SCIENCE MEETS LIFE

MULTI MOT

AN OPEN DATA EXCHANGE ECOSYSTEM FOR CELL MIGRATION DATA

lennart martens

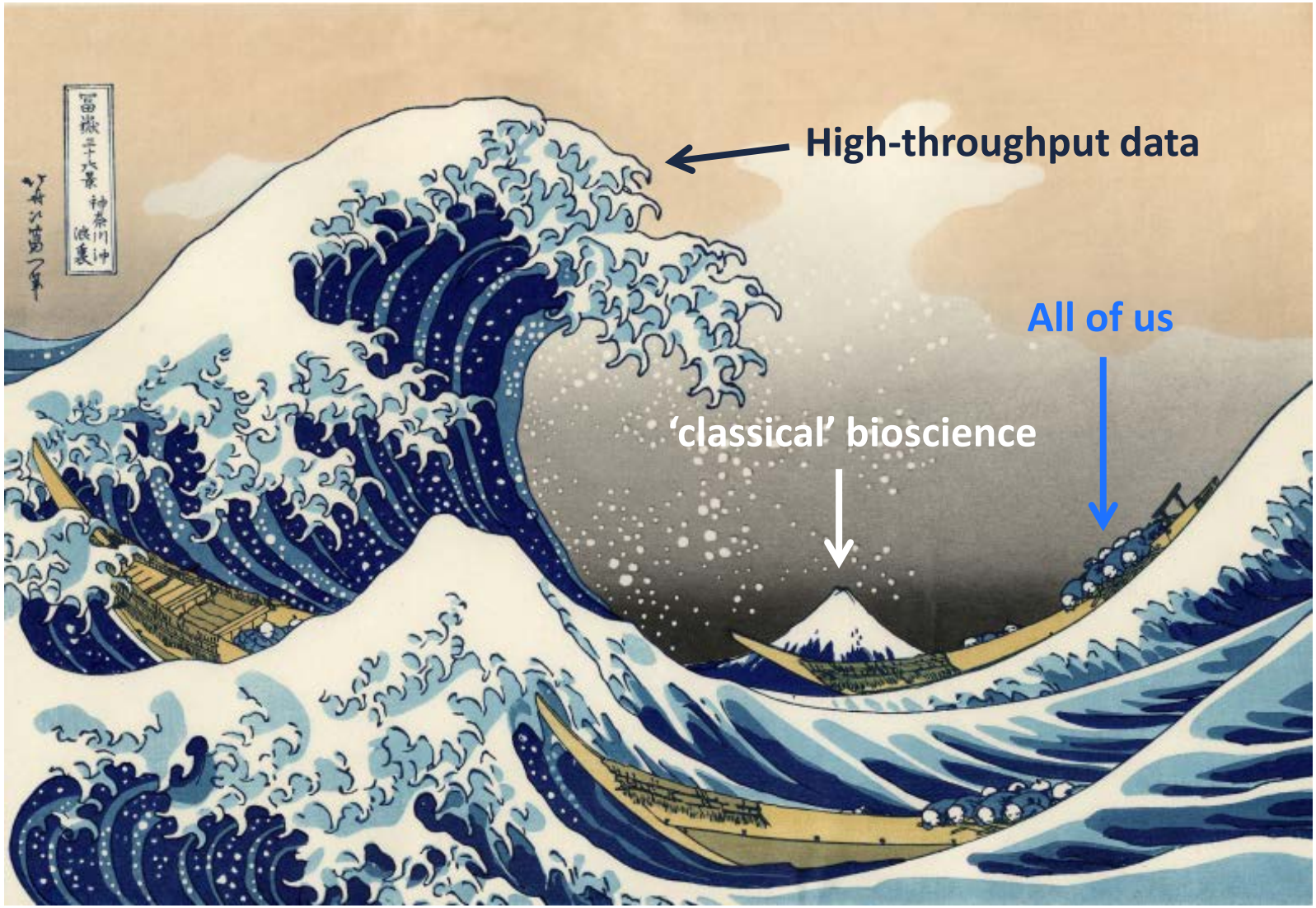
lennart.martens@ugent.be
computational omics and systems biology group
Ghent University and VIB, Ghent, Belgium

FACULTEIT GENEESKUNDE EN GEZONDHEIDSWETENSCHAPPEN

VIB

GHENT UNIVERSITY

Comp omics



"The Great Wave off Kanagawa", by Hokusai, ~1830

Why would we need a data ecosystem?

The origin of the MULTIMOT project

MULTIMOT and CMSO

Being an object of sociological study

CC BY-SA 4.0

Why would we need a data ecosystem?

The origin of the MULTIMOT project

MULTIMOT and CMSO

Being an object of sociological study

CC BY-SA 4.0

Data availability is frequently required by an important third party

Where are the data?

***Nature Biotechnology* now requires data availability statements to be supplied with research papers.**

As the research community embraces data sharing, academic journals can do their bit to help. Starting this month, all research papers published in *Nature Biotechnology*, *Nature* and 11 other Nature titles will include information on whether and how others can access the underlying data.

These statements will report the availability of the 'minimal data set' necessary to interpret, replicate and build on the findings reported in the paper. Where applicable, they will include details about publicly archived data sets that have been analyzed or generated during the study. Where restrictions on access are in place—for example, in the case of privacy limitations or third-party control—authors will be expected to make this clear.

The new policy (http://go.nature.com/2bf4vqn) builds on our long-standing support for data availability as a condition of publication. It also extends our support for data citation, the practice of citing data sets in reference lists via digital object identifiers (DOIs).

