

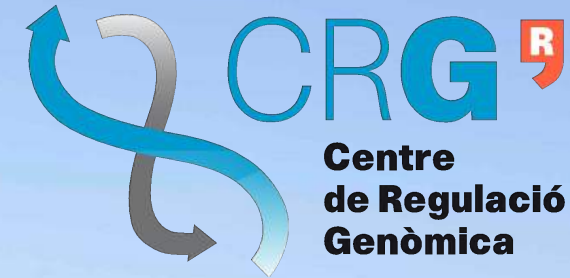
QCloud: A web-based quality control system for the proteomics community

Roger Olivella

CRG-UPF Proteomics Unit

Luca Cozzuto

CRG Bioinformatics Core

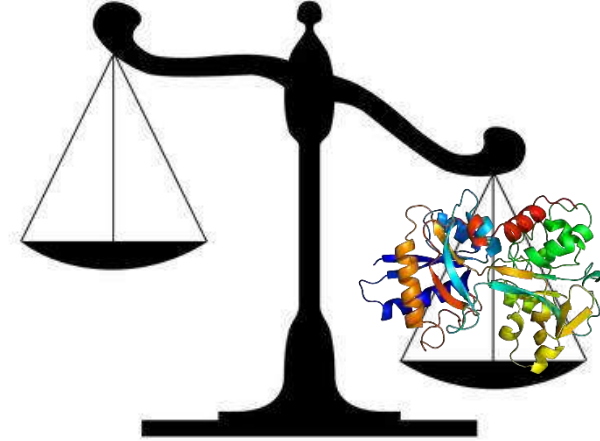


Need for a QC system in Proteomics



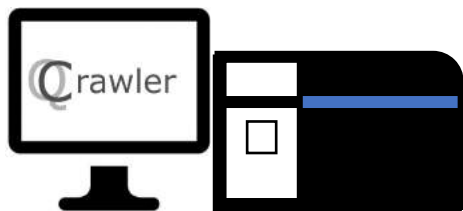
Mass spectrometer

≈



A tool to control and assure
the quality of mass spectrometric acquisition

.Net



QCloud 1.2

Not scalable
Low running time
...but still quite robust

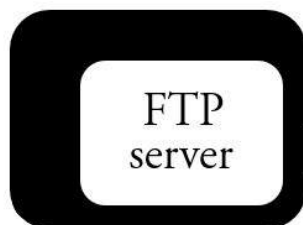
.raw
.wiff



QC1, QC2, QC3

QCrawler

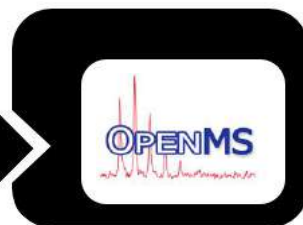
ftp2repository.jar



QConverter



mzml2qcml.jar



qcml2db.jar



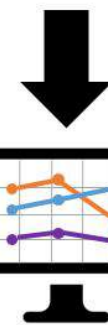
.raw
.wiff
.d

mzML

qcML

LAMP

Java, bash script



Website

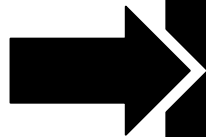
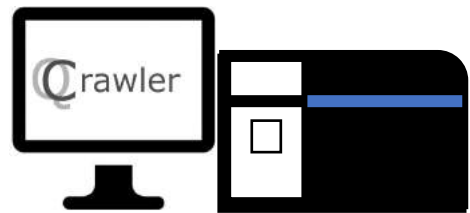
<http://qcloud.crg.eu>

