



KubeCon



CloudNativeCon

Europe 2021

# Mining large data sets of biomedical Omics Data made easy with managed Kubernetes!

*Dr. Theodore Alexandrov - Team Leader EMBL  
Doug Davis - PM Code Engine*

Virtual



# European Molecular Biology Laboratory (EMBL)



*Virtual*



- International research organization
- Focused on life sciences
- 27 member states
- Sites in Heidelberg, Cambridge, Rome, Barcelona, Grenoble, Hamburg

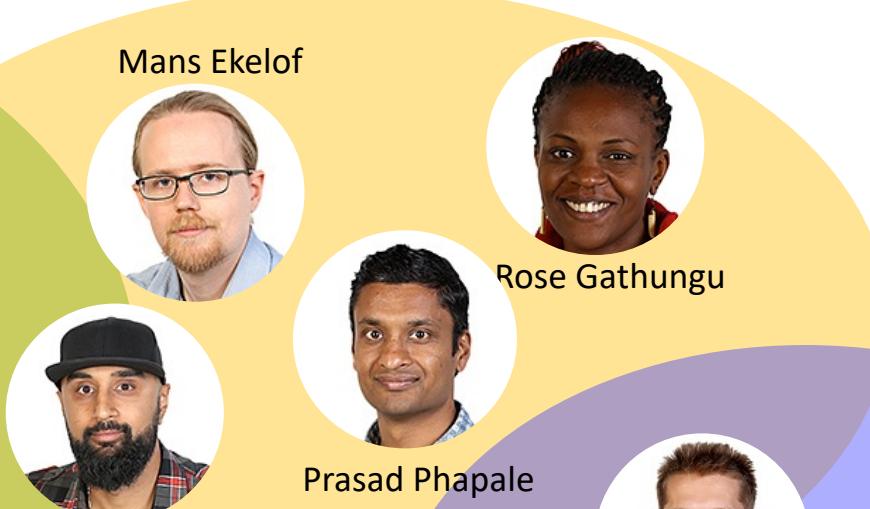


