

# DD-DeCaF

## Bioinformatics Services for Data-Driven Design of Cell Factories and Communities

**Markus Herrgård**

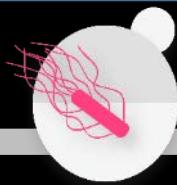
DD-DeCaF Project Coordinator

Professor & Director of Data Science and Automation

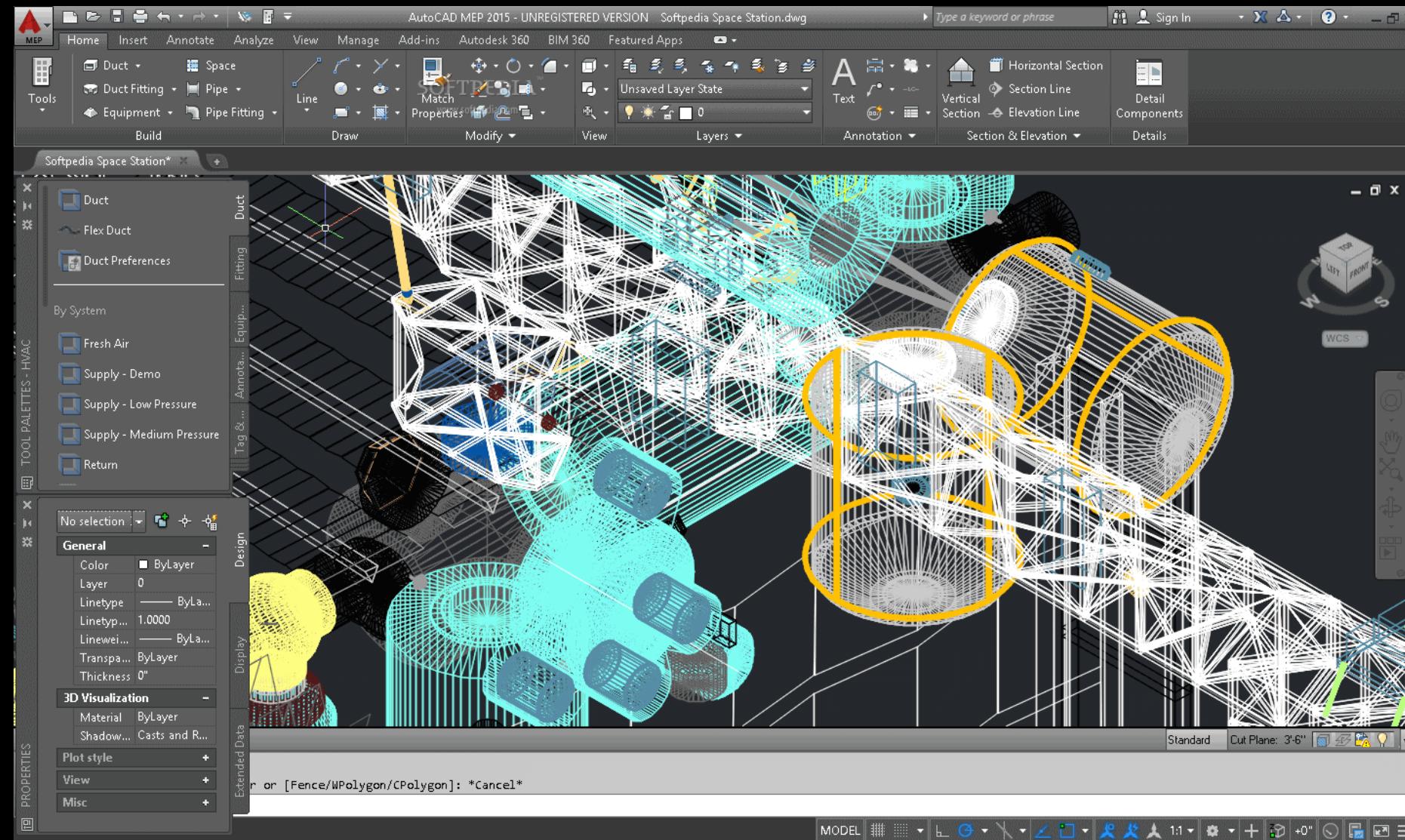
The Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark





# Design in Traditional Engineering Fields



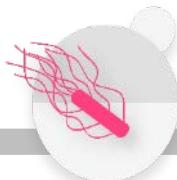


# Design in Biological Engineering

Our models are mostly qualitative and lack predictive power needed for engineering design

Omics data can help in obtaining more predictive models and enables improved design

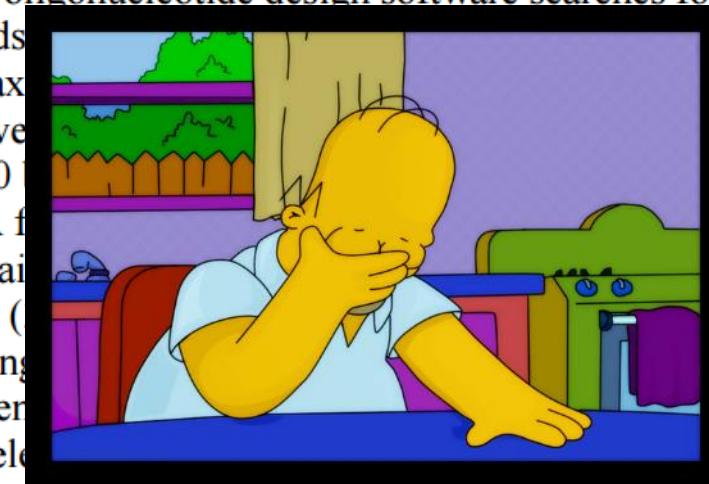




# But we can even design full bacterial genomes, right?

In the supplementary methods section for the paper:

**A. Oligonucleotide Design Software.** The oligonucleotide design software searches for a combination of parameters. These include dsDNA assembly stages, maximum fragment size, maximum number of oligos assembled per assembly stage and appended vector sequences. The software yields overlapping oligos that do not exceed 80% of the total oligo length, a size equal to half the oligo length and dsDNA fragments. In each assembly stage, with the exception of the eighth molecules which contain unique vector sequences, each molecule contains a unique set of vector sequences (40). At the end of assembly, a unique set of vector sequences (40) is appended to the 5' and 3' ends of each resulting oligo. The unique vector sequences at the ends of the oligos are unique to each sequence. These appended vector sequences are used for cloning or PCR amplification and insert recombination. The unique vector sequences also expose overlaps for subsequent assembly stages (40).