HTML, CSS, JQuery,
Google Charts

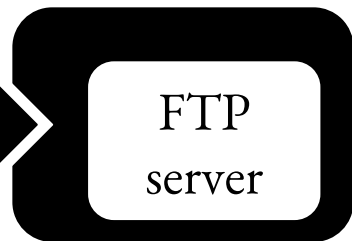
Similar to

MultiQC

.Net



ftp2nf.sh

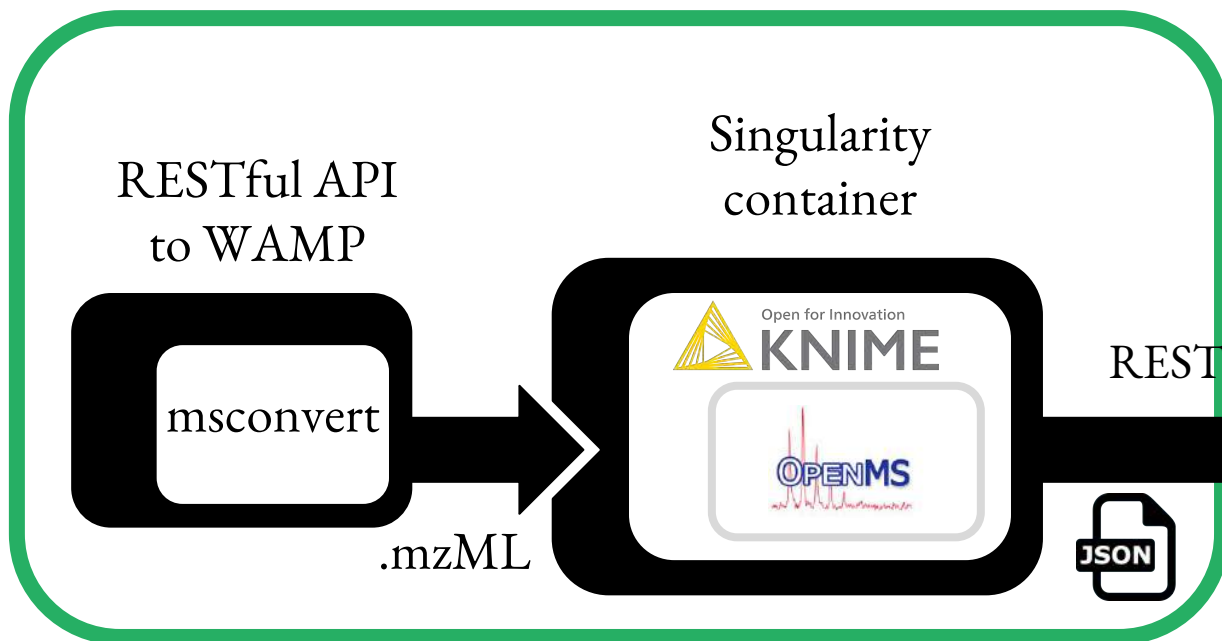


.raw
.wiff

QC1, QC2, QC3



nextflow



RESTful API
to WAMP

msconvert

.mzML

Singularity
container



