benefits of open science

lennart martens

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

Why should we share our work?

What does sharing our work entail?

What can we do once we start sharing our work?

Of sedimentation, opportunity, and dragons

Why should we share our work?

What does sharing our work entail?

What can we do once we start sharing our work?

Of sedimentation, opportunity, and dragons

We usually think we need open science to prevent bad things from happening

- While open science helps prevent some cases of fraud or low quality work being published, it is certainly not a panacea (cfr. peer review)

- Simultaneously, fraud is regularly detected:
  - in the absence of the source data
  - from papers published in closed access journals
  - without any of the code or metadata available

Instead, we should rather focus on the good that comes from open science

- Open science makes the work accessible to anyone
- Open science allows people to build much more efficiently on previous work
- Open science helps maximize the usefulness of each individual research effort
- Data tend to have a (much!) longer shelf life than interpretation
- Open science fosters creativity, and stimulates revolutionary research

Why should we share our work?

What does sharing our work entail?

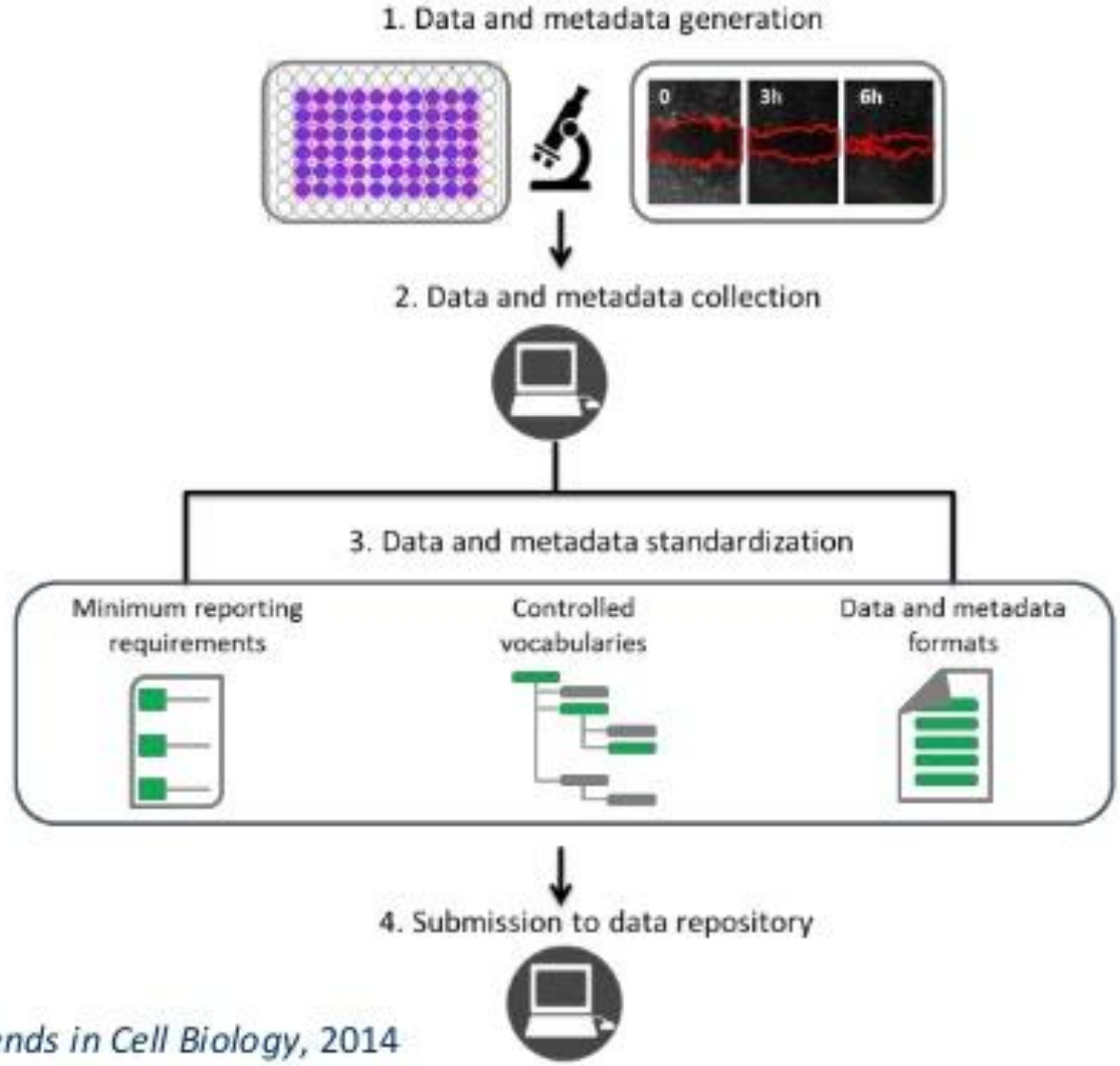
What can we do once we start sharing our work?