RE

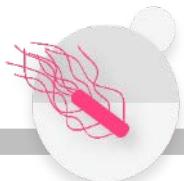
SYN

De  
ba

Clyd  
Thor  
Li M  
Lijie  
Chuc

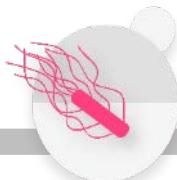
Hut

The oligonucleotide software was initially implemented as two WordPerfect macros: 1) Overlapping fragments.wcm, and 2) Oligonucleotides.wcm. The code for these macros is presented below. A streamlined version of the software is embedded in the Archetype® software package, which is commercially available through Synthetic Genomics, Inc.

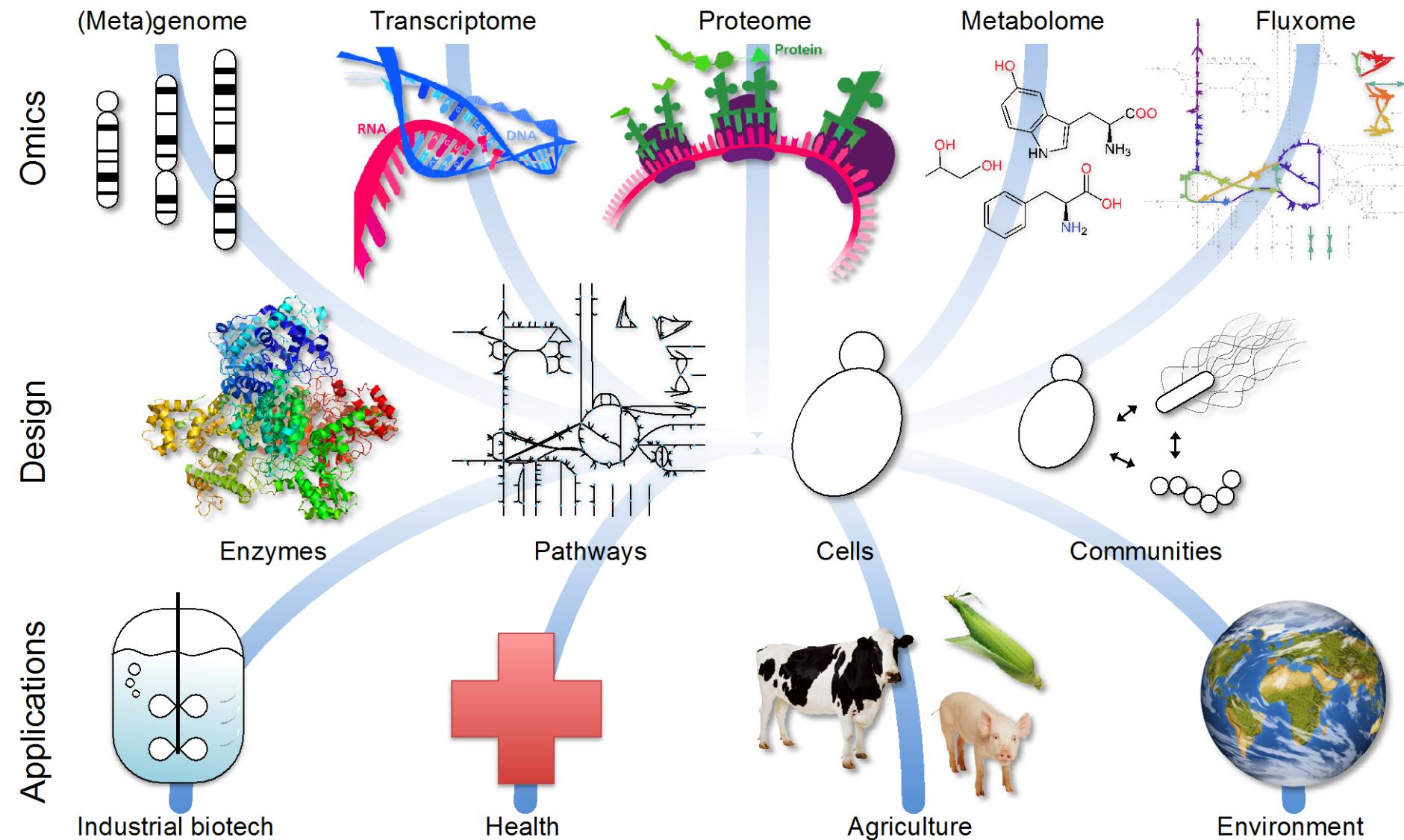


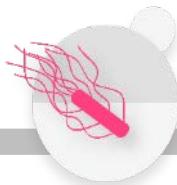
# Design complexity

Level	Length	Number of choices	Unit	Design space
DNA	50	4	Base pair	$10^{30}$
Protein	300	20	Amino acid	$10^{390}$
<i>Pathway</i>	<b>15</b>	<b>5</b>	<i>Part</i>	<b><math>10^6</math></b>
<i>Organism</i>	<b>1000</b>	<b>3</b>	<i>Gene</i>	<b><math>10^9</math></b>
<i>Community</i>	<b>3</b>	<b>50</b>	<i>Organism</i>	<b><math>10^{23}</math></b>
Process	10	5	Unit operation	$10^5$
Industry	1	10000	Chemical product	10000