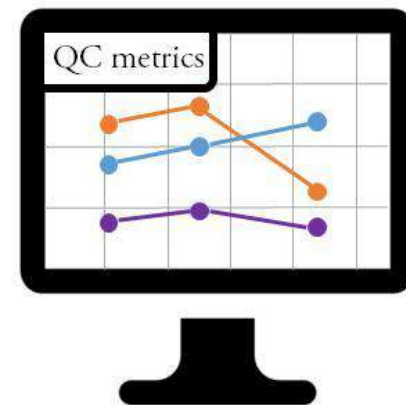
Open for Innovation
KNIME



RESTful API

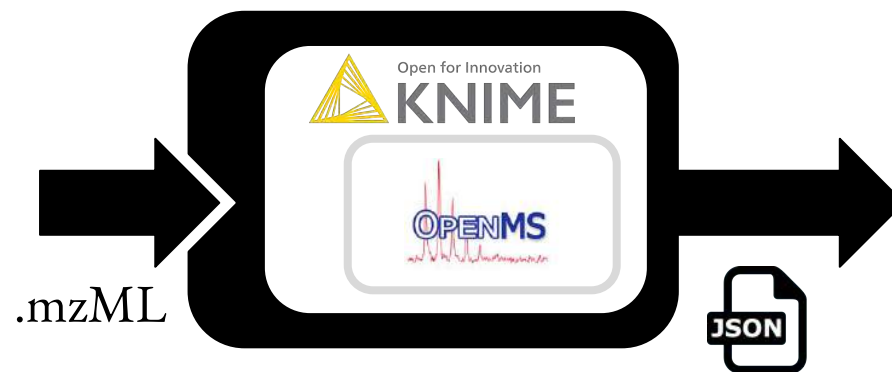


MySQL

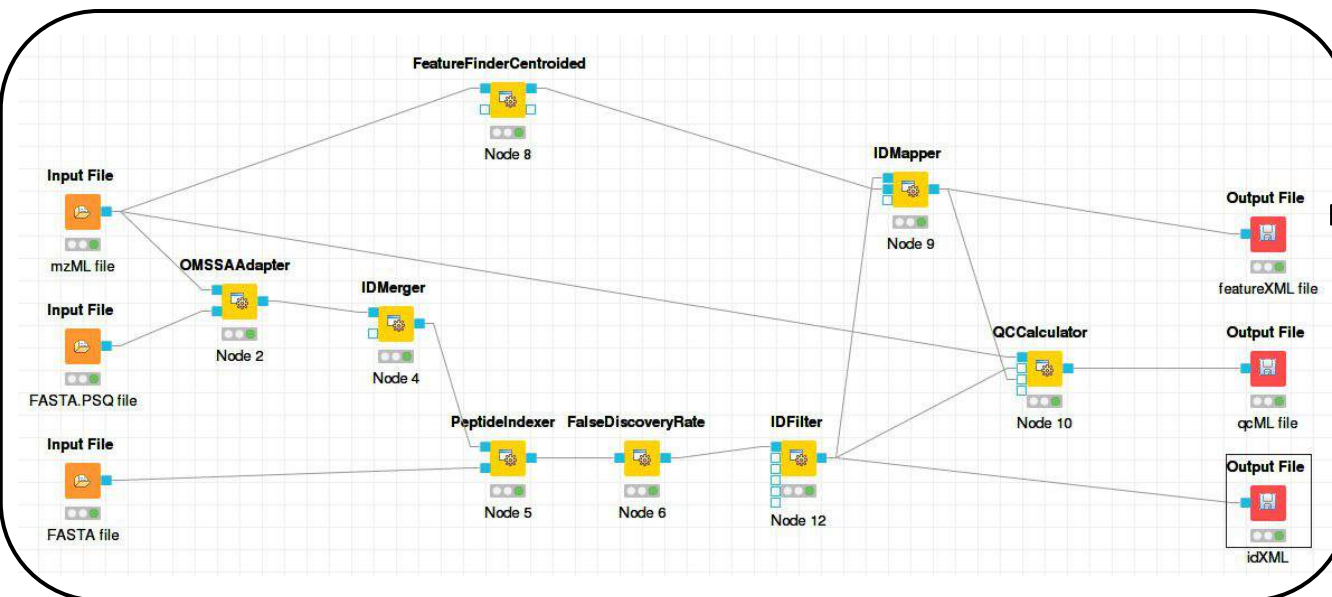


QCloud 2.0

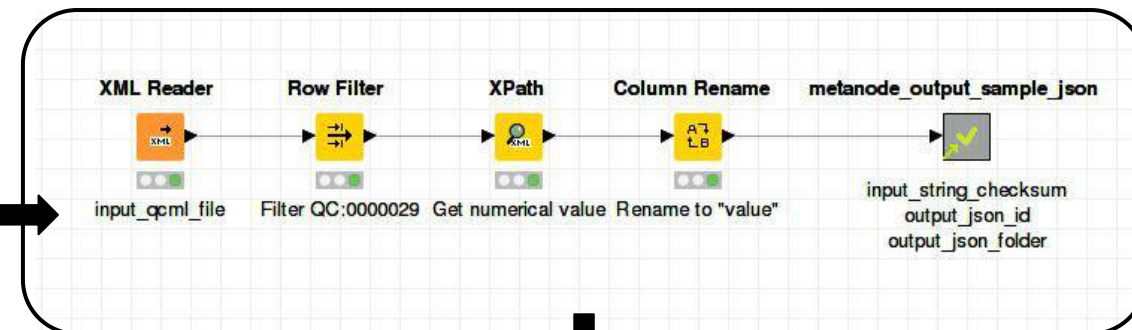
Spring boot
Hibernate
Angular 7
Plotly.js