Of sedimentation, opportunity, and dragons

We can share our data, our code, our protocols, and our interpretations

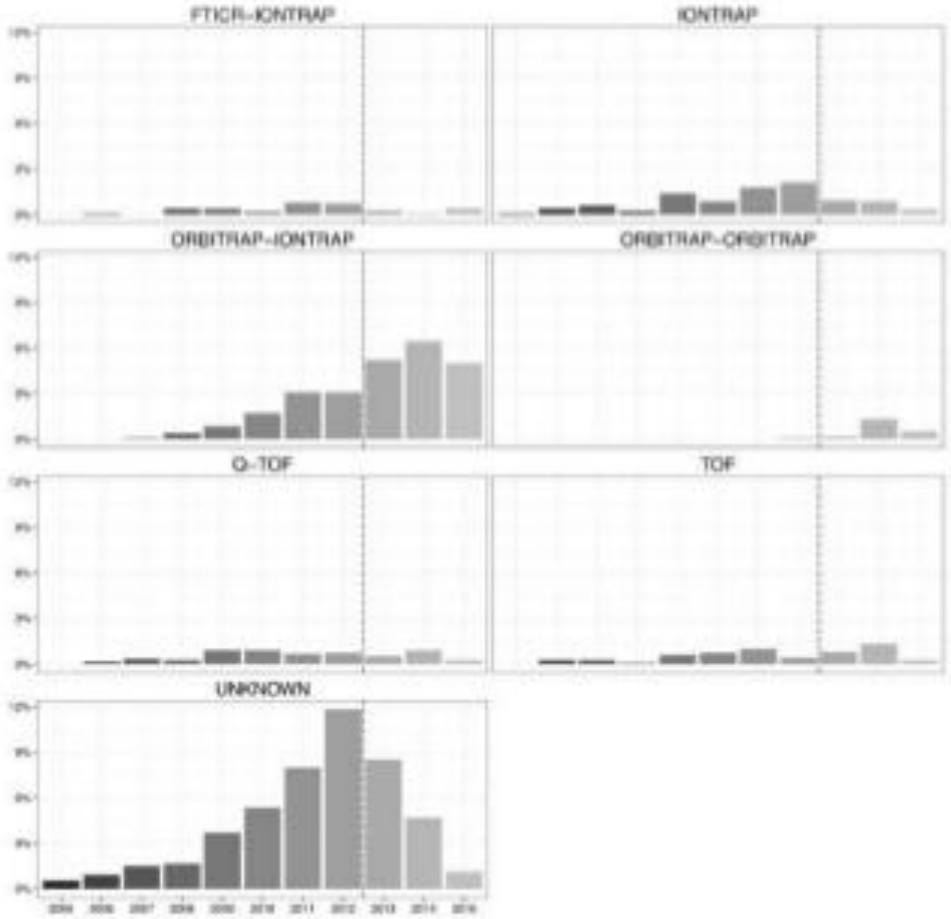
- The data that is obtained should be accompanied by the associated metadata
- The code that is written should be understandable, documented, and hosted at a reliable site
- The protocols should be provided clearly and in full
- The interpretations should be clearly linked to the data (*full provenance*)
- Everything should be licensed in a permissible way

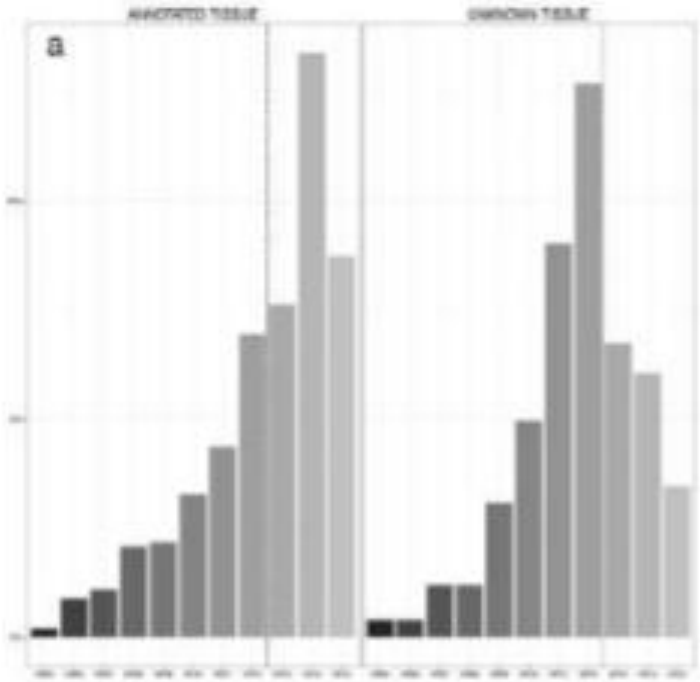
Any open data exchange ecosystem requires standardization



Masuzzo, *Trends in Cell Biology*, 2014

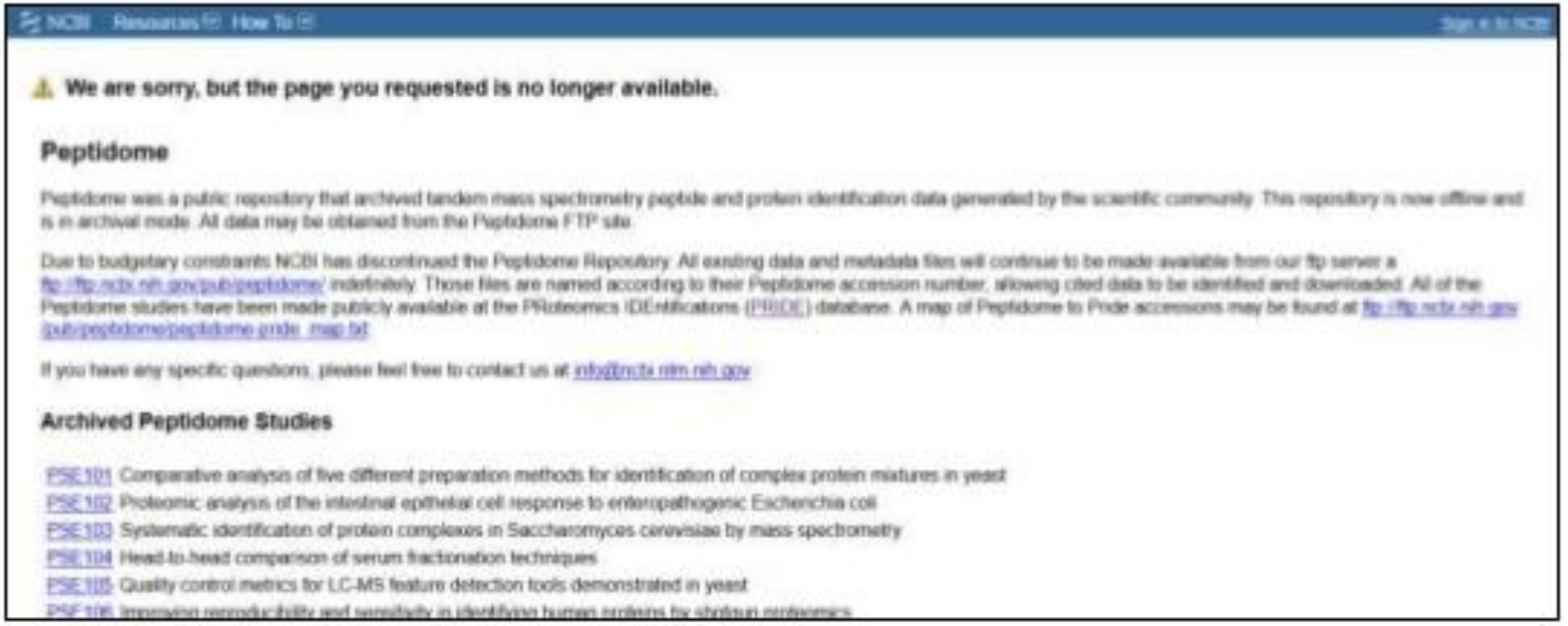
Metadata is often problematic, as it requires a bit of additional effort





Verheggen, *EuPA Open Proteomics*, 2015

Sharing information effectively is not that easy – if you want to do it right!