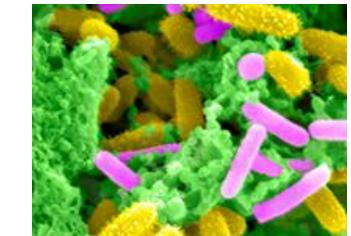
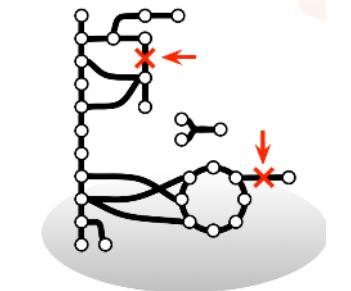
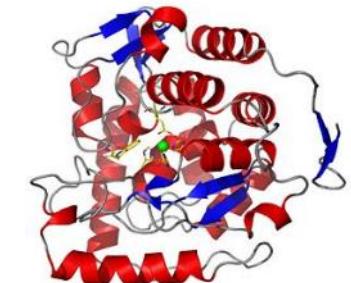
# Concept: From data through design to biotech products

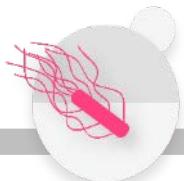




# Application areas

- 1. Design of novel enzymes and biochemical pathways by mining large-scale metagenomic data**
- 2. Design of cell factories for chemical and protein production by integrating multiple omics data types using next generation genome-scale models**
- 3. Design and analysis of microbial communities in industrial, health and environmental applications by building community-scale models based on metagenomic data**



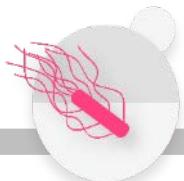


# Three key ideas of the project

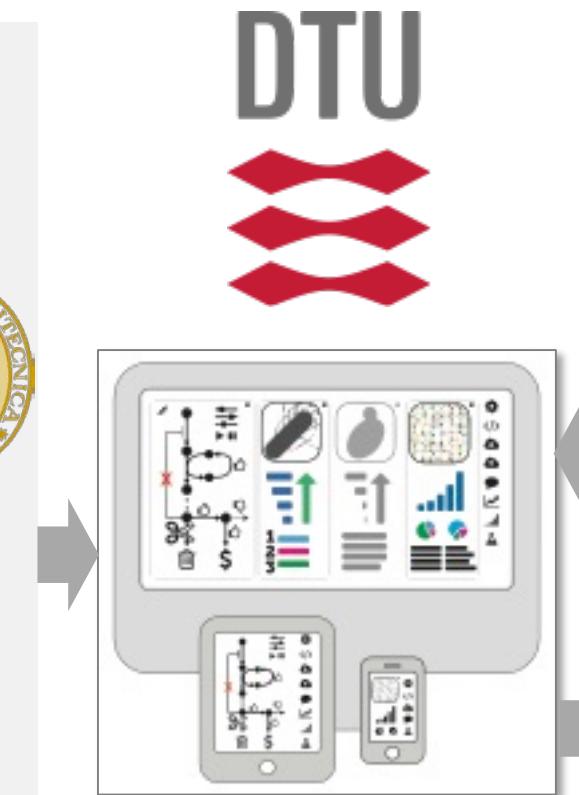
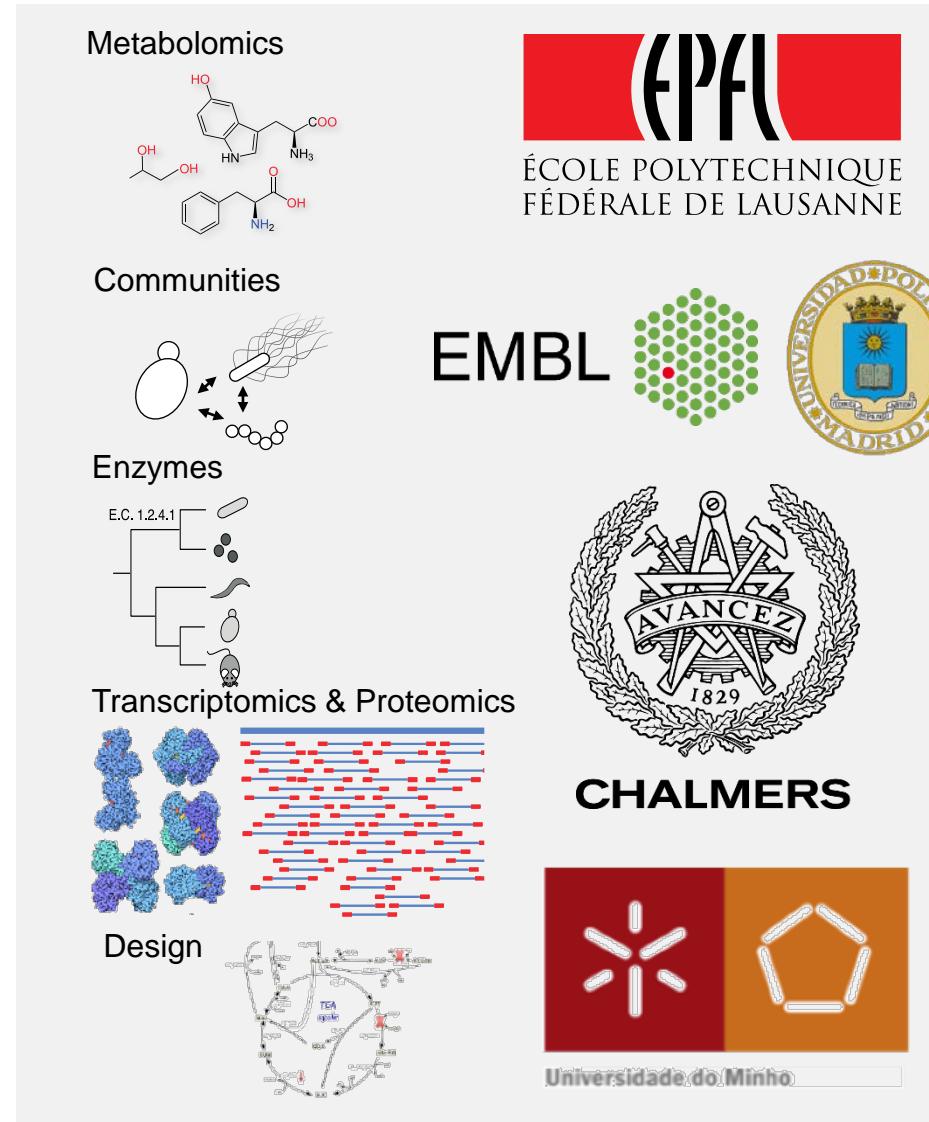
Use of genome-scale models as the scaffolds to integrate diverse data types