# Alexandrov Team at EMBL

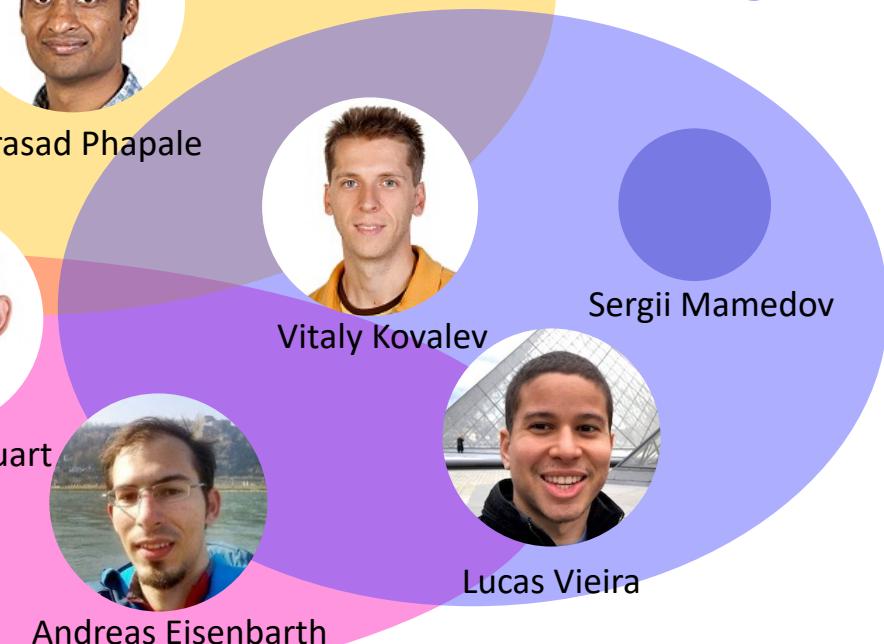
Biology



Chemistry



Software  
engineering



Computer science

# Spatial metabolomics technology



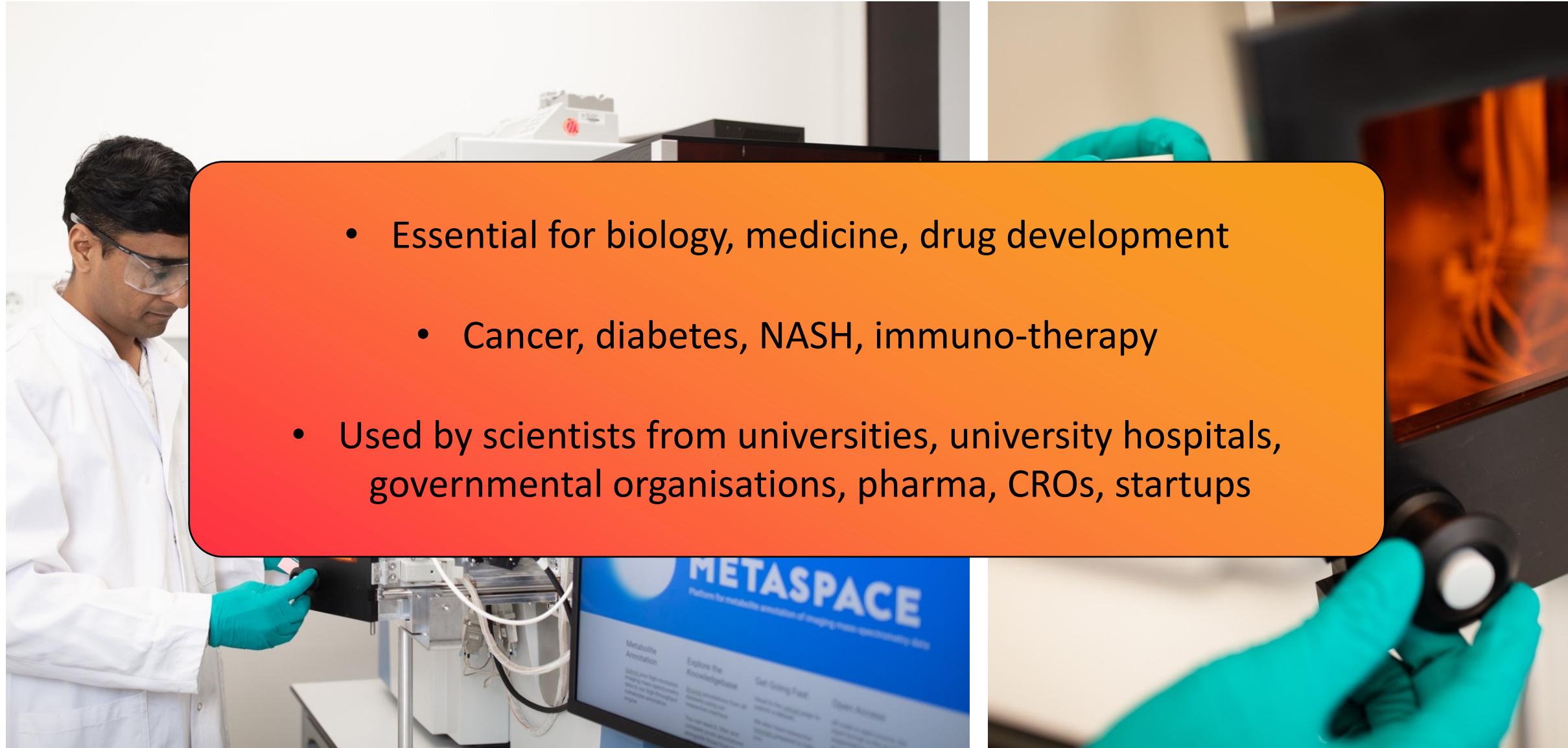
KubeCon



CloudNativeCon

Virtual

Europe 2021



- Essential for biology, medicine, drug development
  - Cancer, diabetes, NASH, immuno-therapy
- Used by scientists from universities, university hospitals, governmental organisations, pharma, CROs, startups