CC BY-SA 4.0

But data availability is only the start of a data exchange ecosystem

Data sharing makes the work accessible to all of us

Data sharing allows us to build much more efficiently on previous work

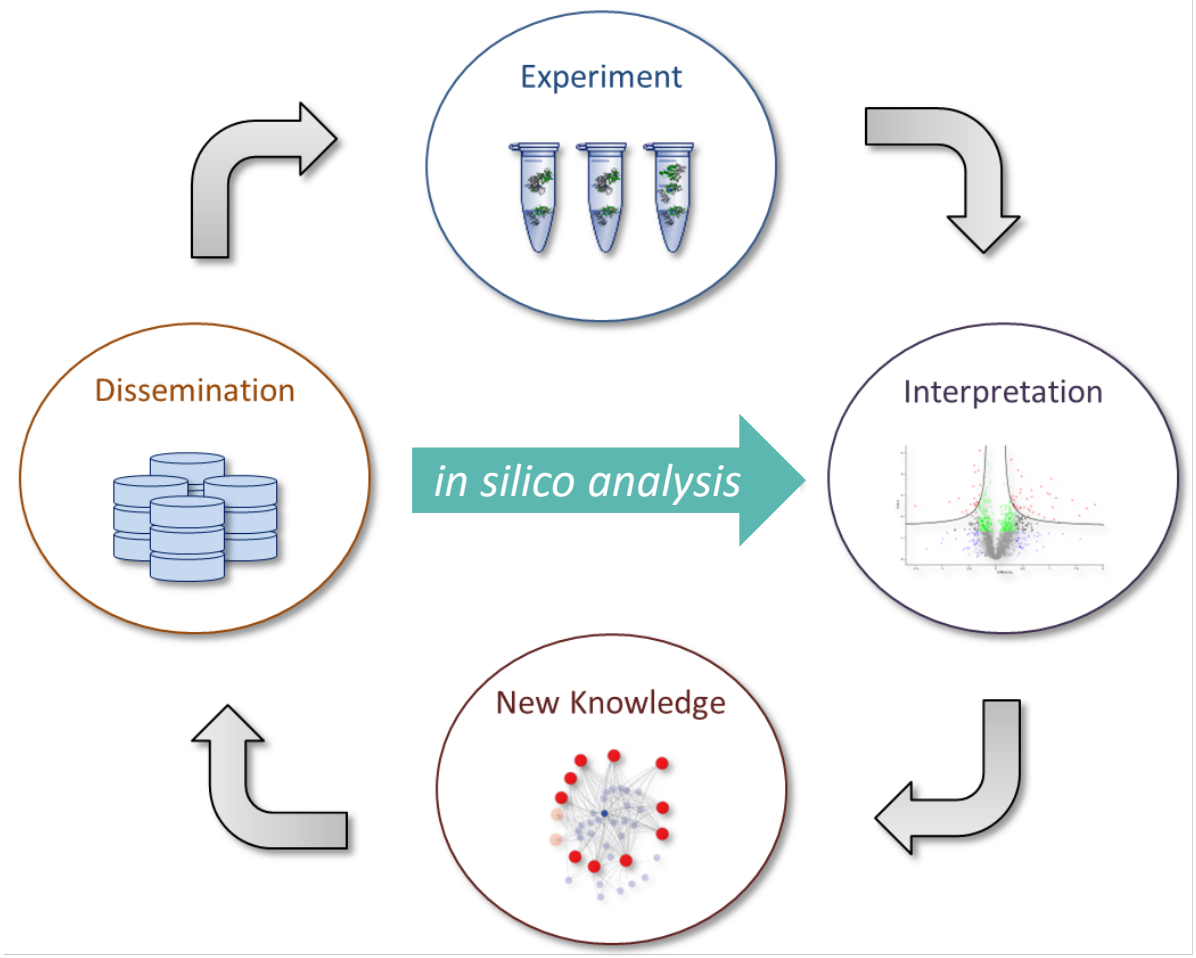
Data sharing helps maximize the usefulness of each individual research effort

Data tend to have a (much!) longer shelf life than our (limited) interpretations

Data sharing fosters creativity, and stimulates revolutionary research

CC BY-SA 4.0

A mature data exchange ecosystem provides many means for data re-use



Vaudel, Proteomics, 2016

CC BY-SA 4.0

Why would we need a data ecosystem?