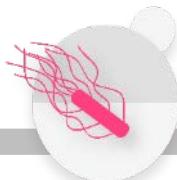
Use of microservices to integrate diverse databases, tools and methods

Direct involvement of end-user partners as part of the project

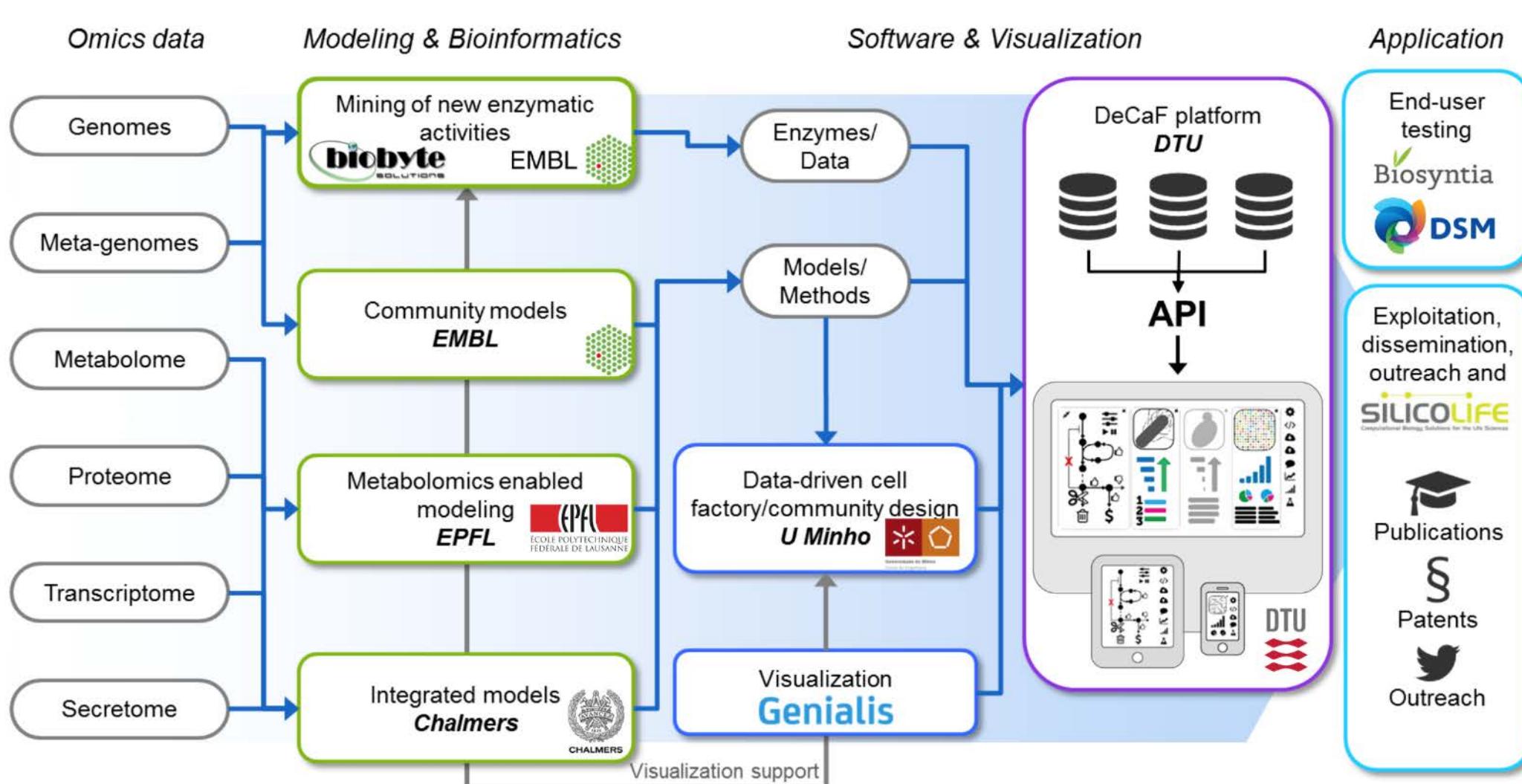


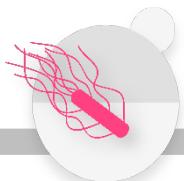
# Consortium





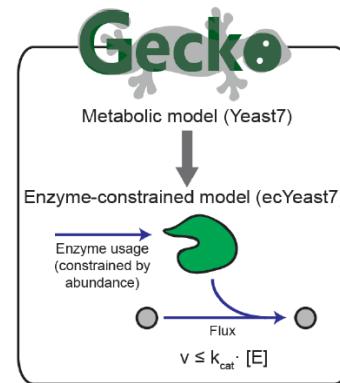
# Work plan



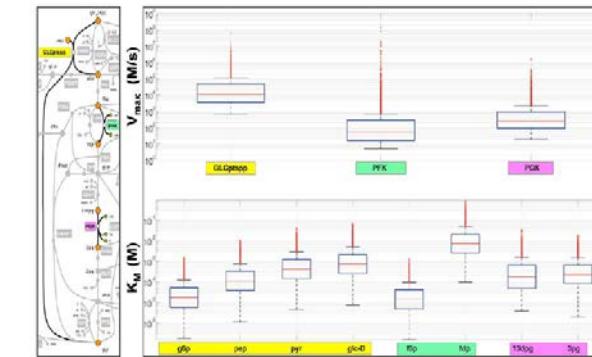


# Models and methods developed within DD-DeCaf

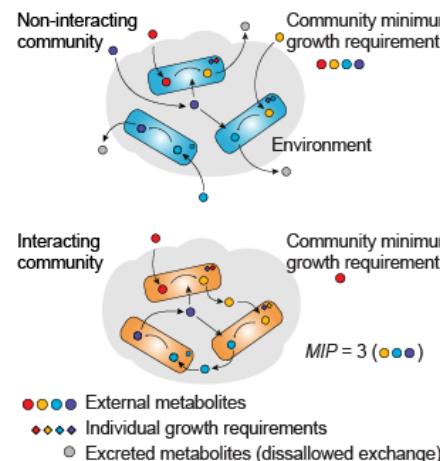
## GECKO – models incorporating proteomics data



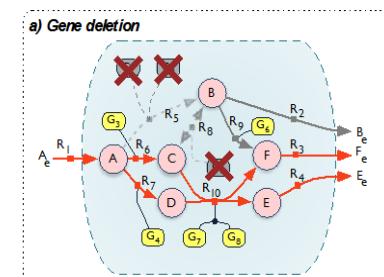
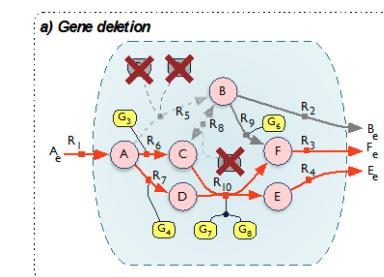
## Large-scale kinetic models incorporating metabolomics data

