to the users of the BiolinX Partnering platform. INNEN has provided a technical interface to the Intelligence Platform to deliver the expected services/output. The BioLinX Intelligence search features are a new part of the BioLinX Partnering Platform, the search results are retrieved from the BioLinX Intelligence Platform, sent to the BioLinX Partnering Platform, formatted accordingly to the BioLinX Partnering Platform specification and displayed there. The advantage of this kind of integration between the BioLinX platforms is that the users just register once at the BioLinX Partnering Platform. They have access to services and content of both platforms by logging into their personal partnering area (password protected).

The BioLinX Partnering and Intelligence platform provides access to:

1. Patents: 9 million patents from the European patents database;
2. Scientific papers: 18 million scientific open source papers from literature databases;
3. Research projects: 94 thousand European funded research projects from the CORDIS database

Via the newly integrated search functionalities BioLinX users can run searches on the BioLinX Intelligence Platform using parameters like

- Free test search for patents, projects and literature
- Search in the invention title, name of applicants, name of inventors and patent category (patents)
- Search in name of organisation, status of project and start/end of project (CORDIS projects)

Screenshots of the BioLinX Intelligence Platform

The following pages show screenshots of the BioLinX Intelliegence Platform. Each section is displayed. This gives an overview of the structure, the infomation that can be displayed on the platform, the functionalities and the "look and feel" of the BioLinX Intelligence Platform.