Open for Innovation
KNIME module_workflow_shotgun.knwf



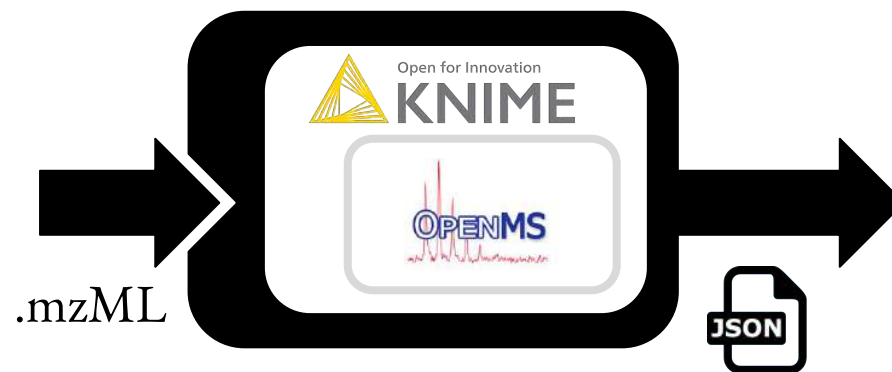
Open for Innovation
KNIME module_parameter_QC_0000029.knwf



```

{
  "file" : {
    "checksum" : "06f2e678e83fb704fc461f17b5566227"
  },
  "data" : [{
    "parameter" : {
      "qCCV" : "QC:9000001"
    },
    "values" : [{
      "value" : "411",
      "contextSource" : "QC:0000029"
    }]
  }]
}
  
```

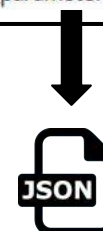
JSON



- ▲ module_workflow_qc4l_cid
- ▲ module_workflow_qc4l_etcid
- ▲ module_workflow_qc4l_ethcd
- ▲ module_workflow_qc4l_hcd
- ▲ module_workflow_shotgun
- ▲ module_workflow_srm