The origin of the MULTIMOT project

MULTIMOT and CMSO

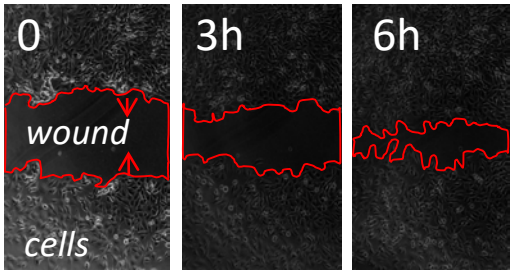
Being an object of sociological study

CC BY-SA 4.0

We developed CellMissy, to manage data from cell migration experiments

Experiment Manager

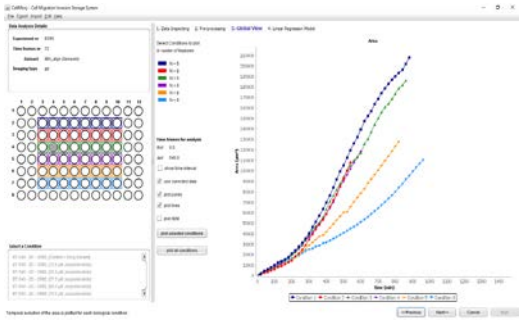
Experiment



Data Loader

Data Analyzer

Collective cell migration



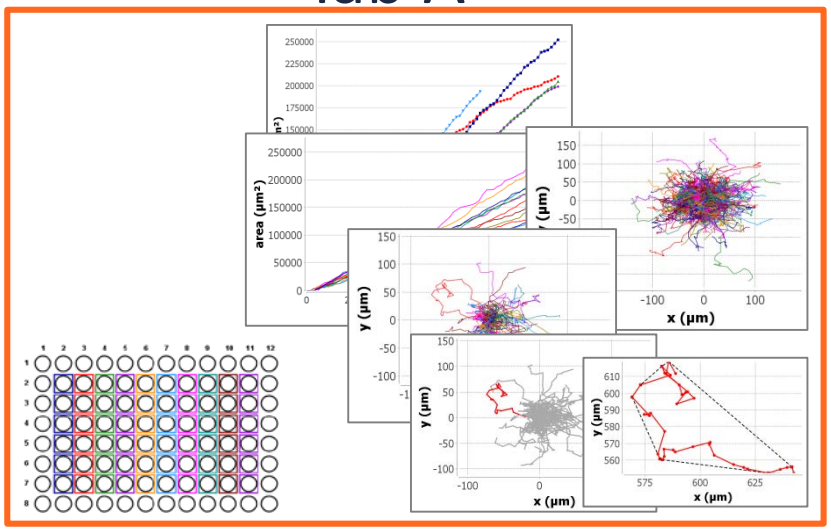
Single-cell migration

Masuzzo, Bioinformatics, 2013 and Masuzzo, Scientific Reports, 2017

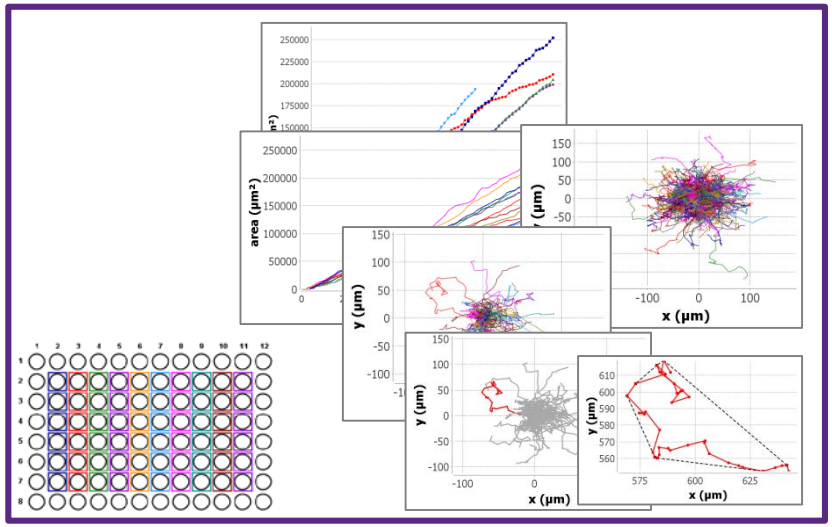
CC BY-SA 4.0

CellMissy can also export an experiment so it can be shared with other researchers