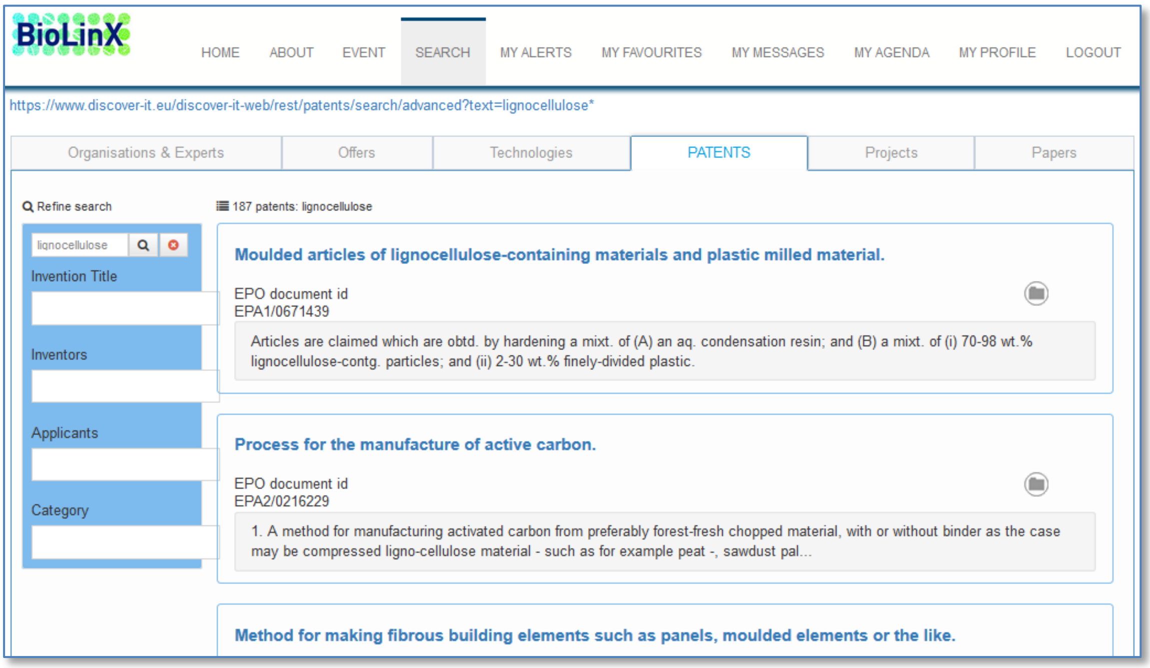

Figure 1: Search for patents in BioLinX Intelligence Platform

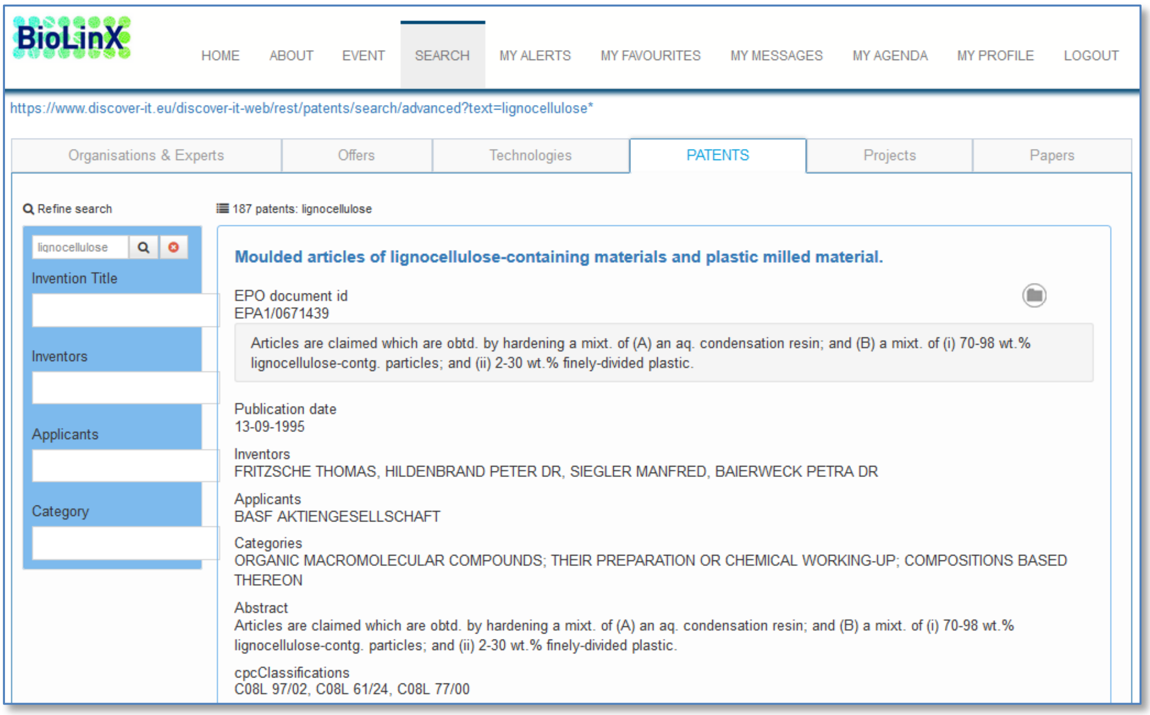

Figure 2: Search for patents, view on opened search result

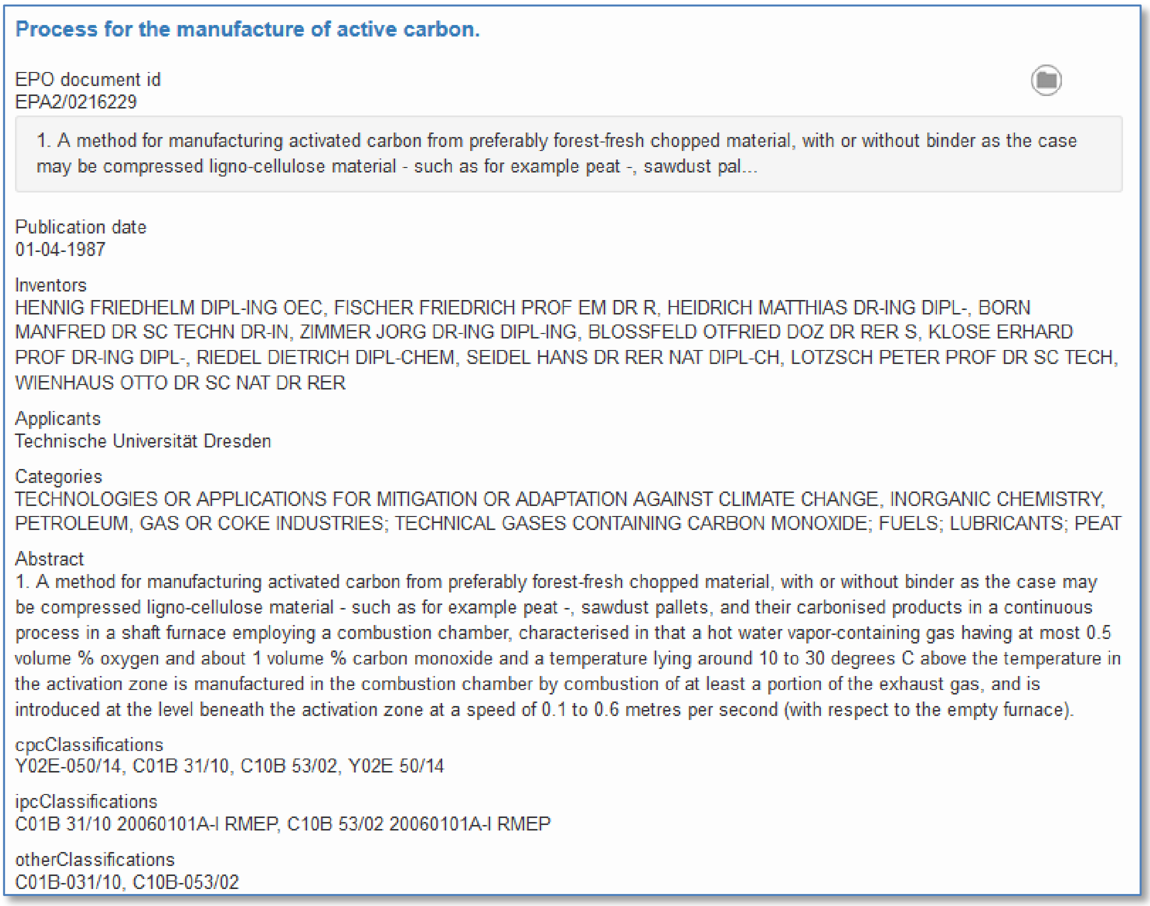

Figure 3: Search for patents, detail view of opened hit

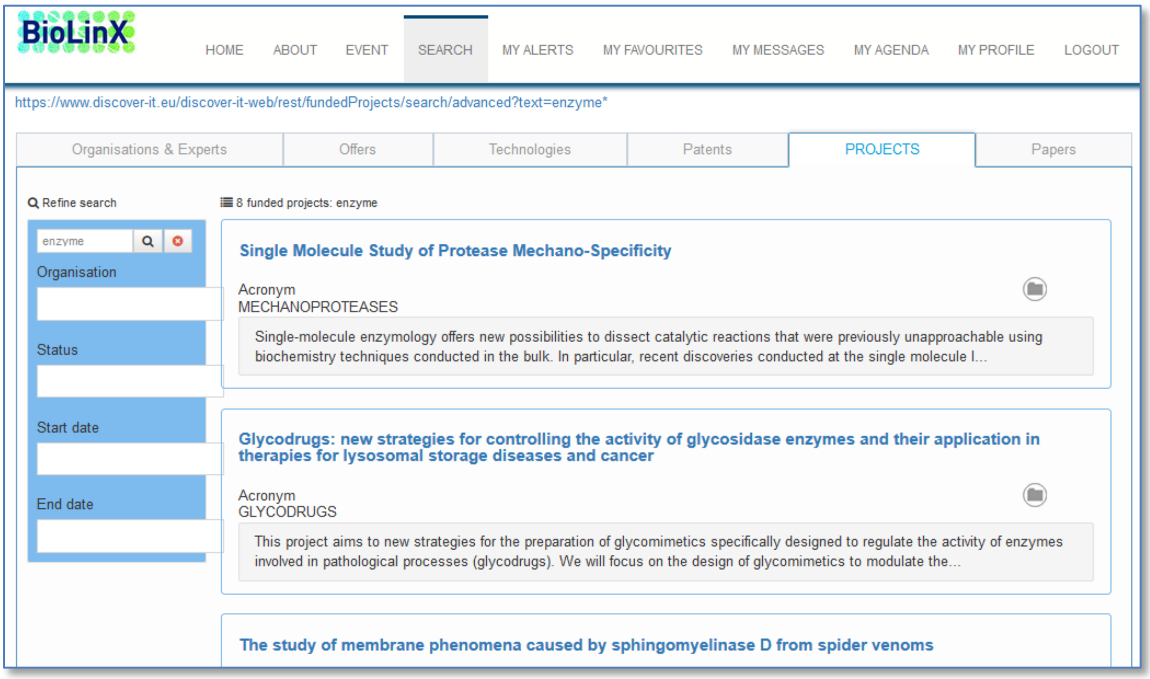

Figure 4: Search for projects – hit list