# Spatial metabolomics data



KubeCon

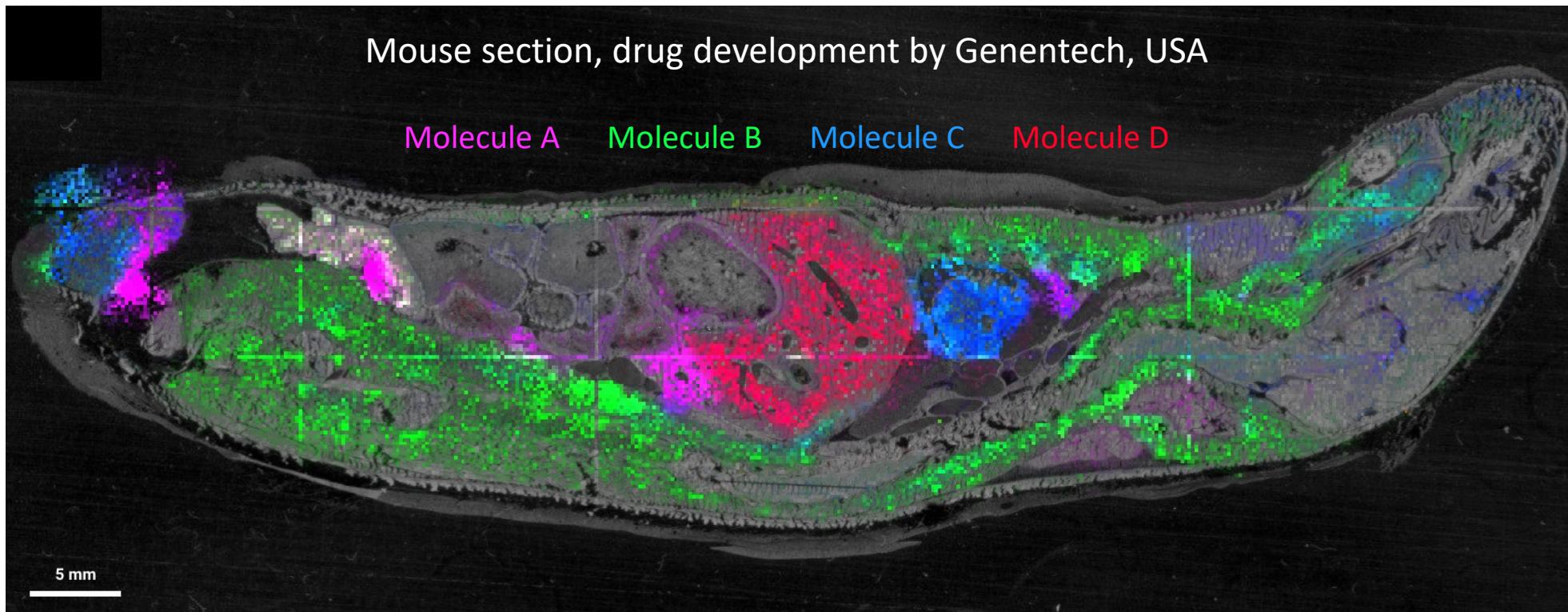


CloudNativeCon

Europe 2021

*Virtual*

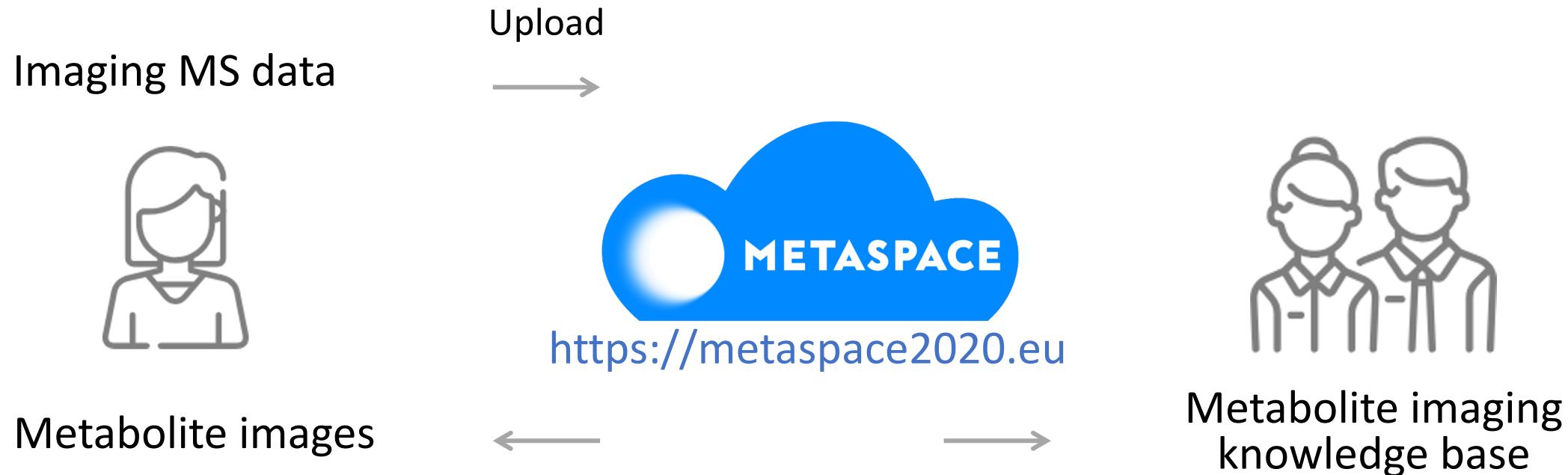
- Hyperspectral image with 100.000+ pixels, 10.000+ channels
- Every channel is associated with a molecule
- 10 GB – 1 TB