lab A



lab B

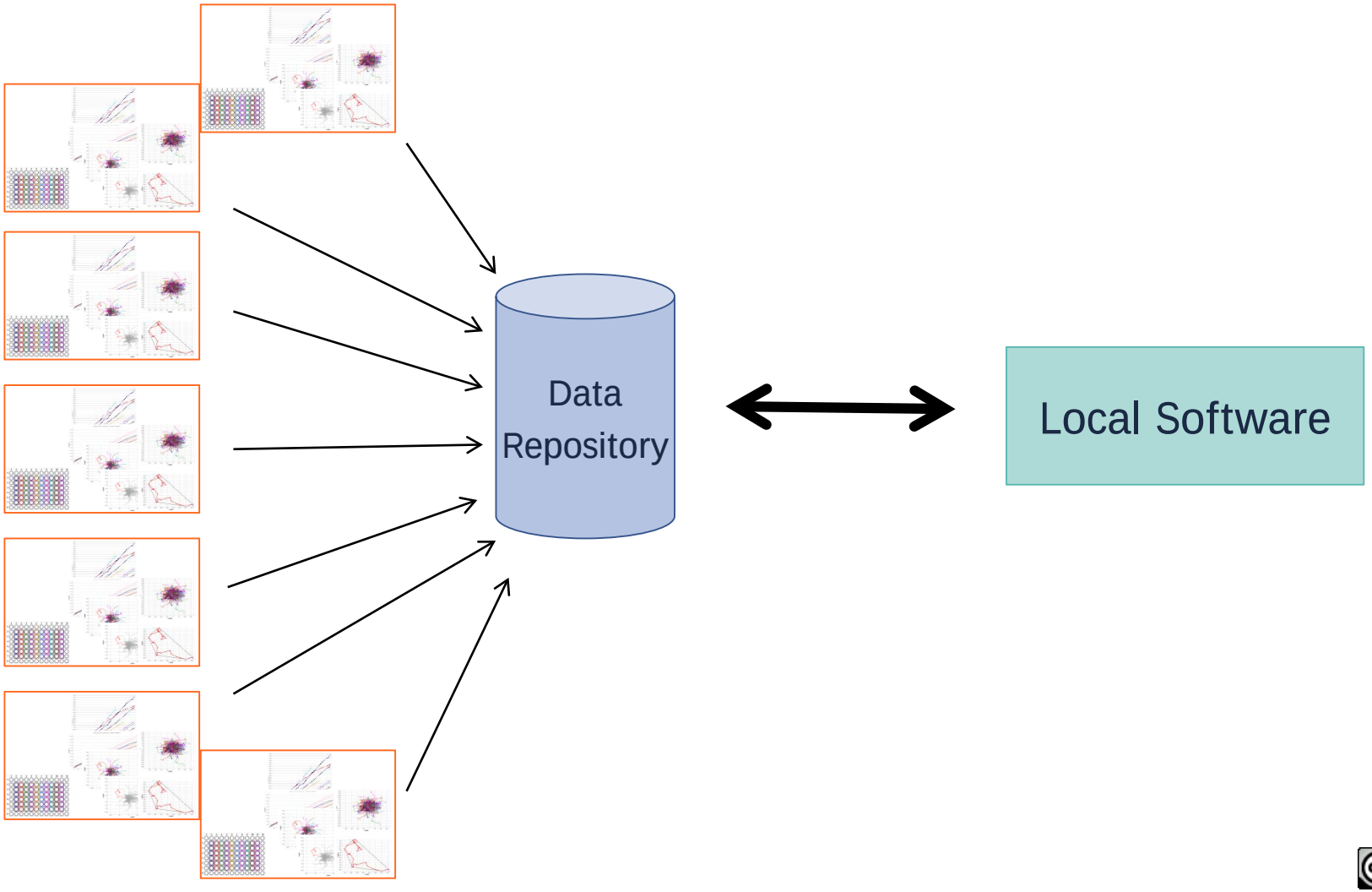


This is one file in CellMissy! (≈10 MB)

Masuzzo, Bioinformatics, 2013 and Masuzzo, Scientific Reports, 2017

CC BY-SA 4.0

With this in hand, it was only a small step to envision a true data sharing ecosystem



CC BY-SA 4.0

So we called a meeting in Ghent in 2014 to discuss this idea in more depth

Forum

CellPress

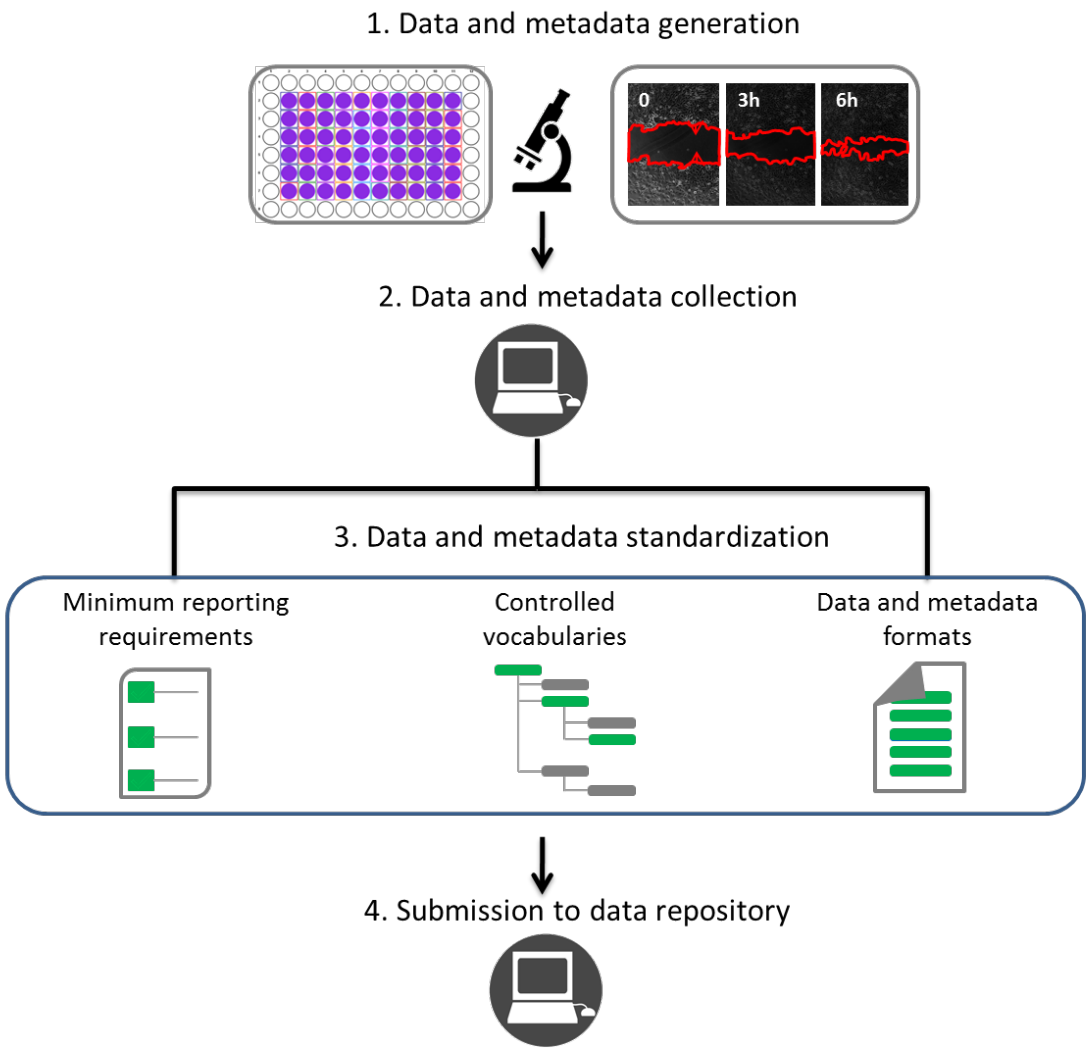
An open data ecosystem for cell migration research

Paola Masuzzo[1,2], Lennart Martens[1,2], and The 2014 Cell Migration Workshop Participants[3]

Cell migration workshop, Ghent, March 2014; Masuzzo, Trends in Cell Biology, 2015

CC BY-SA 4.0

Which led to a vision for an open data exchange ecosystem for cell migration data