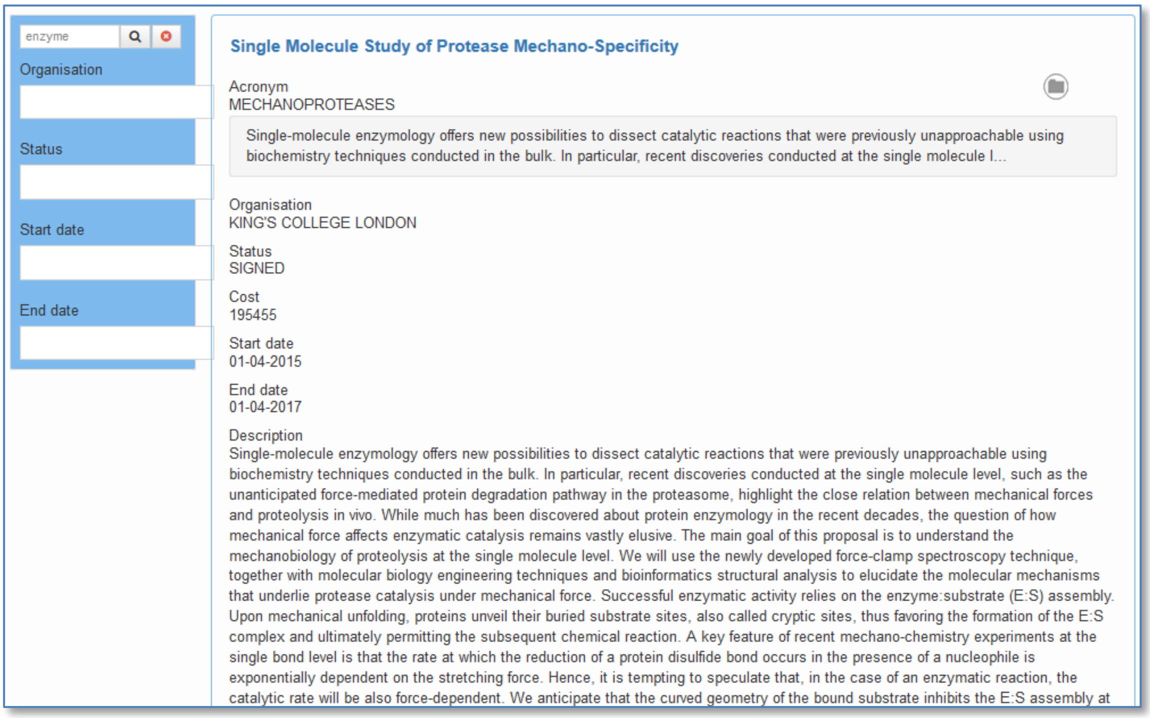

Figure 5: Search for projects – view on opened search result

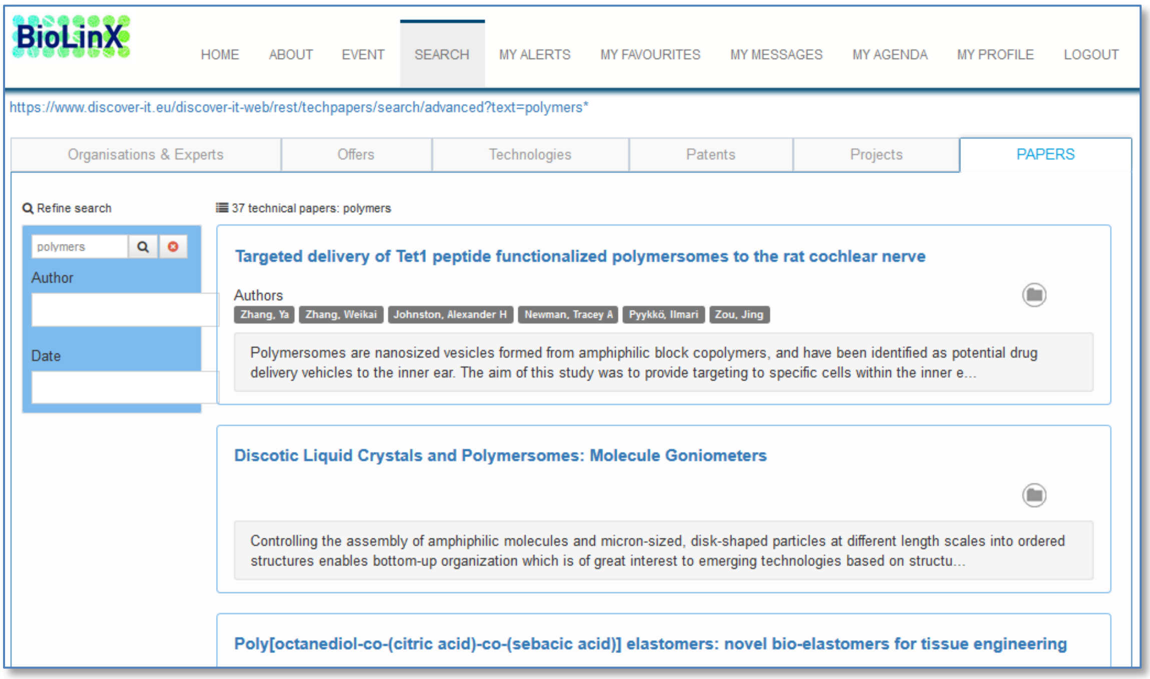

Figure 6: Search for literature (papers) – hit list

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 652692

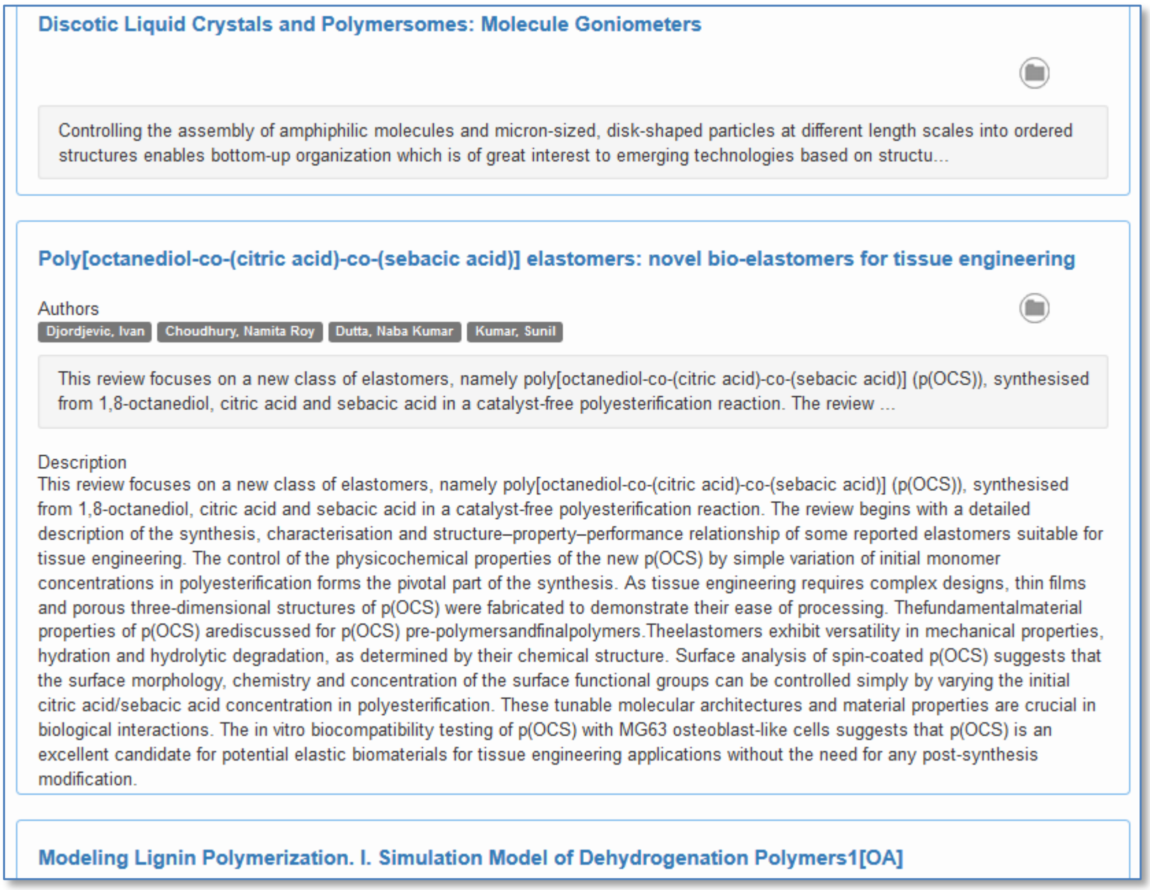

Figure 7: Search for literature (papers) – detail view of hit

2.2 Tagging functions

Based on the scope of the BioLinX project, a taxonomy has been developed that covers biomass sources and target industries (s. Annex 1). This taxonomy was originally to be used for the tagging of content of the BioLinX Intelligence Platform. This tagging feature was supposed to help users to find relevant content related to their sector of interest. The tags were to be used as keywords to boost the matrix used by the search system. The Latent Semantic Analysis approach implemented in the system makes use of vectorial proximity of words to find relevant contents. By adding specific keywords to the overall matrix used for vectorial proximity, the contents retrieved shall be more relevant for the domain of knowledge of the user.