# METASPACE: cloud engine & knowledge base



*Virtual*



10.000+ submissions / 800+ users / 100+ labs / 70+ publications



Palmer et al, Nat Methods 2017  
Alexandrov et al, BioRxiv 2019  
Alexandrov, Annu Rev Biomed Data Science 2020



Welcome to

# METASPACE

Platform for metabolite annotation of imaging mass spectrometry data

## Metabolite Annotation

[Submit](#) your high-resolution imaging mass spectrometry data to our high-throughput metabolite annotation engine

## Explore the Knowledgebase

[Browse](#) annotations from all datasets using our interactive interface

You can search, filter and compare your annotations alongside those from the whole imaging mass spectrometry community

## Get Going Fast

Head to the [upload](#) page to submit a dataset.

We also have interactive [tutorials](#) prepared to help you.

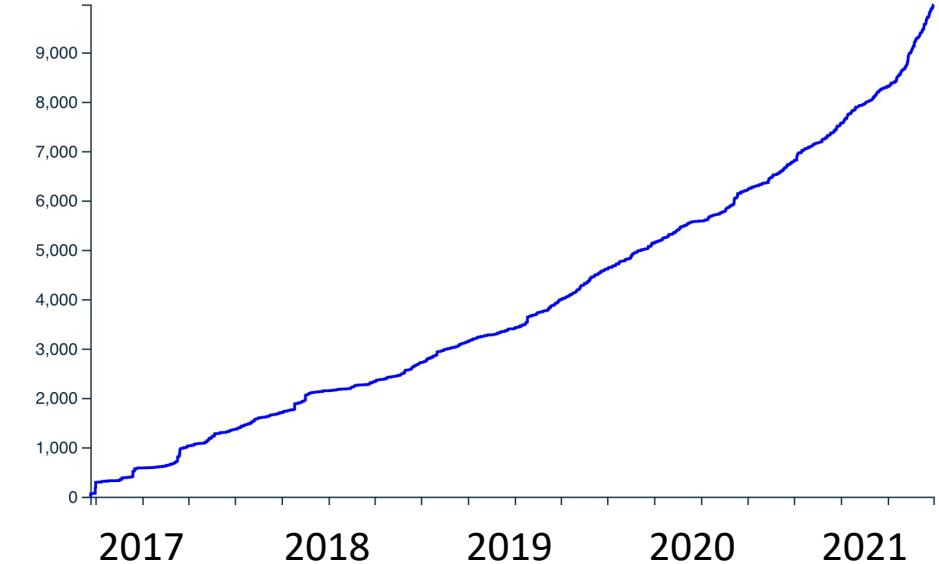
## Open Access

All code is open-source, the input format is the [imzML](#) supported by all mass spec major vendors, the metabolite annotations from the community datasets are public and can be browsed or exported.