NCBI Resources How To

We are sorry, but the page you requested is no longer available.

Peptidome

Peptidome was a public repository that archived tandem mass spectrometry peptide and protein identification data generated by the scientific community. This repository is now offline and is in archival mode. All data may be obtained from the Peptidome FTP site.

Due to budgetary constraints NCBI has discontinued the Peptidome Repository. All existing data and metadata files will continue to be made available from our ftp server a ftp://ftp.ncbi.nih.gov/pub/peptidome/ indefinitely. Those files are named according to their Peptidome accession number, allowing cited data to be identified and downloaded. All of the Peptidome studies have been made publicly available at the PRoteomics IDEntifications (PRIDE) database. A map of Peptidome to Pride accessions may be found at ftp://ftp.ncbi.nih.gov/pub/peptidome/peptidome_pride_map.txt

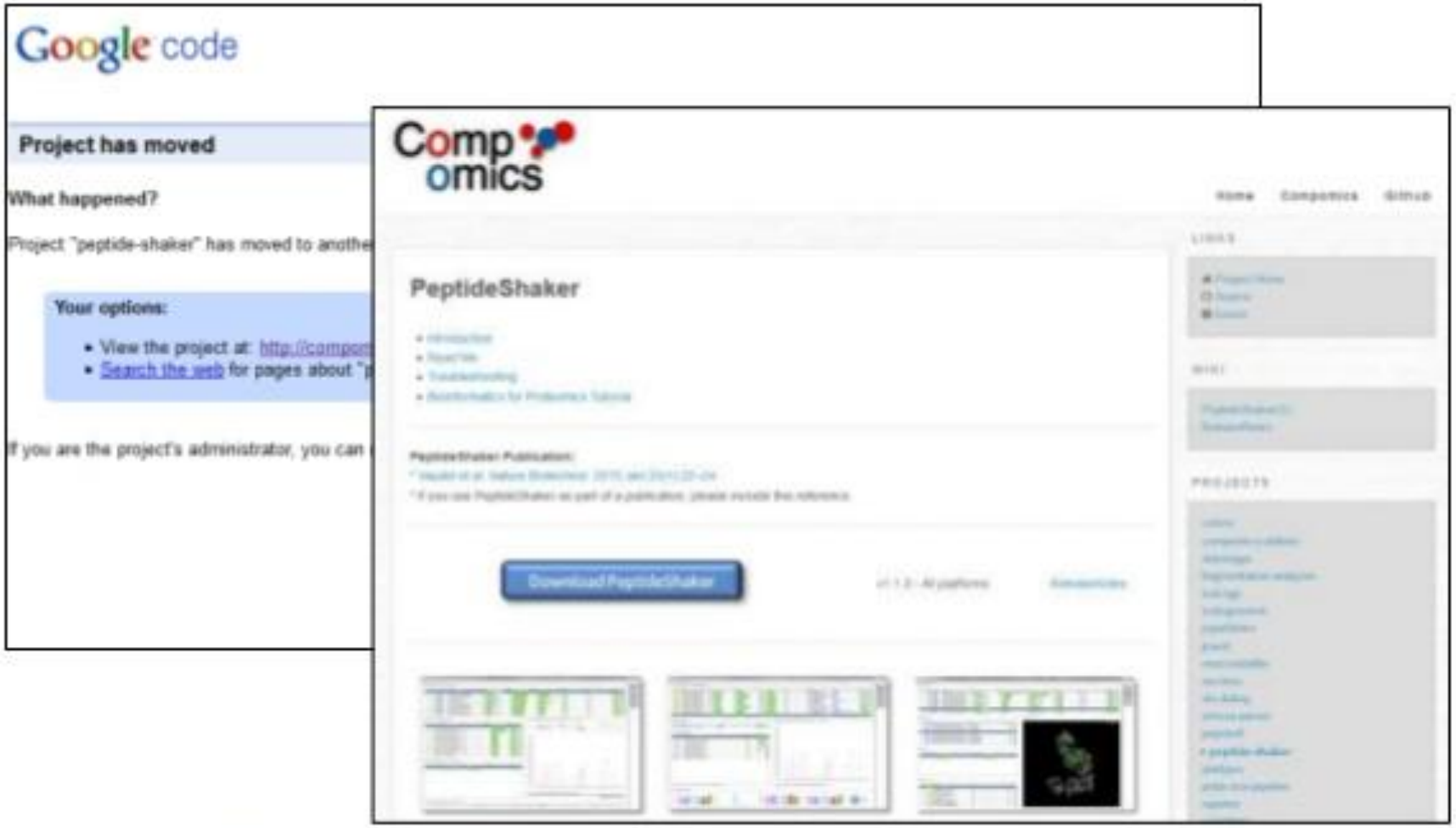
If you have any specific questions, please feel free to contact us at info@ncbi.nlm.nih.gov

Archived Peptidome Studies

PSE101 Comparative analysis of five different preparation methods for identification of complex protein mixtures in yeast
PSE102 Proteomic analysis of the intestinal epithelial cell response to enteropathogenic Escherichia coli
PSE103 Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry
PSE104 Head-to-head comparison of serum fractionation techniques
PSE105 Quality control metrics for LC-MS feature detection tools demonstrated in yeast
PSE106 Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics

Slotta, *Nature Biotechnology*, 2009; Csordas, *Proteomics*, 2013 and Martens, *Proteomics*, 2013

And as responsible caretaker of your stuff, you sometimes need to take action as well