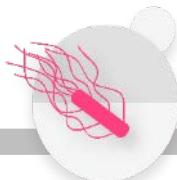


## Pipelines for building community models from metagenomic data

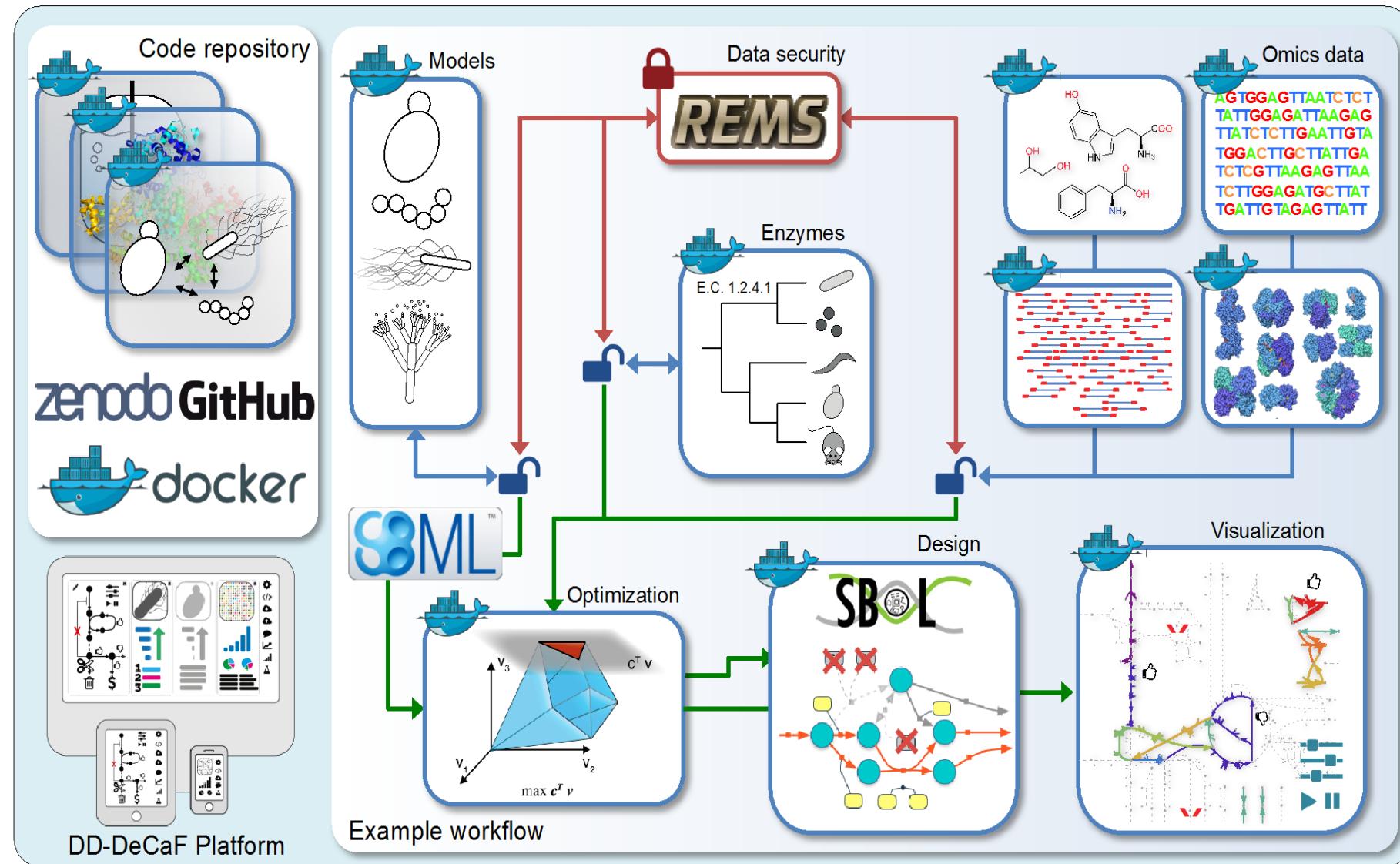


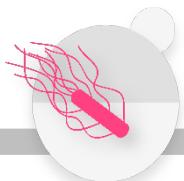
## Optimization-based cell factory design methods using advanced models



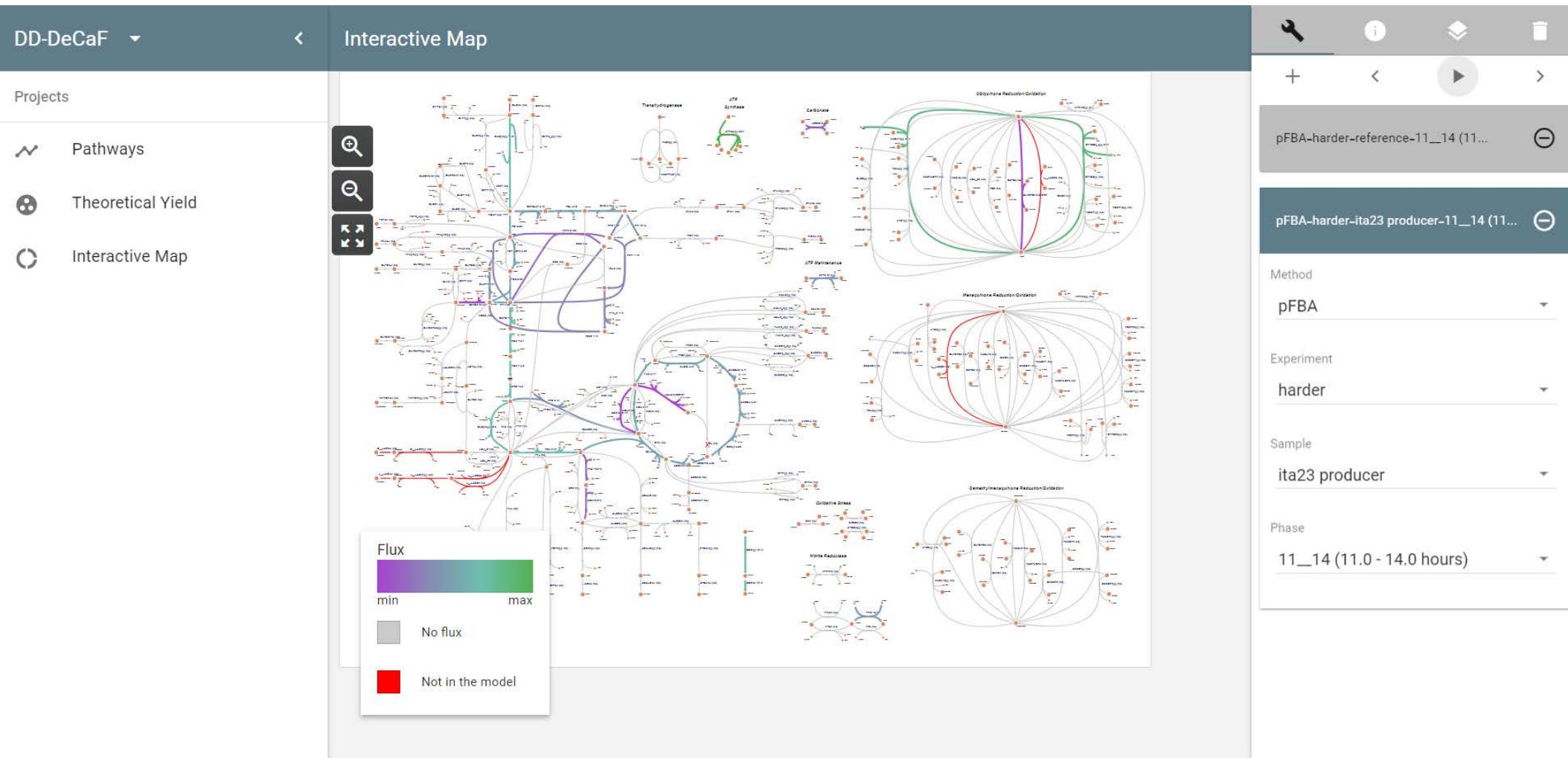


# Platform architecture: Modular microservices

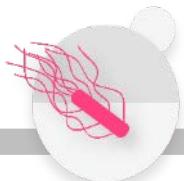




# Publicly available online software platform



<http://app.dd-decaf.eu>



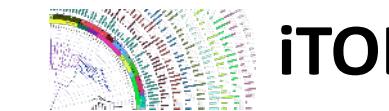
# Tools & databases associated with the project