# Need for new computing technologies

- **METASPACE uptake and growth**
  - Super-linear growth of submissions
  - Submissions come irregularly
- **Pain points**
  - Deployment delays
  - Infrastructure & queue management
  - resource planning
- **Motivation for new solution**
  - The problem is embarrassingly parallelable



# Serverless as a solution



*Virtual*

## • CloudButton

- European project to develop user-friendly serverless framework, 2019-2022
- IBM and EMBL are partners
- <https://cloudbutton.eu>



Gil Vernik, IBM



Lachlan Stuart, EMBL

## • Lithops

- Serverless computing framework
- Key development in CloudButton
- <https://github.com/lithops-cloud>





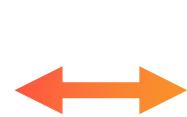
KubeCon



CloudNativeCon

Europe 2021

*Virtual*



**IBM Cloud** Code Engine

# Where are we?

- Lesson Learned
  - Users want to focus on producing value
    - For their customers
    - To get their job done
  - Infrastructure is a means to an end
    - Learning & managing it is an undesirable burden
- State of the Cloud Native community
  - Managing infrastructure is hard
    - E.g. Spark, Kubernetes - great but complex
  - Varying platforms mean varying features and constraints
- Developers should code, not manage infrastructure
  - IBM has a better option...