Vaudel, *Nature Biotechnology*, 2015

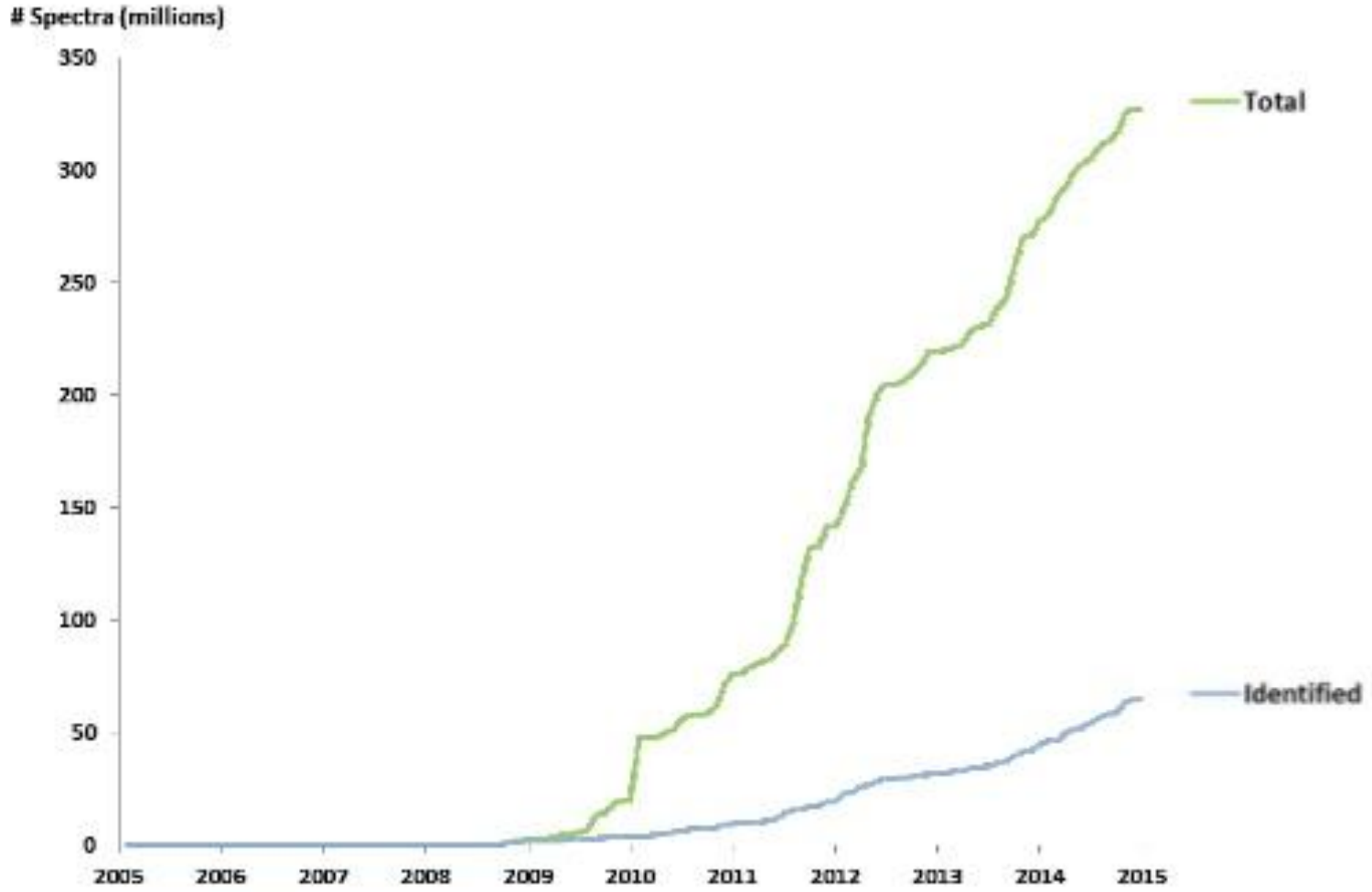
Why should we share our work?

What does sharing our work entail?

What can we do once we start sharing our work?

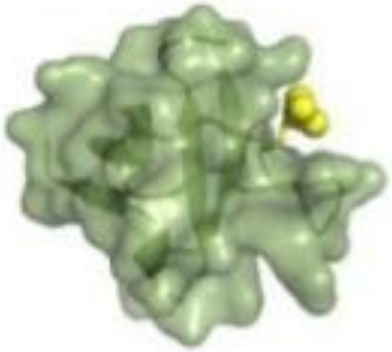
Of sedimentation, opportunity, and dragons

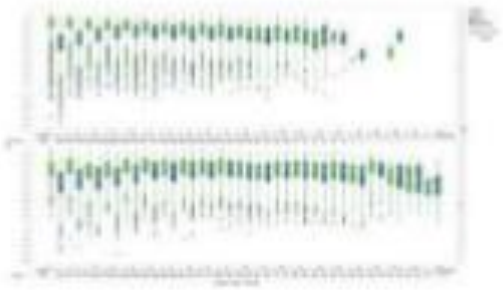
A large amount of post-consumer MS data is collected in public databases such as PRIDE