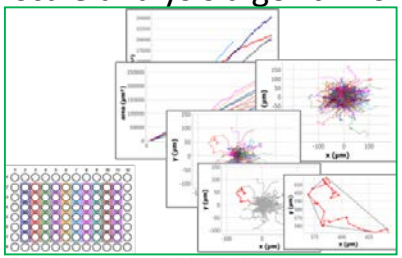


Masuzzo, Trends in Cell Biology, 2015

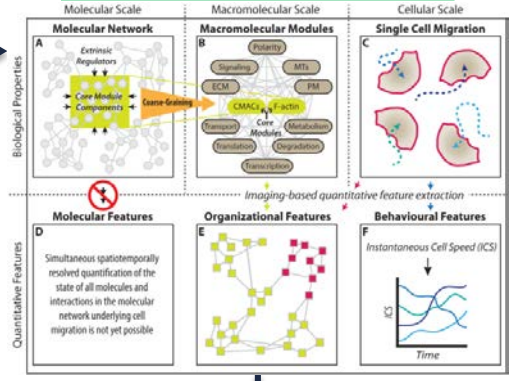
CC BY-SA 4.0

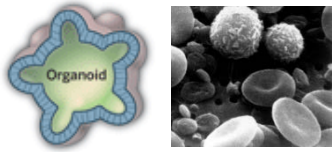
From the start, we included data re-use as a key feature of such an ecosystem

6. Multiscale and meta-scale analysis algorithms



5. Retrieval /dissemination from data repository





7. Application to proof-of-concept studies

Masuzzo, Trends in Cell Biology, 2014; Lock, PLoS ONE 2014

CC BY-SA 4.0

Why would we need a data ecosystem?

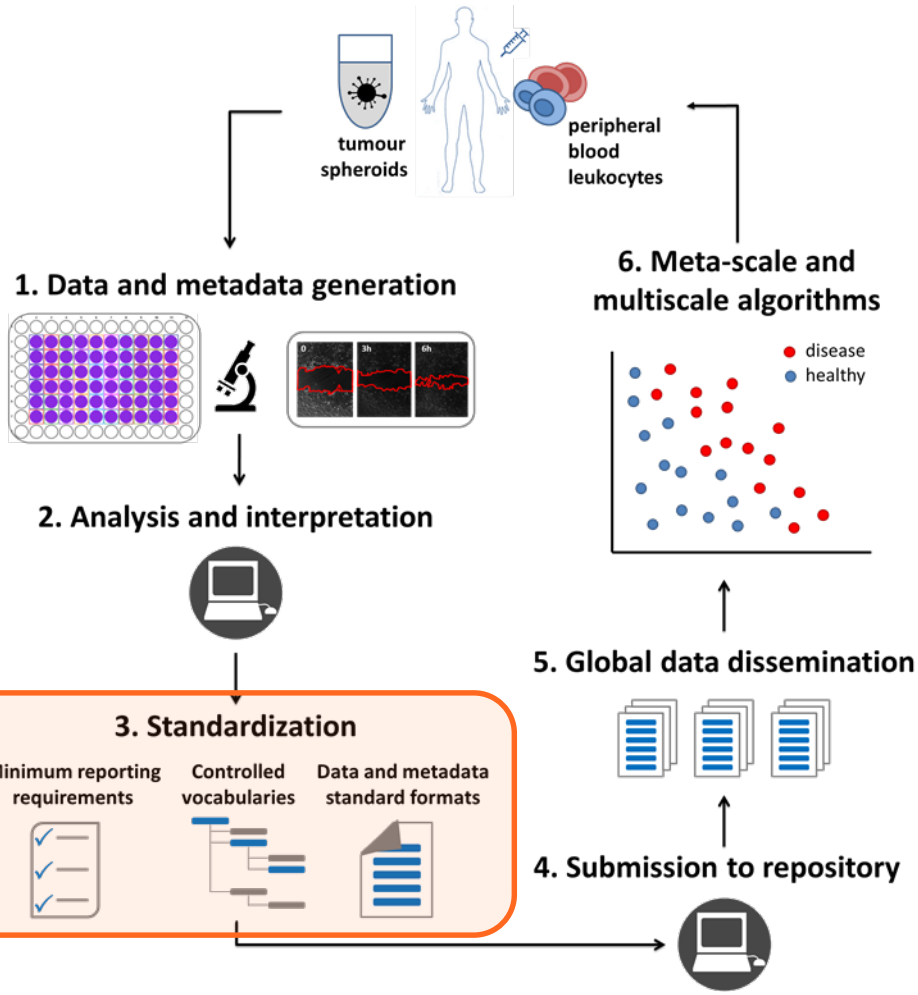
The origin of the MULTIMOT project

MULTIMOT and CMSO

Being an object of sociological study

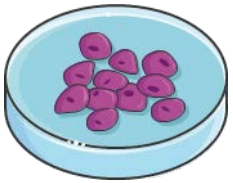
CC BY-SA 4.0

The H2020 MULTIMOT project provided the means to start implementing this vision



CC BY-SA 4.0

The various steps in a cell migration experiment deliver a lot of data in many distinct formats



sample preparation

- paper laboratory notebooks
- electronic laboratory notebooks
- spreadsheets
- text files
- protocols
- papers...