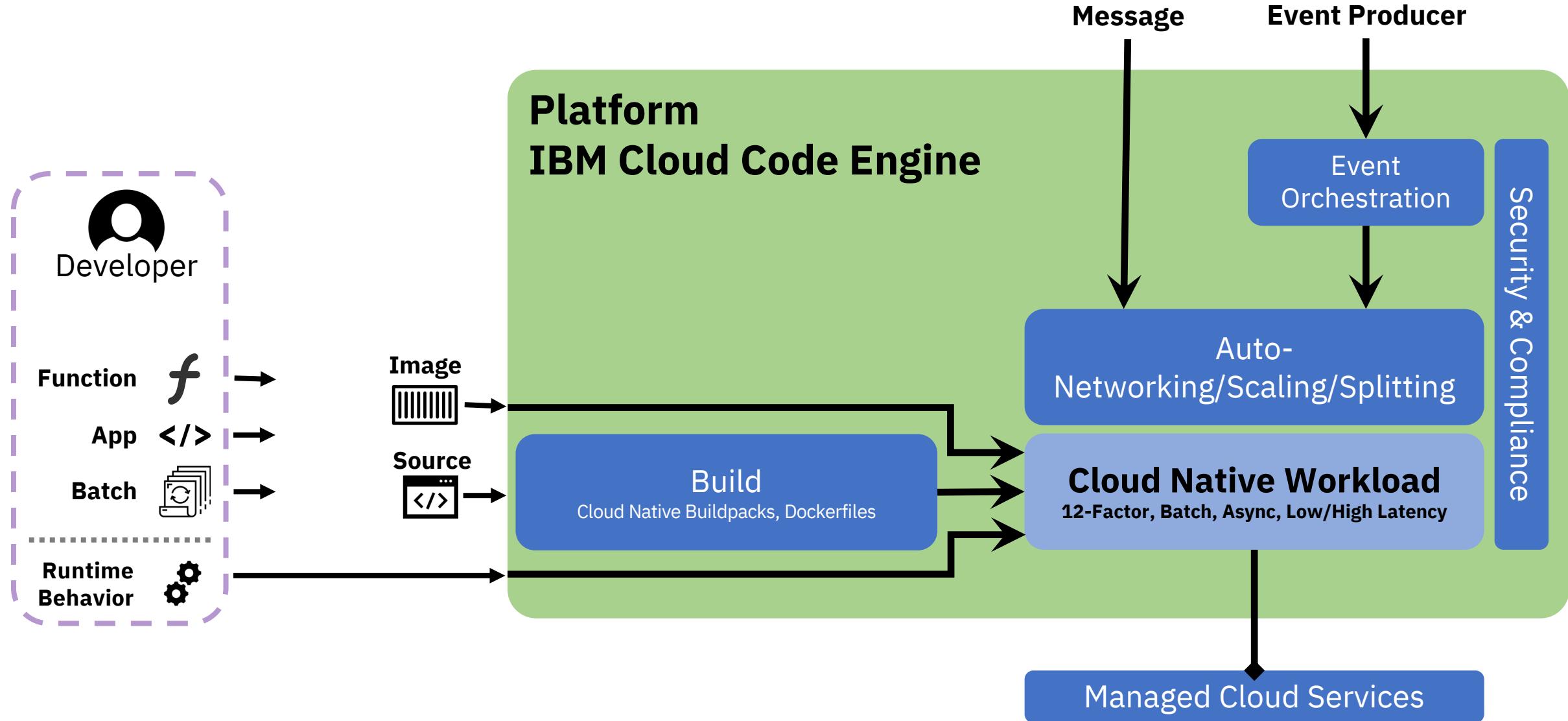


# What if ... ?



CloudNativeCon  
Europe 2021

Virtual



# With IBM Cloud Code Engine, you can!



KubeCon



*Virtual*

- Provide your code and you get:
  - App: Internet exposed, securely hosted workload
    - Scales automatically up and down based on load - even down to zero
    - Zero downtime, blue/green, upgrades of your application
  - Batch: Scaled batch job that scales down when done
  - Connect to managed hosted services
  - **Pay only for when the code is running**
- Need some customizations? Configure:
  - Min, max scale of your apps
  - Resource usage: e.g. memory, CPU, timeouts
  - And more...
- Integrate all your workloads in a private secured network



# IBM Cloud Code Engine

A developer can deploy **any type of application**

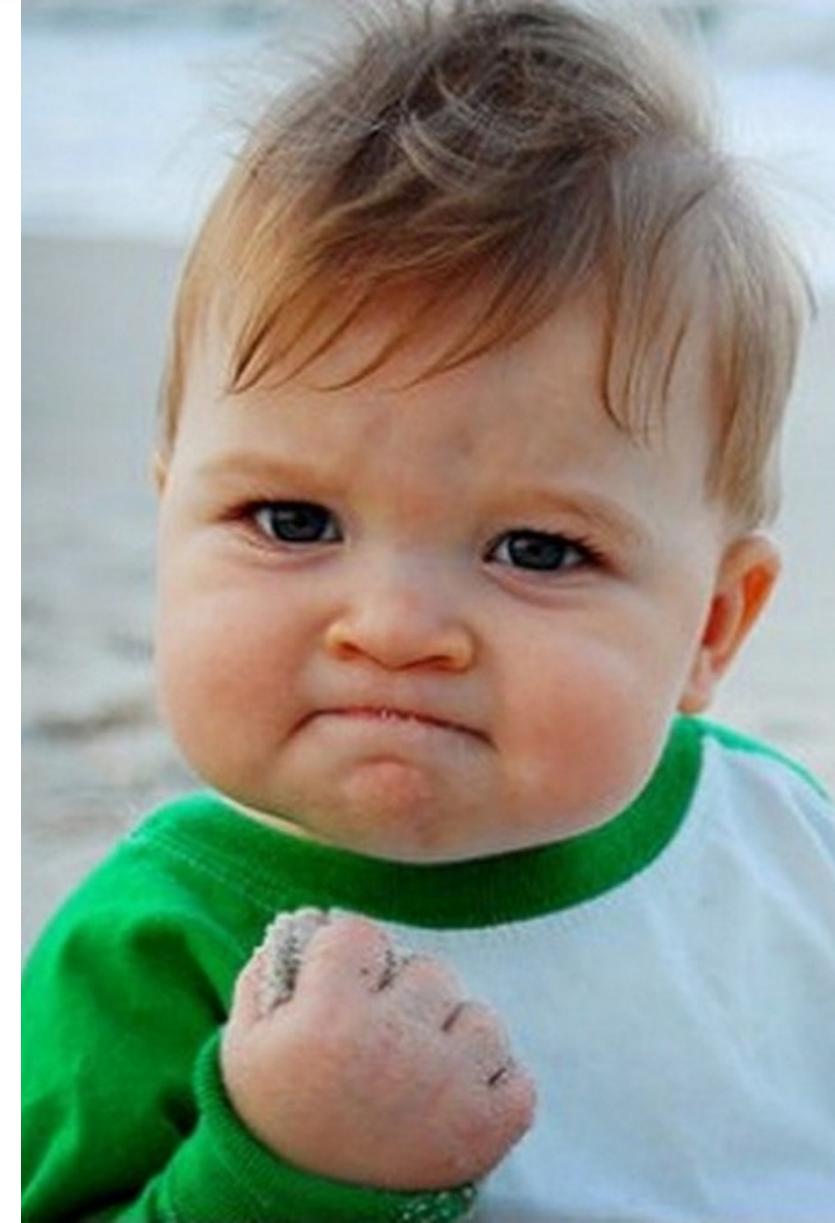
{ container  
batch job  
source code  
function

on a unified platform **without**

{ provisioning  
configuring  
managing  
securing      **any** { clusters  
networks  
VMs  
certificates

and **only pay** when their application is active.