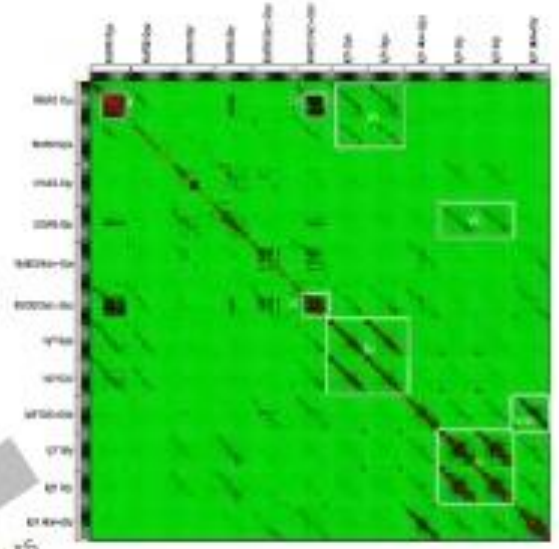


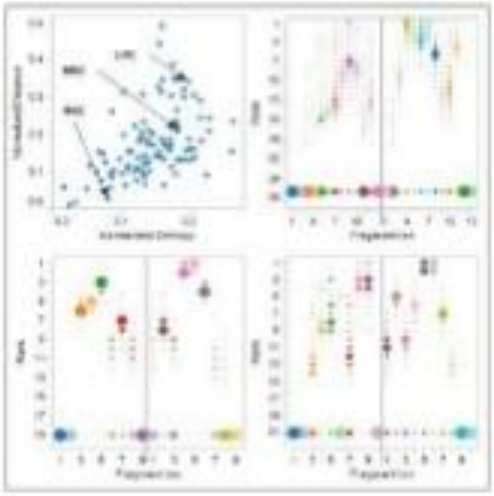
Martens, *Proteomics*, 2005

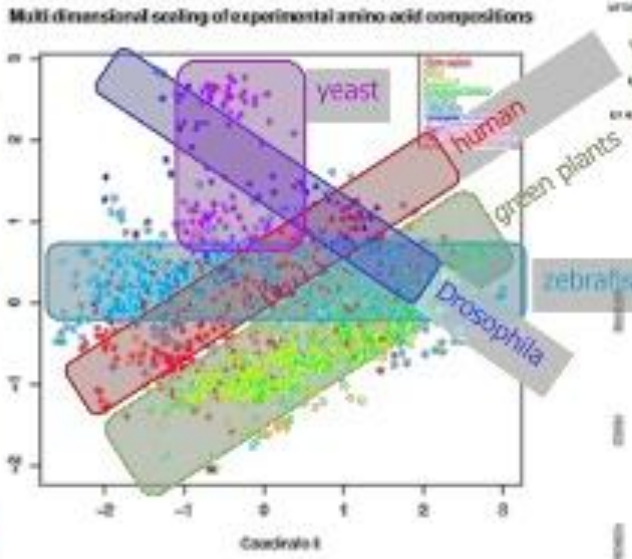
My group published a lot of data re-use, mostly orthogonal, always cross-experiment

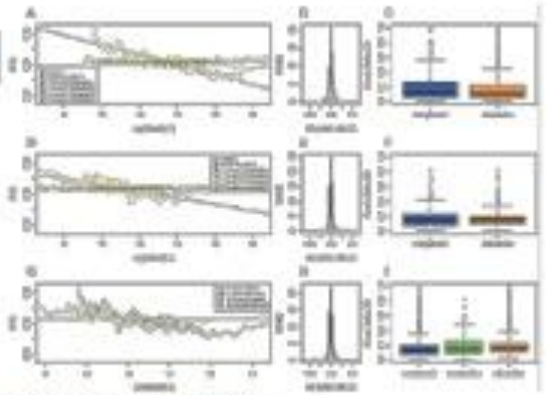






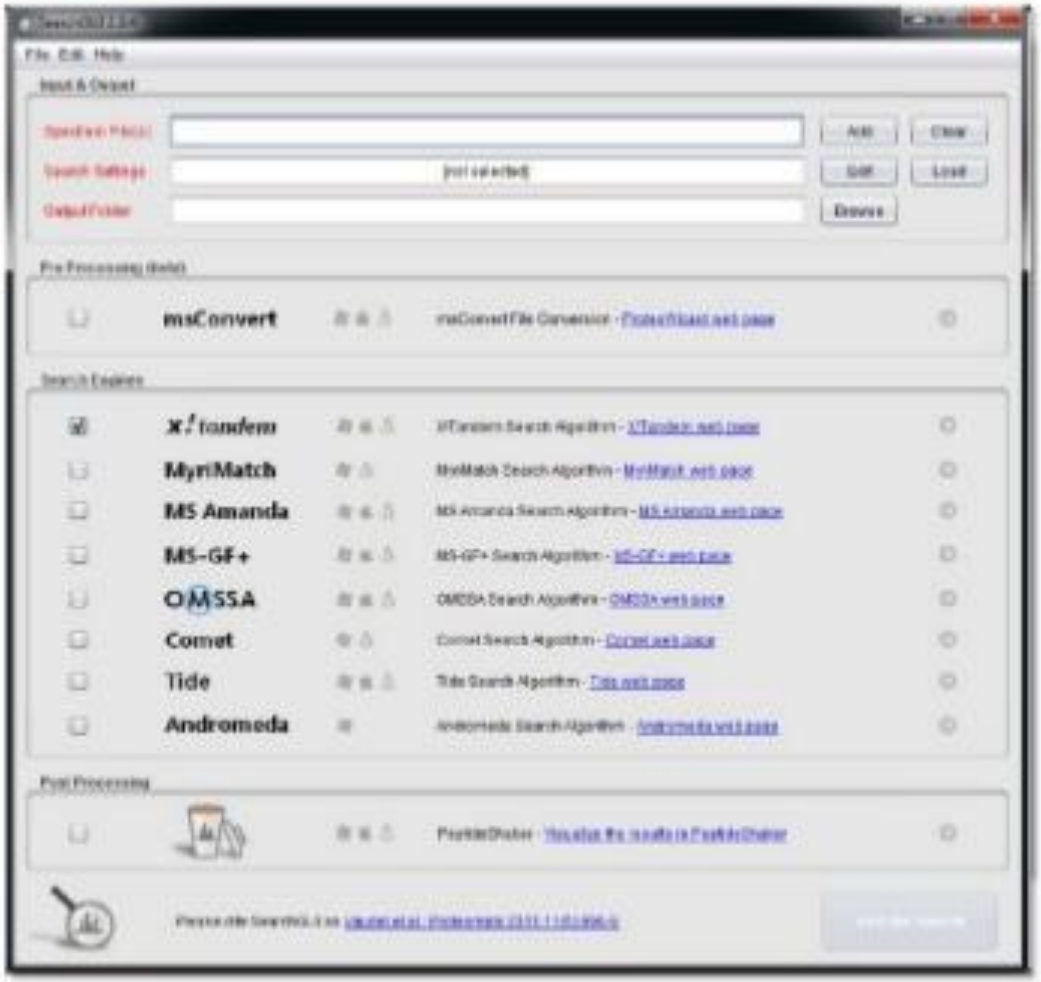






Foster, *Proteomics*, 2011; Colaert, *Nature Methods*, 2011; Barsnes, *Proteomics* 2011, Vandermarliere, *Proteomics* 2013; Degroeve, *Bioinformatics* 2013