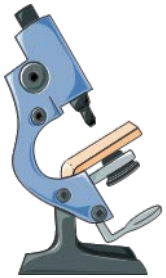


image acquisition

- raw files
- XML files
- proprietary microscope or acquisition software files → ND2 for Nikon, LIF for Leica, OIB or OIF for Olympus, LSM or ZVI for Zeiss

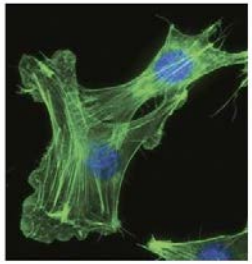
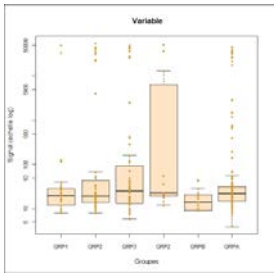


image processing

- image files with pixel values and metadata
- png, jpeg, tiff, avi
- text files describing processing algorithms
- text files describing extracted features



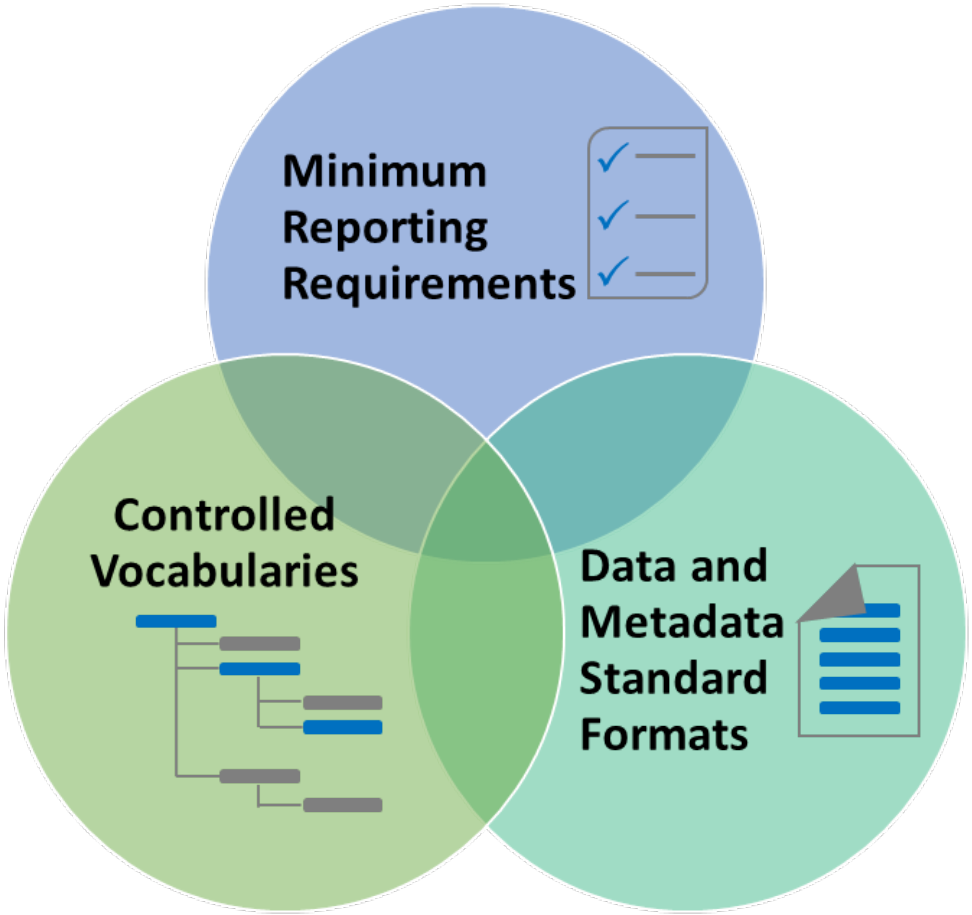
data analysis

- graphs, plots
- analysis pipelines
- text files describing computational algorithms...