In the course of the project, the aim and practicability of the envisioned tagging functionalities has been discussed more thoroughly. The lead question in this discussion has been what functionality offers the most benefits for users of the platform and will likely be accepted by users. From a user's point of view, tagging features should benefit the user directly and they should be fast and easy to use.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 652692

To tag content within the BioLinX Intelligence Platform, two methods are theoretically feasible:

1) Users tag content manually. When a user gets a result from a search on the platform, he is given the taxonomic list and asked to assign tags from this list to the search result. This requires extensive effort by the user.
2) Automatic tagging: For this functionality, additional user information has to be exchanged between the two platforms. When registering to the platform the user needs to submit a full and valid profile and needs to identify all relevant classifications. The user needs to mark (tag) the relevant information and this information is handed back from the Partnering to the Intelligence Platform along with the areas of interest the user has indicated in his profile. This is then used to apply tags to the document.

Both methods have considerable disadvantages: Manual tagging requires extensive effort by users as the user needs to assign tags based on the taxonomic list to every document. The project partners agree that this concept is unrealistic, as hardly any user will be willing to perform a lengthy tagging process at every document.

Therefore, the starting point of the BioLinX project was to work with automated tagging. The project team has however come to the conclusion that the automatic tagging based on the user profiles requires all users to submit a full and valid profile when registering. From the experience with former EU funded projects where similar platforms were used, this is a serious drawback for registrations; it might even cause potential participants in the BioLinX events to decide against participation, which could be harmful for the other deliverables. Alternatively, users might select classifications at random to fulfil the requirements superficially, leading to misinformation and in the end a falsified and therefore useless tagging.

Further discussion with the project team has revealed that the benefits of automatic tagging of content with taxonomic keywords, seem to be relatively scarce. Performing a taxonomic search on "enzyme" results in a plentitude of not very accurate hits as there is a multitude of enzymes. More specific searching would require to include all enzyme classifications (or best: tens of thousands of enzymes), which would make the taxonomic list unworkable.

During the discussions the project team has identified that 'free text search' would be the optimal solution. It does not require a lot of effort from the user nor does it create an unworkable taxonomic list. In the enzyme example the free text search allows users to find results for specific enzymes such as "lipase" or "amylase", narrowing down the search from the very beginning. It is the believe of the project team that the free text search makes the database much more user friendly and useful than a taxonomic search could ever do.

The project team therefore decided to use the free text search as the best solution. It requires a mininmum of effort from the user, prevents an overcomplicated registration process and / or an unworkable taxonomic list. The project team also realized that an important part of the tagging functionality for users is the possibility to retrieve their search results at any later time.

This part of the tagging functionality is implemented in such a way that it allows users to bookmark their personal search results such as patents, projects and scientific papers in order to get a faster

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 652692

access to previously identified relevant information. All bookmarked information can be easily stored in the “My favourites” section of the BioLinX Partnering Platform and deleted at any time. This newly integrated user-specific bookmarking functionality is user-friendly, helpful and time saving.