Many mass spectrometry data identification algorithms are free and open source

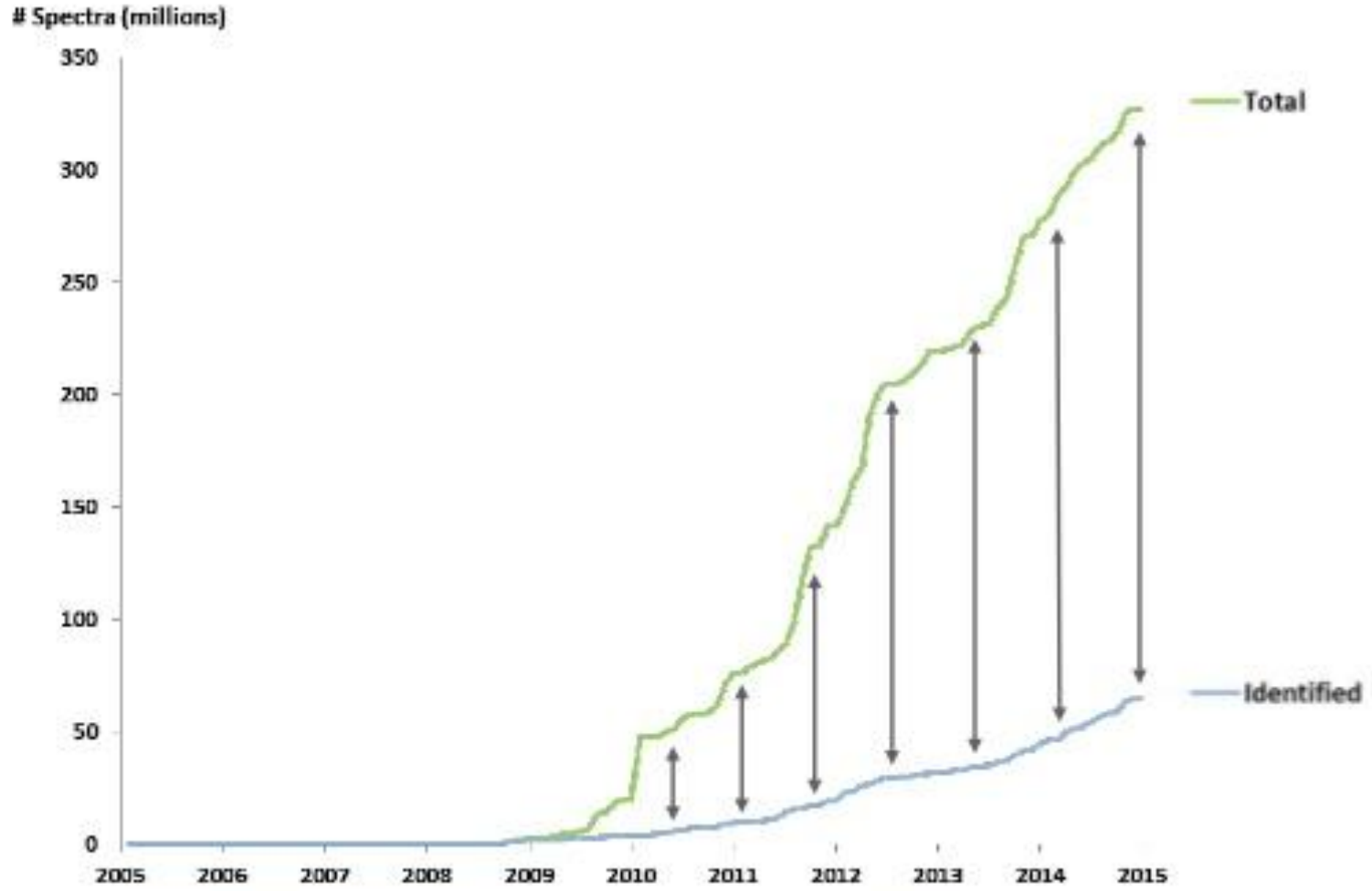


http://searchgui.googlecode.com
Vaudel, *Proteomics*, 2011

Our PeptideShaker tool provides direct access to all public data in PRIDE