**Run Your Code. Easily. At Scale**  
<https://cloud.ibm.com/codeengine>



# The solution



**IBM Cloud** Code Engine

- Scalable solution for science
- Less infrastructure overhead
- No need for resource planning
- Orchestrates the partitioning and processing of data
- Hides the runtime platform
- Allows dev to focus on business logic
- Native support for serverless
- On demand scaling with a pay-as-you-go model
- No artificial memory/CPU constraints

# Thank You!



*Virtual*

## Q&A

**More questions:**

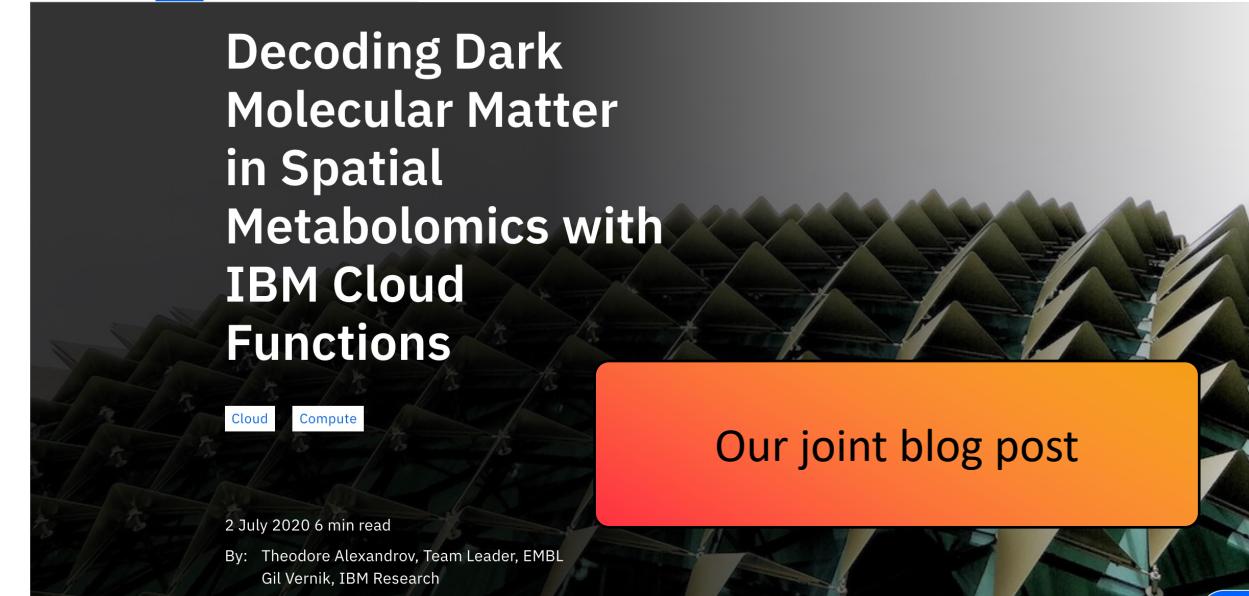
Visit the IBM Booth at KubeCon

Or IBM Cloud Code Engine: <https://cloud.ibm.com/codeengine>

EMBL: Dr. Theodore Alexandrov : [@thalexandrov](https://twitter.com/thalexandrov)

IBM: Doug Davis : [@duginabox](https://twitter.com/duginabox)

IBM | Cloud Products Solutions



The image is a screenshot of an IBM blog post. The title is "Decoding Dark Molecular Matter in Spatial Metabolomics with IBM Cloud Functions". Below the title are two small blue buttons labeled "Cloud" and "Compute". At the bottom left, it says "2 July 2020 6 min read" and "By: Theodore Alexandrov, Team Leader, EMBL Gil Vernik, IBM Research". At the bottom right, there is a red button with the text "Our joint blog post". The background of the post features a dark, abstract geometric pattern.

<https://www.ibm.com/cloud/blog/decoding-dark-molecular-matter-in-spatial-metabolomics-with-ibm-cloud-functions>





KubeCon



CloudNativeCon

Europe 2021

Virtual



Forward Together »