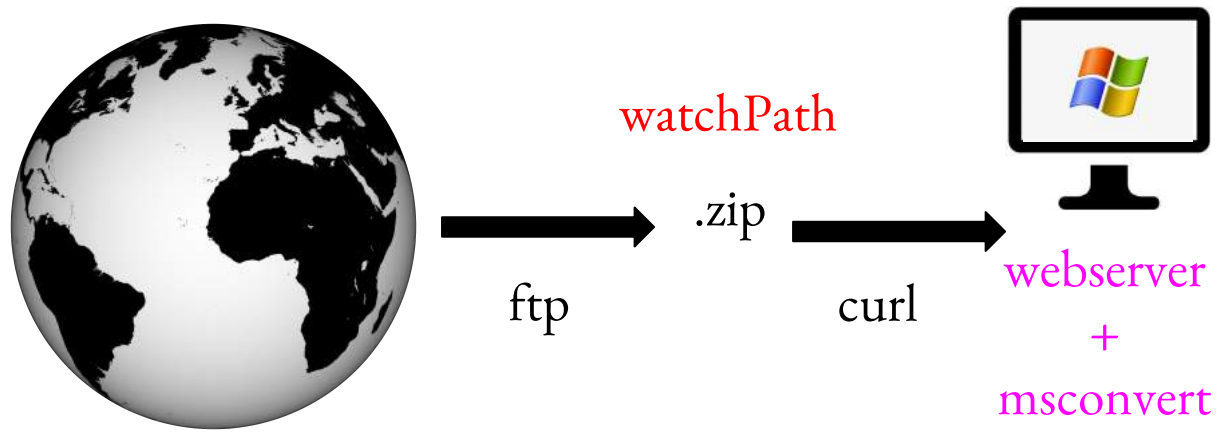


- ▲ module_check_peptides
- ▲ module_parameter_QC_0000007
- ▲ module_parameter_QC_0000029
- ▲ module_parameter_QC_0000031
- ▲ module_parameter_QC_0000032
- ▲ module_parameter_QC_0000048
- ▲ module_parameter_QC_1000014
- ▲ module_parameter_QC_1000285
- ▲ module_parameter_QC_1000927
- ▲ module_parameter_QC_1000928
- ▲ module_parameter_QC_1001844
- ▲ module_parameter_QC_1001844_qc4l
- ▲ module_parameter_QC_1010086



▲ module_api_conn

nextflow





ftp

watchPath

.zip

curl



webserver
+
msconvert

curl

.mzML



NF Processes
wrapping
Knime workflows





watchPath



.zip

ftp



curl



webserver
+
msconvert



curl

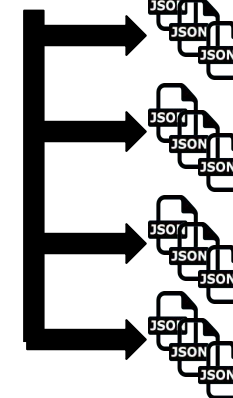
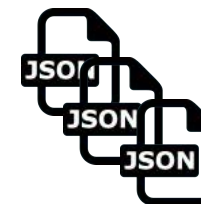
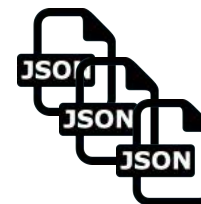
.mzML



NF Processes
wrapping
Knime workflows



3 different QC workflows
2 different db species



.fasta



makeblastdb



storeDir



watchPath



.zip

ftp



curl



webserver
+
msconvert



curl

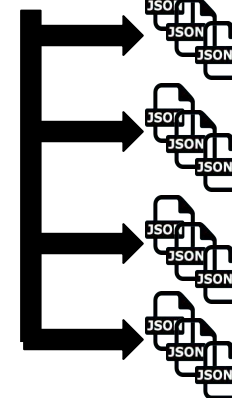
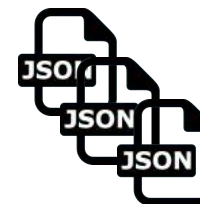
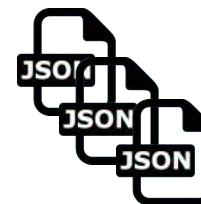
.mzML



NF Processes
wrapping
Knime workflows



3 different QC workflows
2 different db species



errorStrategy = 'ignore'

.fasta



makeblastdb




storeDir

Future developments

- Better communication between the NF pipeline and the web/database
 - Storing configuration / process identifiers within the database
 - Once a failure is detected, send info to the database
- Storing the process ontology information within the DB
- Creating a smarter system for cleaning the **work** directory (currently a cron job is removing older folders)
- Storing the docker image in Docker Hub
- Replacing the Windows version of msconvert with the Linux one
- To run the whole pipeline in a cloud system

Acknowledgements



 roger.olivella@crg.eu
luca.cozzuto@crg.eu

CRG-UPF

Proteomics Unit

Eduard Sabidó

Cristina Chiva

Roger Olivella

Daniel Mancera

Eva Borràs

Guadalupe Espadas

Amanda Solé

Olga Pastor

Julia Morales

Melissa Bradbury

CRG

Bioinformatics Unit

Luca Cozzuto

Toni Hermoso

Julia Ponomarenko

OpenMS Community

Hannes Roest

Lars Nilse

Chris Bielow

IMP-Vienna

Karl Mechtler

VIB-Ghent

Francis Impens

FGCZ-Zurich

Paolo Nanni