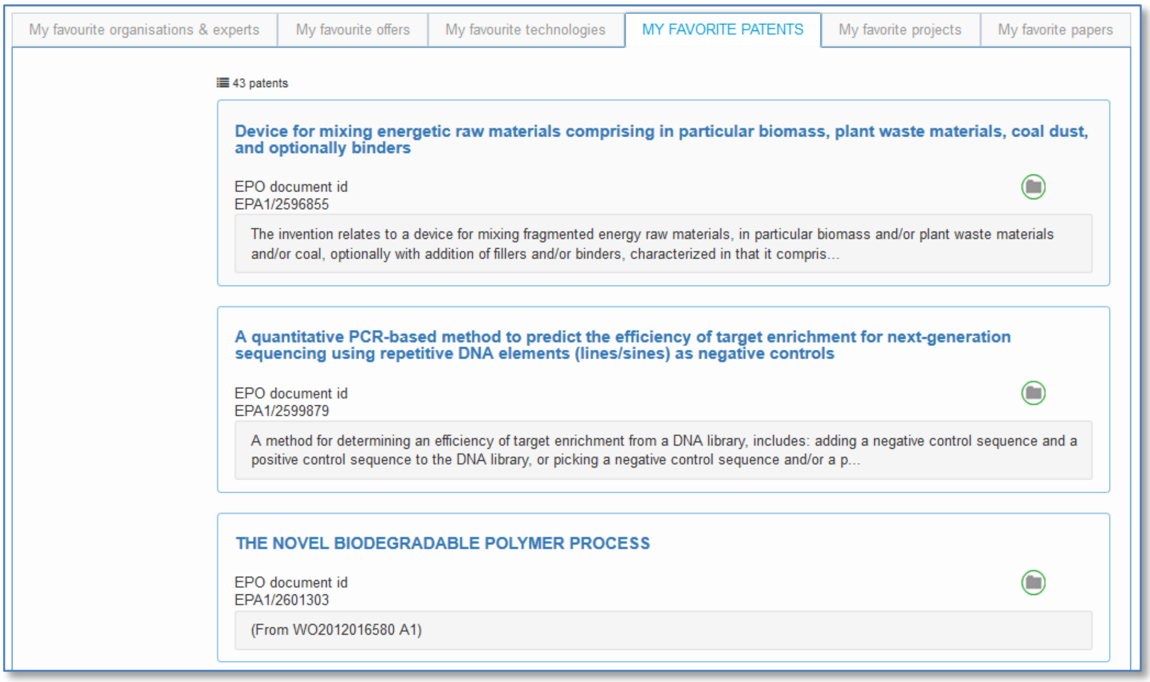

Figure 8: My Favourites – tagged patents

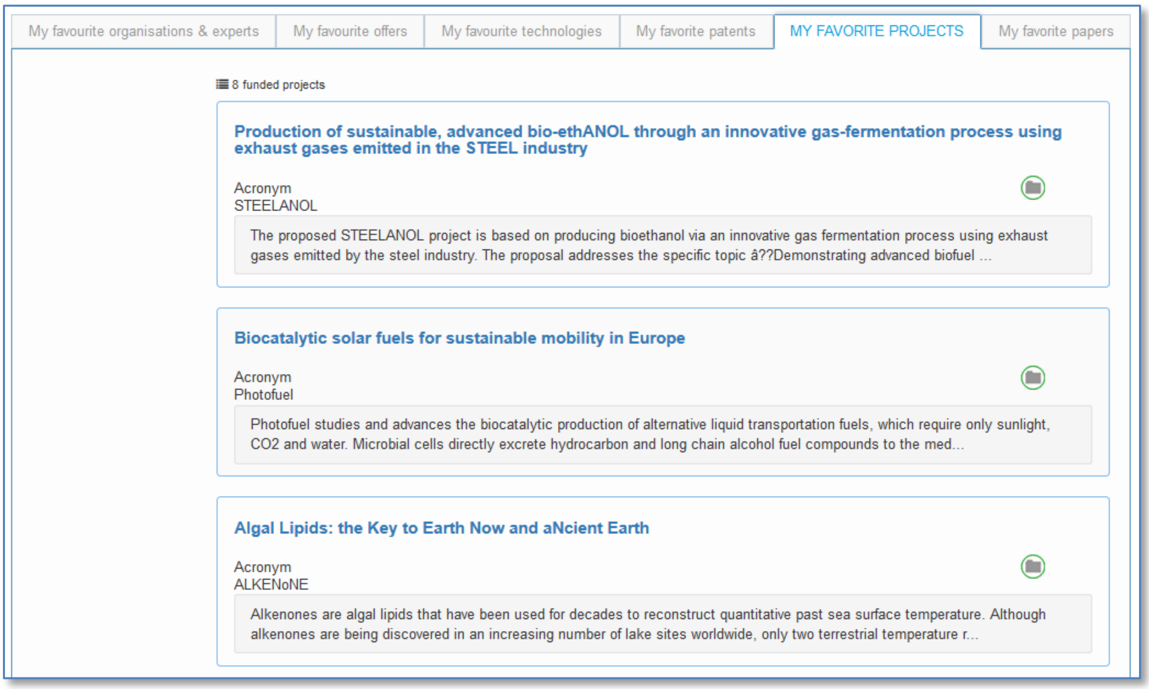

Figure 9: My Favourites – tagged projects

3.