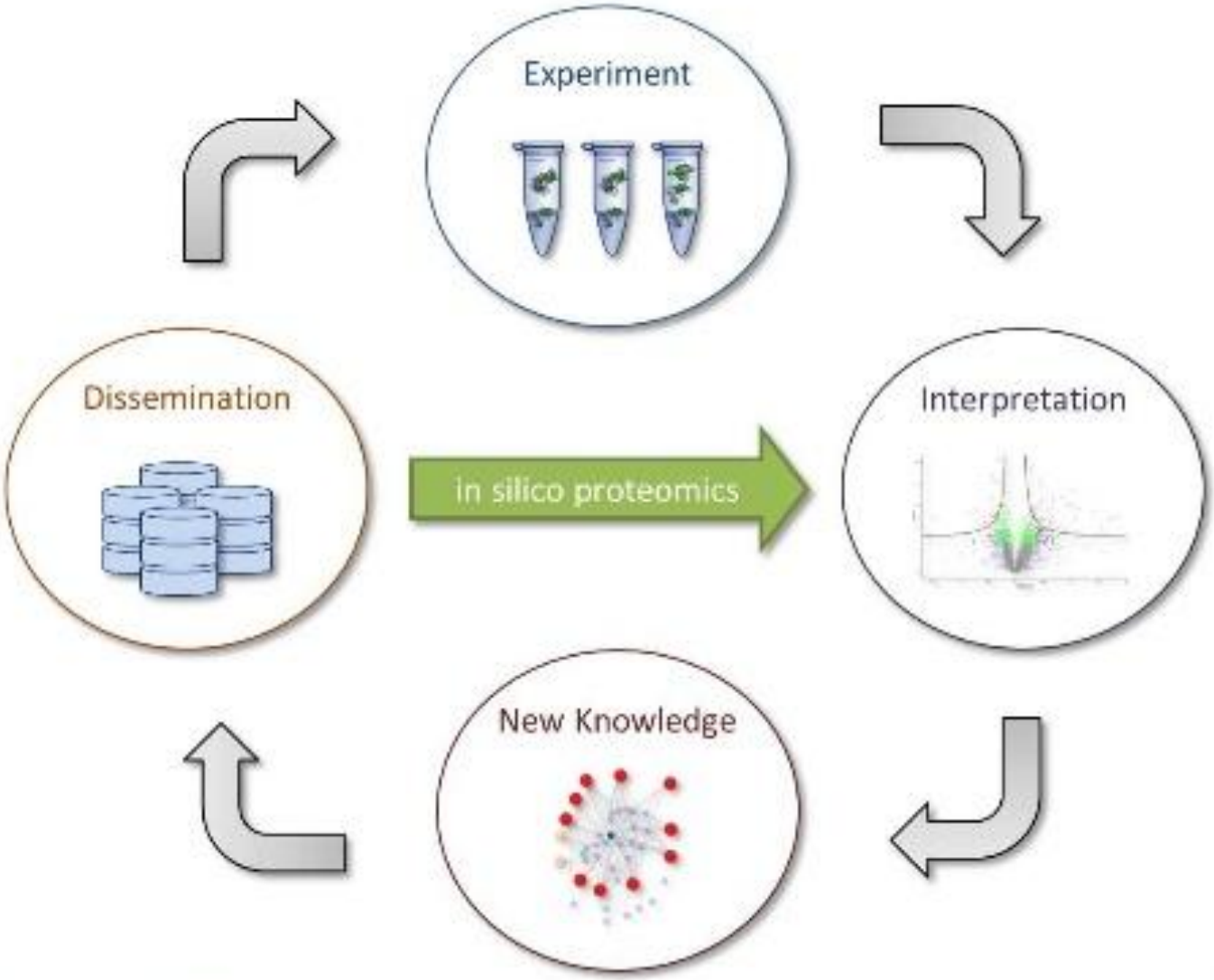
Vaudel, *Nature Biotechnology*, 2015

About 75% of MS based proteomics data currently remains unknown



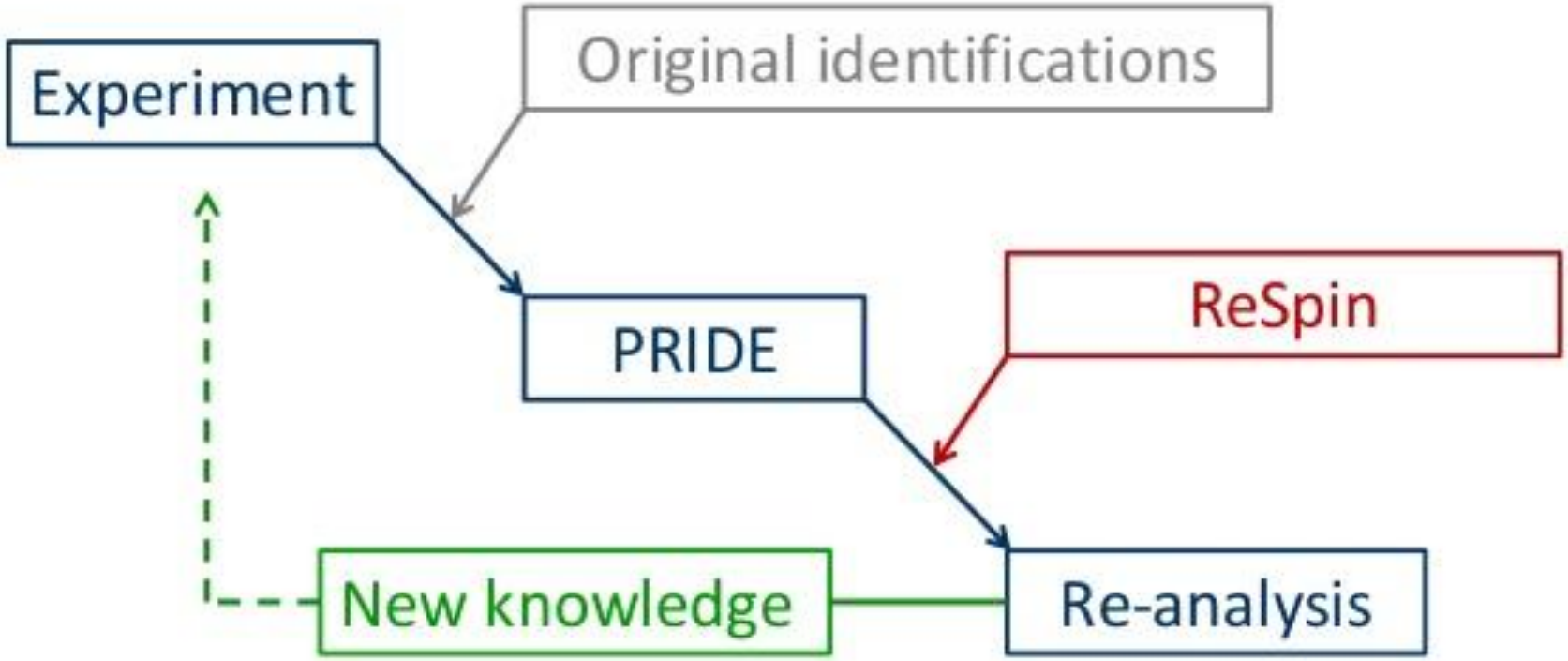
Martens, *Proteomics*, 2005 and Vizcaíno, *Nature Biotechnology*, 2014