Servier Medical Art, CC-BY 3.0; Cell Image Library, CC-BY 3.0

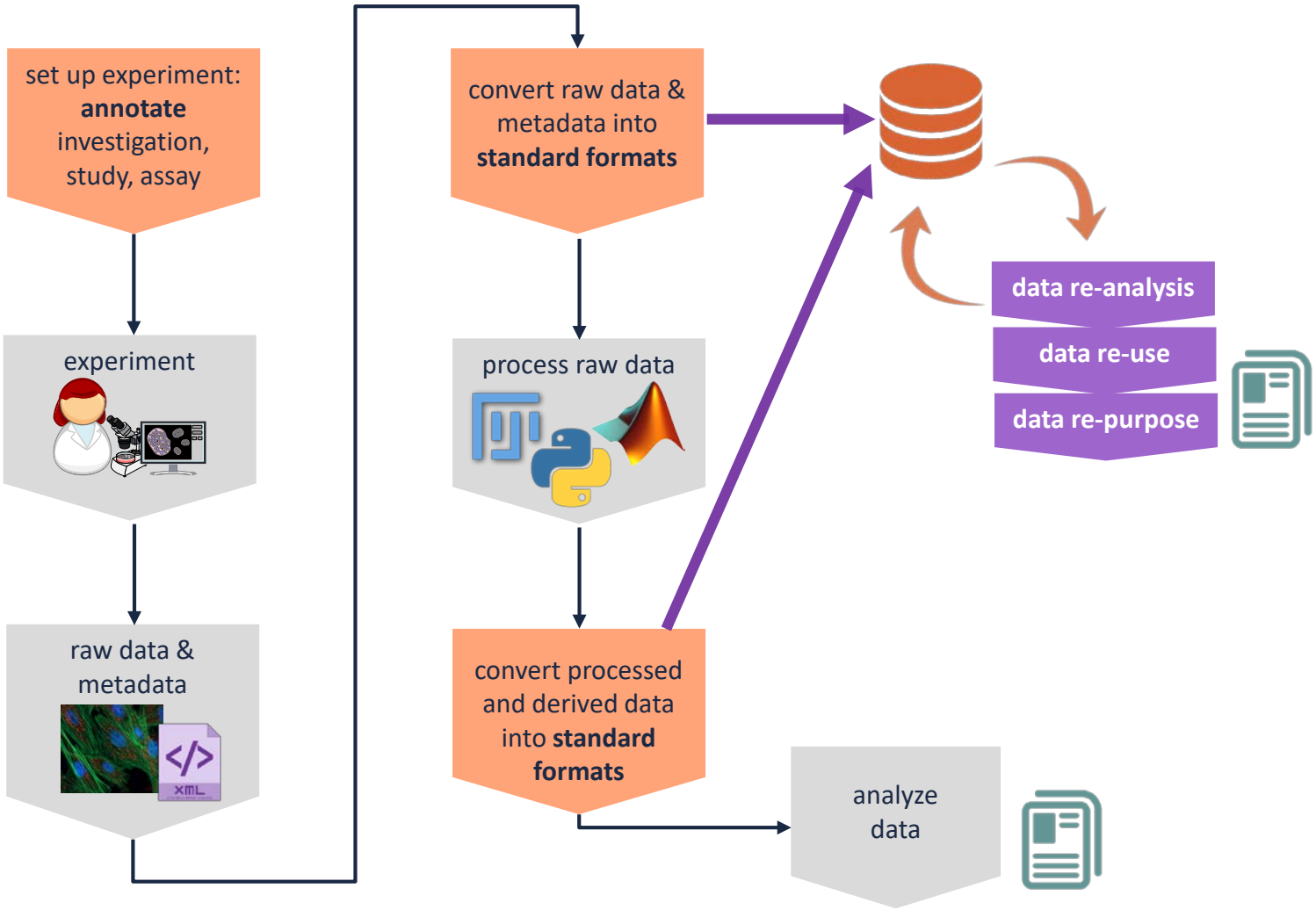
CC BY-SA 4.0

Data reporting requires 3 building blocks: minimal requirements, CVs, and formats



CC BY-SA 4.0

The goal of standardization is to maximize data re-use



CC BY-SA 4.0

A lot of standards are already available for the life sciences

672 Standards

Terminology Artifact	345
Model/Format	213
Reporting Guideline	114

840 Databases

Protein	167
Genome	129
DNA	86

93 Policies

Funder	20
Journal	67
Society	2

https://biosharing.org/

CC BY-SA 4.0

Common formats allow data to be accessed, and cell migration can build upon what exists

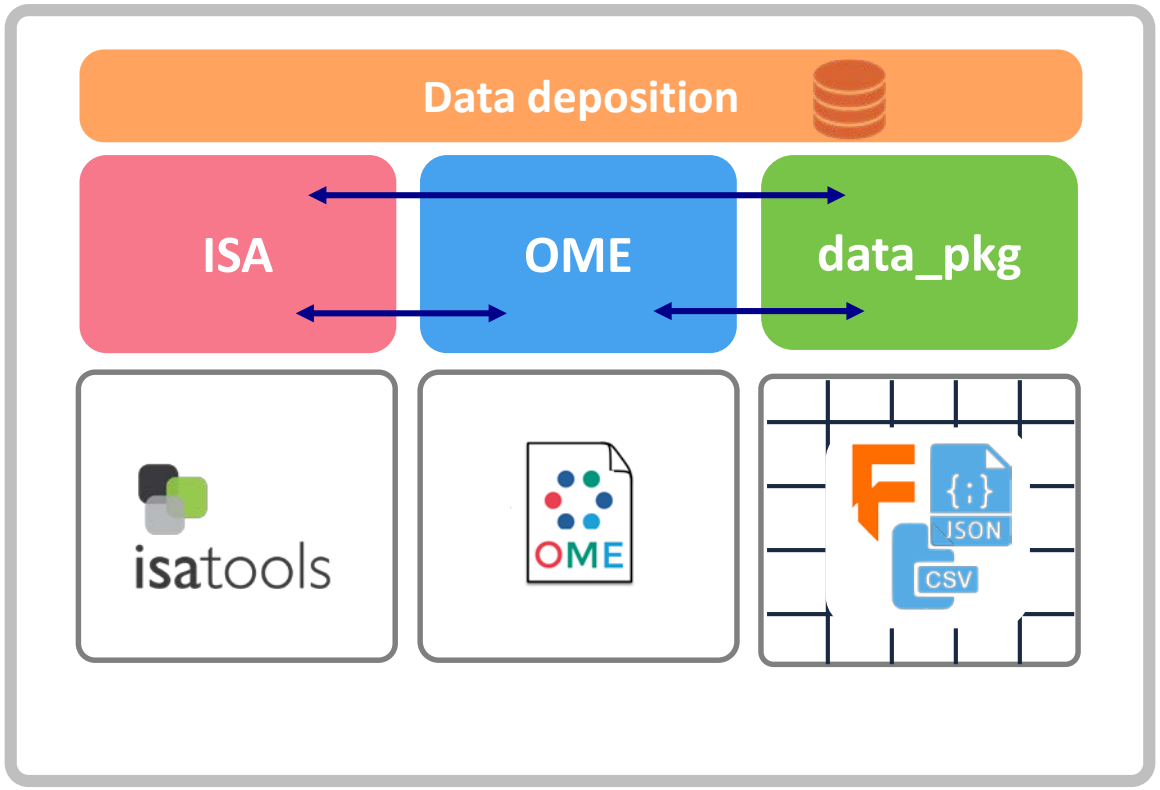


Image adapted from Josh Moore, CMSO workshop 2017

CC BY-SA 4.0

Minimum reporting requirements and CVs improve data verification and accessibility