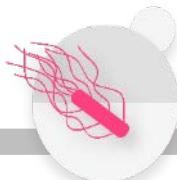


**EggNOG** eggNOG-mapper

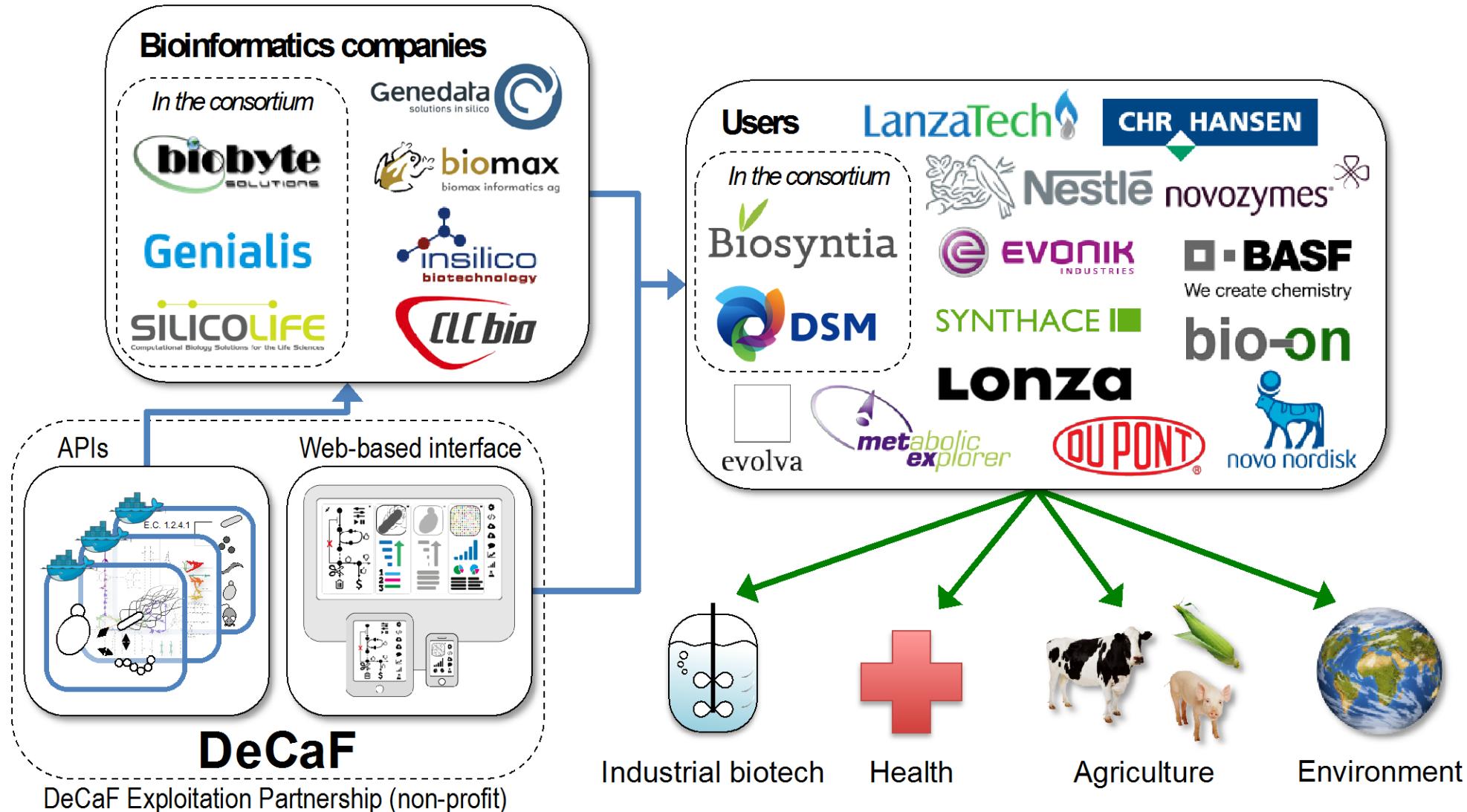
**NGLess**  
**optlang**

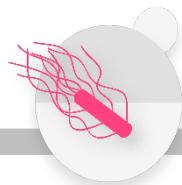
**cameo**



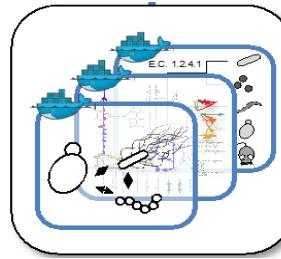
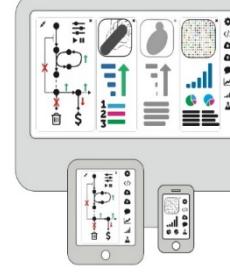


# Exploitation model

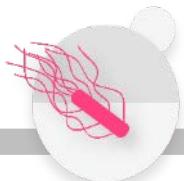




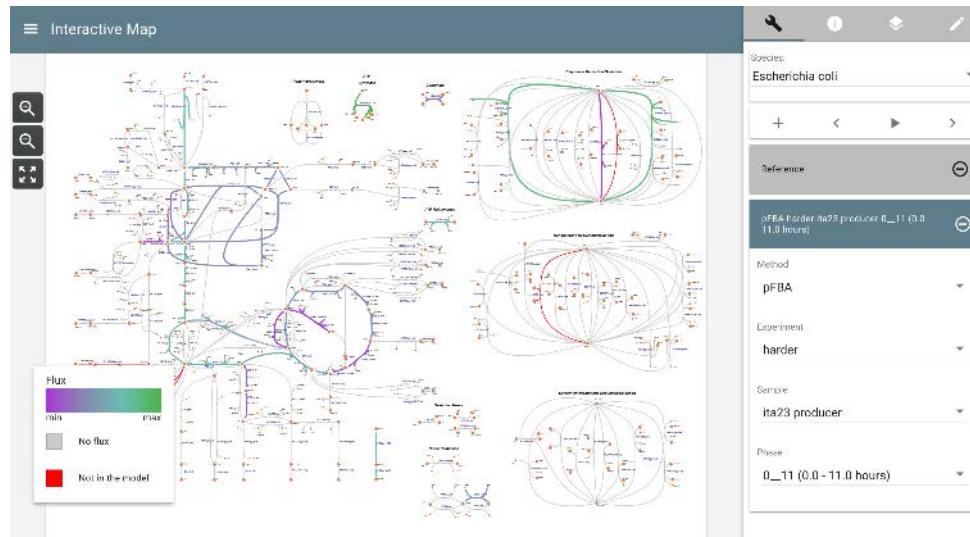
# How to benefit from the project?