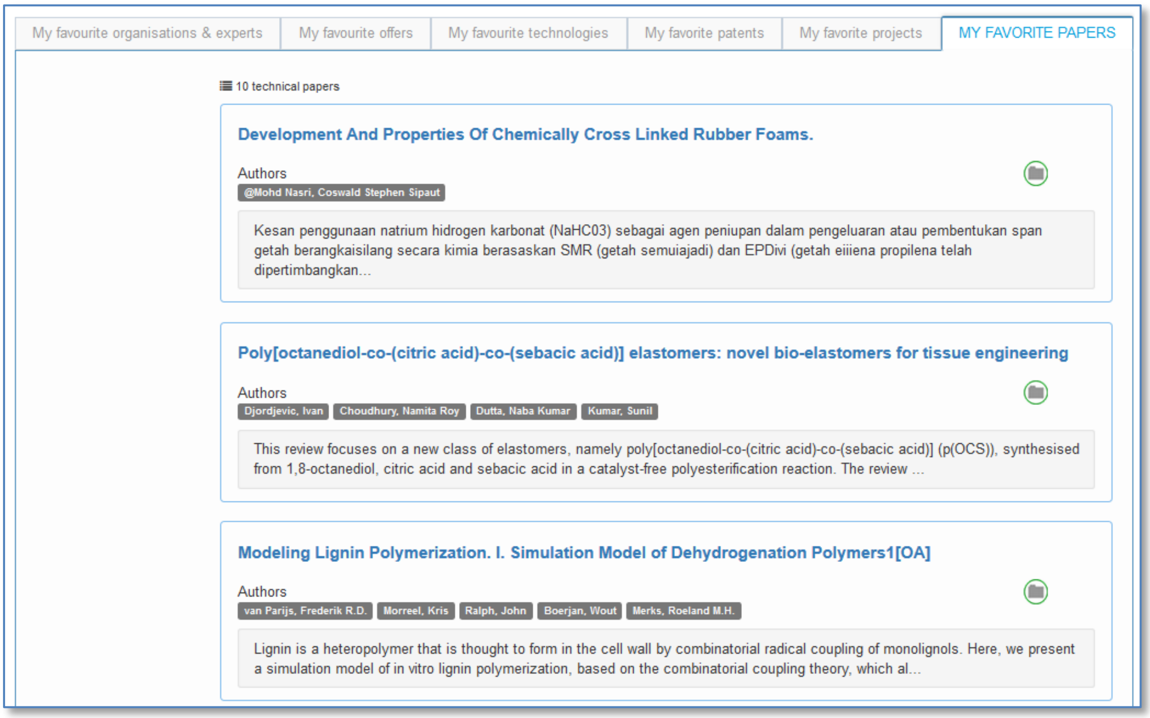

Figure 10: My Favourites – tagged literature

3 Next steps

The next steps in WP 3 are the ongoing Administration of the BioLinX Partnering and Intelligence Platform. The BioLinX Partnering Platform will in addition be used to support registration and partnering activities for the BioLinX regional events.

4 Summary

The Intelligence Platform provided by Innovation Engineering has been integrated successfully in the BioLinX Partnering Platform. BioLinX users can search for patents, projects and open source literature and find information on 9 million patents, 18 million scientific open source papers and 94 thousand European funded research projects.

The BioLinX Partnering Platform is ready for use in connection with the BioLinX regional events for registration and partnering, both on-site and online.

The tagging functionalities have been specified (free text search in combination with bookmarking) in the best interest of the users of the BioLinX service to ensure maximum benefit for the European biotechnology community.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 652692

Annex 1: Taxonomy

Classifikation 1: Biobased feedstocks & products (Headline)

Active ingredients

Algae biomass

Animal biomass wastes

Bio-based additives

Bio-based chemical intermediates

Biobased product

Bio-based resins

Biocomposites

Biofuels

Biogas

Biolubricants

Biopolymers

Biosolvents

Biosurfactants

Cellulose, hemicelluloses & lignin

Digestate

Enzymes

Food biomass

Herbaceous and agricultural biomass

Industrial biomass wastes

Lipids

Natural fibres

Oils & fats

Pharmaceutical products

Proteins

Pyrolysis oils

Seaweeds

Sewage sludge

Starch & sugars

Syngas

Wood and woody biomass

Classifikation 2: Target industry & application sectors (Headline)

Adhesives

Agriculture

Bio-based Industry

Construction materials

Chemistry

Coatings

Cosmetics

Design / Engineering

Detergency & home care

Electronic

Energy

Environment

Food & Nutrition

Human Health

Maintenance

Metallurgy

Packaging

Phytosanitary

Plastics / Rubber / Composites

Textile

Transport

Wood / Paper / Cardboard

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 652692