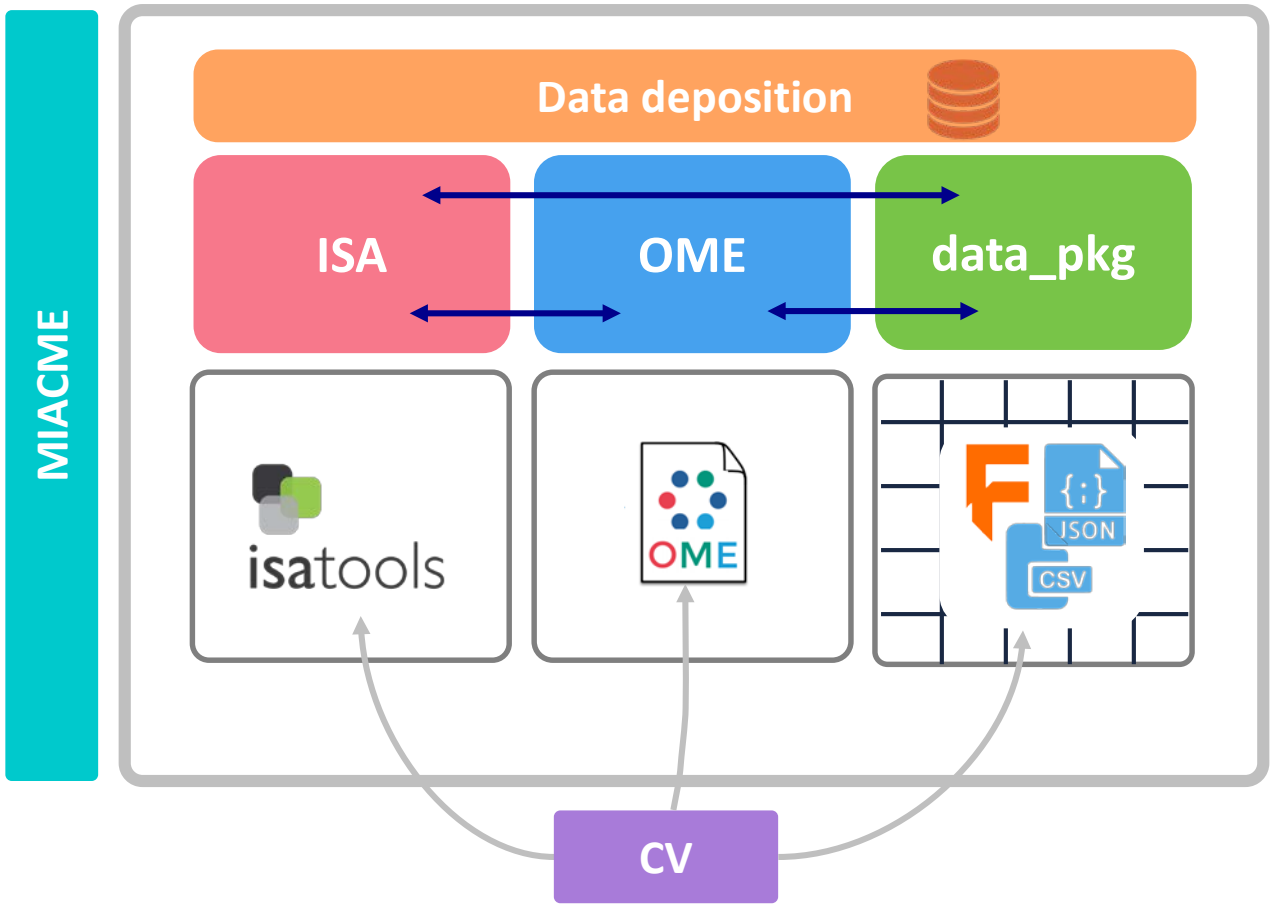


Image adapted from Josh Moore, CMSO workshop 2017

CC BY-SA 4.0

Why would we need a data ecosystem?

The origin of the MULTIMOT project

MULTIMOT and CMSO

Being an object of sociological study

CC BY-SA 4.0

A sociologist's take on our efforts towards (orthogonal) data reuse

"This desire to reactivate data is widespread, and Klie et al. are not alone in wanting to show that 'far from being places where data goes to die' (Klie et al., 2007: 190), **such data collections can be mined for valuable information that could not be obtained in any other way**."

"In attempting to **reactivate sedimented data** in order to enable its re-use, their first step was ..."

"... they are experiments in seeing, in furnishing ways of seeing how data on proteins could become re-usable, could be reactivated as **collective property rather than the by-product of publication**."

Mackenzie and McNally, *Theory, Culture and Society*, 2013

CC BY-SA 4.0

Comp omics

VIB
science meets life

GHENT UNIVERSITY

MULTI MOT

FACULTEIT GENEESKUNDE EN GEZONDHEIDSWETENSCHAPPEN

CC BY-SA 4.0

Participant organisation	PI	Country
VIB (VIB)	Lennart Martens	BE
Weizmann Institute of Science (WIS)	Benjamin Geiger	IL
Ghent University (UG)	Christophe Ampe	BE
Radboud UMC (RUMC)	Peter Friedl	NL
Karolinska Institutet (KI)	Staffan Strömblad	SE
The Chancellor, Masters and Scholars of the University of Oxford (UOXF)	Susanna-Assunta Sansone	UK
Dundee University (DU)	Jason Swedlow	UK
Francis Crick Institute (CRICK)	Erik Sahai	UK
Universität Duisburg-Essen (UDE)	Matthias Gunzer	DE
SME 1 – Idea Biomedical (IB)	Shlomo Turgeman	IL

https://multimot.org/
https://cmso.science/
https://github.com/CellMigStandOrg