<b>Standalone tools and databases</b>    	<b>Public APIs</b> 	<b>Integrated platform and user interface</b> 	<b>Customized software and consulting</b>   
---	--	--	--

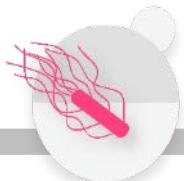
User category	What DD-DeCaF can provide
Academic and industrial users	Software libraries, APIs, standalone tools, integrated platform
Biotechnology end-user SMEs	Customized software solutions and/or standard integrated platform
Large collaborative projects	Platform for data/model management, sharing, analysis, modeling and design
Bioinformatics SMEs	Libraries, tools and APIs for building innovative end-user solutions
Large companies	Customized software based on cutting edge modeling and design methods



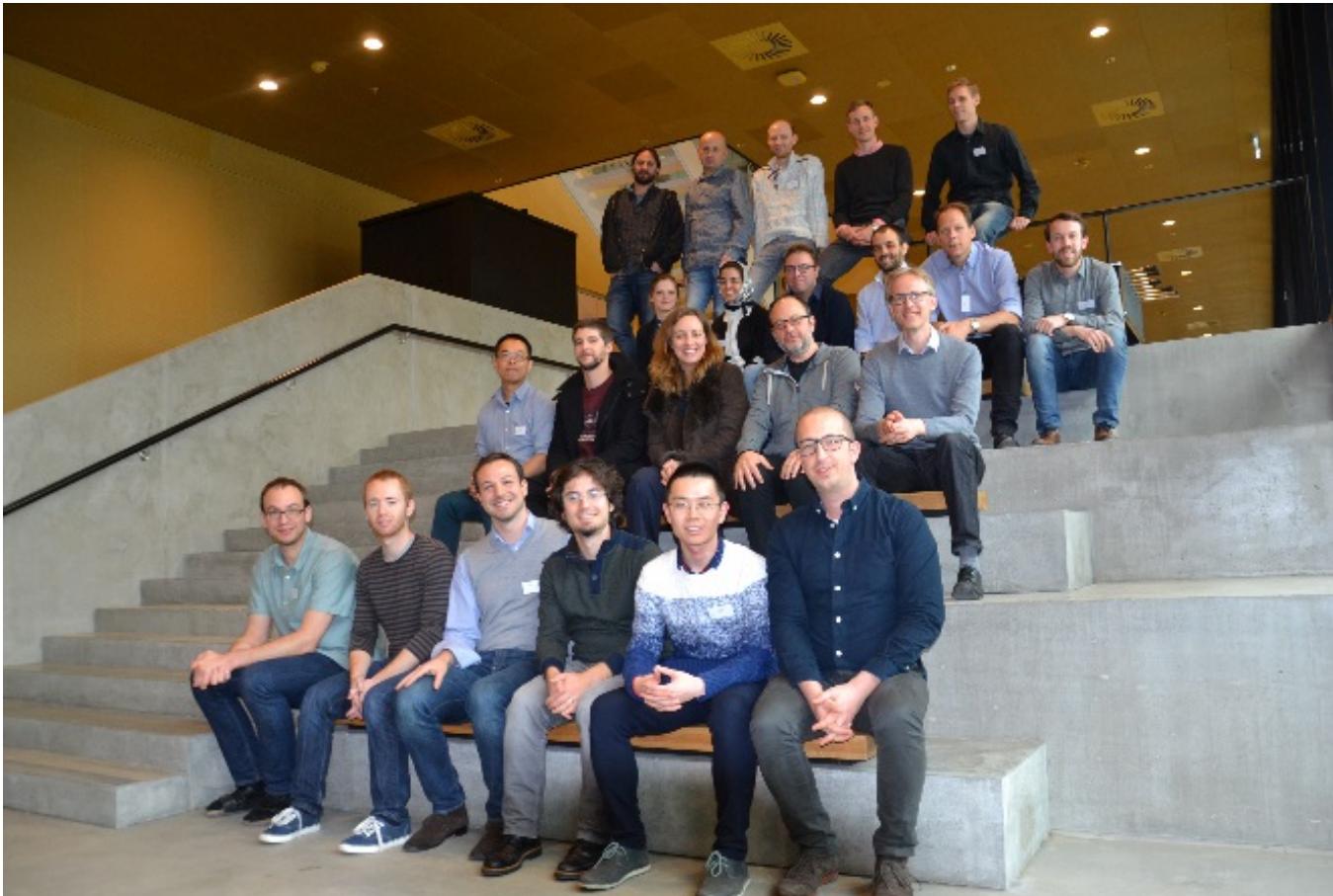
# Following the DD-DeCaF project



**Website:** <http://dd-decaf.eu/>  
**App:** <https://app.dd-decaf.eu>  
**Twitter:** @dddecaf  
**Youtube channel:**  
[https://www.youtube.com/channel/UCg\\_EU0w-OI2YhMm94wUq-Jw](https://www.youtube.com/channel/UCg_EU0w-OI2YhMm94wUq-Jw)



# Acknowledgements



Niko Sonnenschein  
DTU  
Scientific & Technical  
Manager



Simao Soares  
SilicoLife  
Exploitation &  
Dissemination  
Manager



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 686070