These unknown data bristle with opportunity

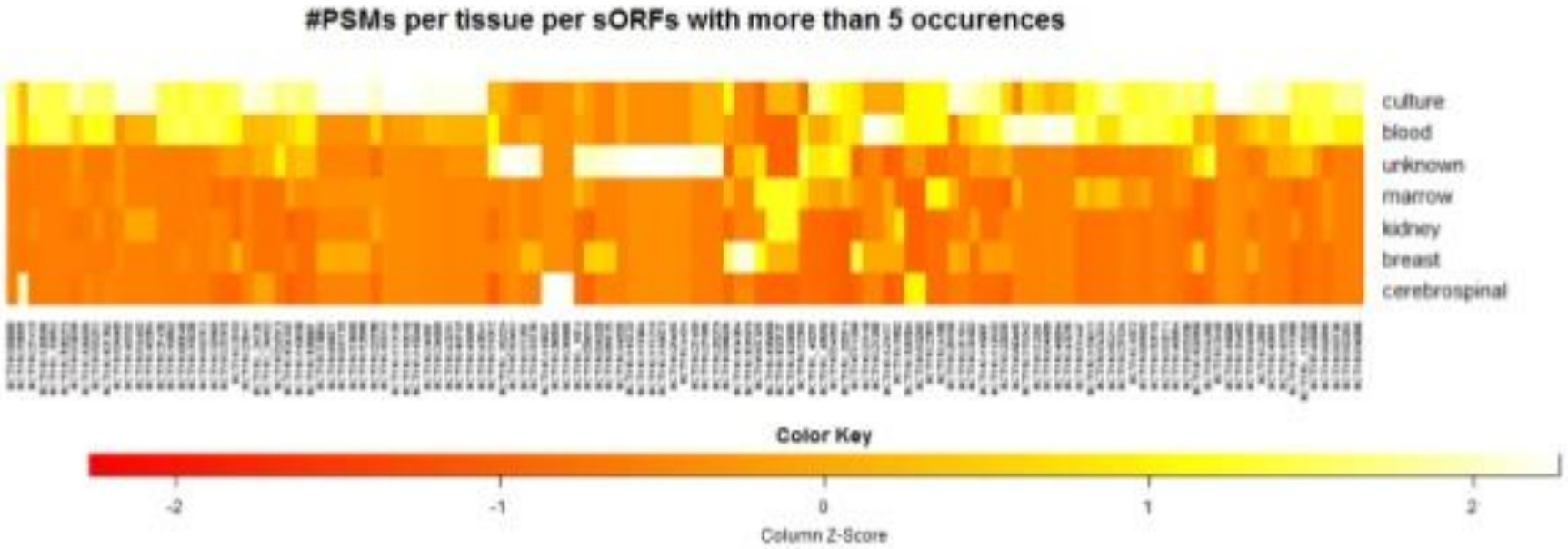


Vaudel, *Proteomics*, 2015

We built the ReSpin pipeline to enable fast re-processing of proteomics data in new ways

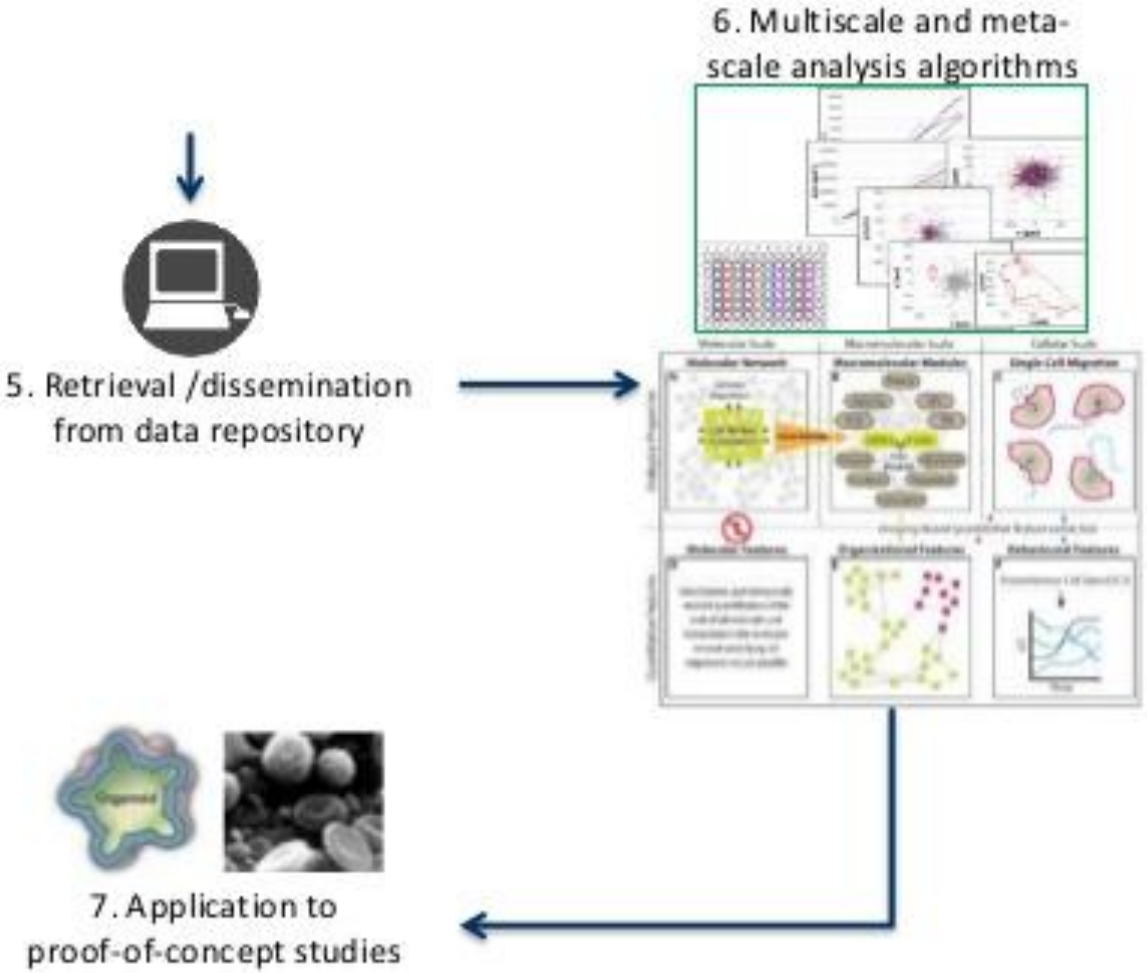


This reprocessing allowed us to detect small ORFs across all public human MS data



Olexiouk, *NAR, in press*

Importantly, deposited data are not the end; an ecosystem enables re-use as well!



Lock, *PLOS ONE*, 2014; Masuzzo, *Trends in Cell Biology*, 2014

Why should we share our work?

What does sharing our work entail?

What can we do once we start sharing our work?

Of sedimentation, opportunity, and dragons

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

And now think about open science, and imagine the opportunities

- What could you do with open science? What could you study? What could you learn?

- What opportunities would present themselves, if…
  - All data (in your field) were available online
  - All algorithms (in your field) were available online
  - All publications (in your field) were open access

- Most of these opportunities are not little steps forward; instead they promise to be revolutionary!

J.R.R. Tolkien, *A Conversation with Smaug*

Compomics

VIB

UNIVERSITEIT GENT

www.compomics.com

@compomics