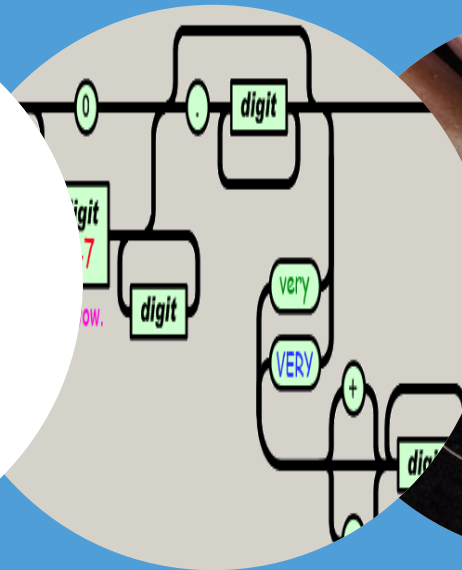
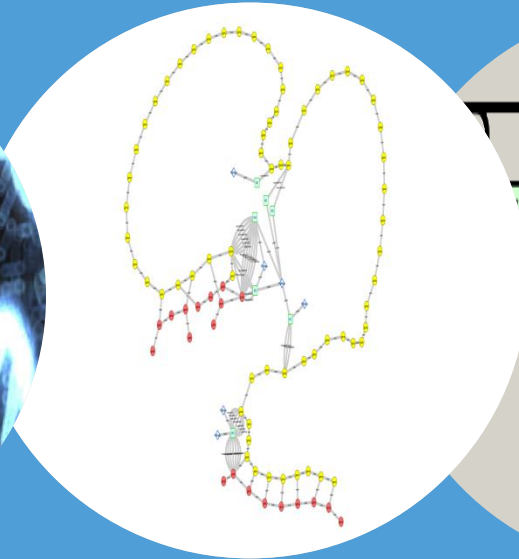


Professional, easy to use and robust bioinformatic tools

Sven Warris, Applied Bioinformatics

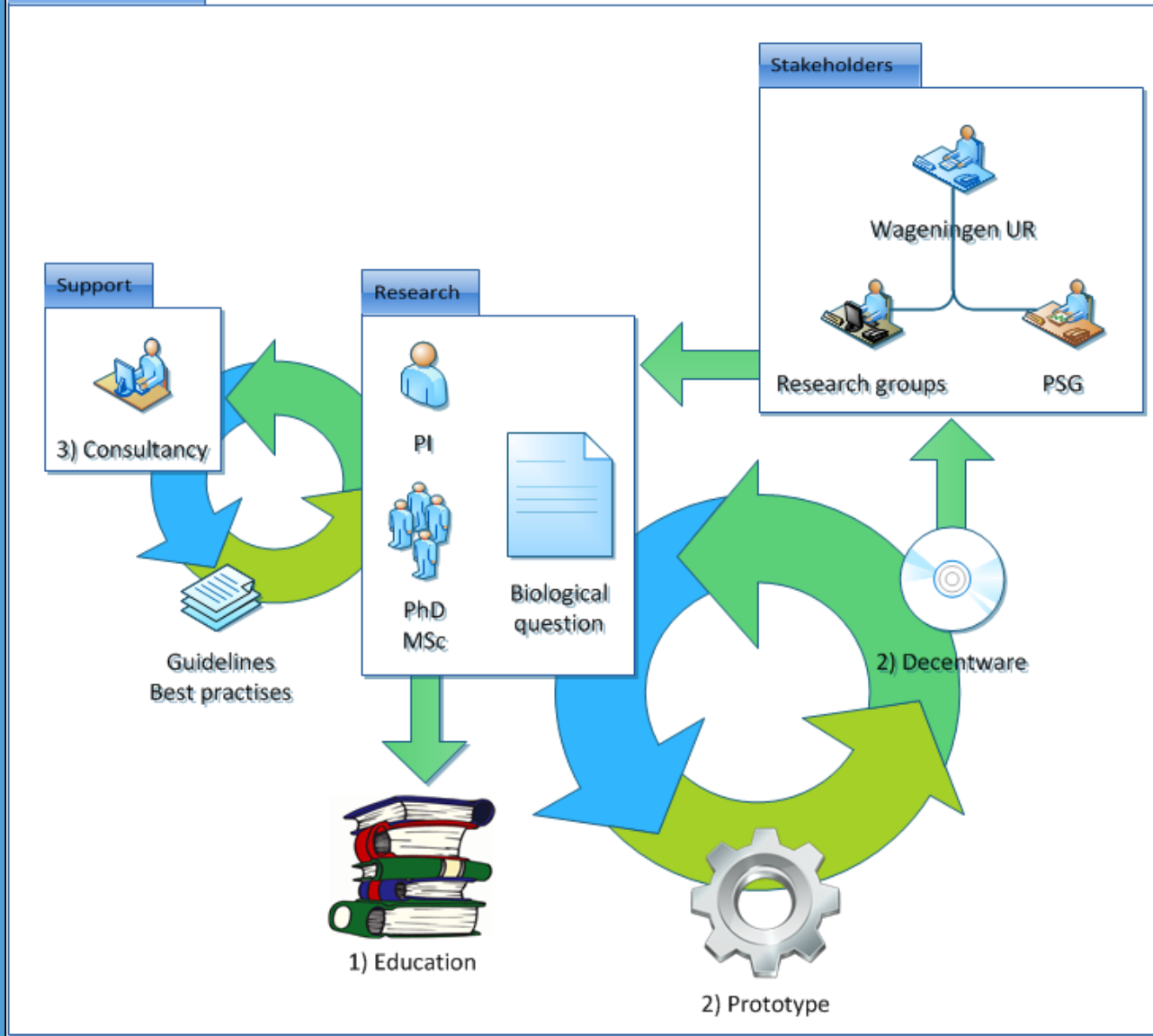
Theo van der Lee, Biointeractions



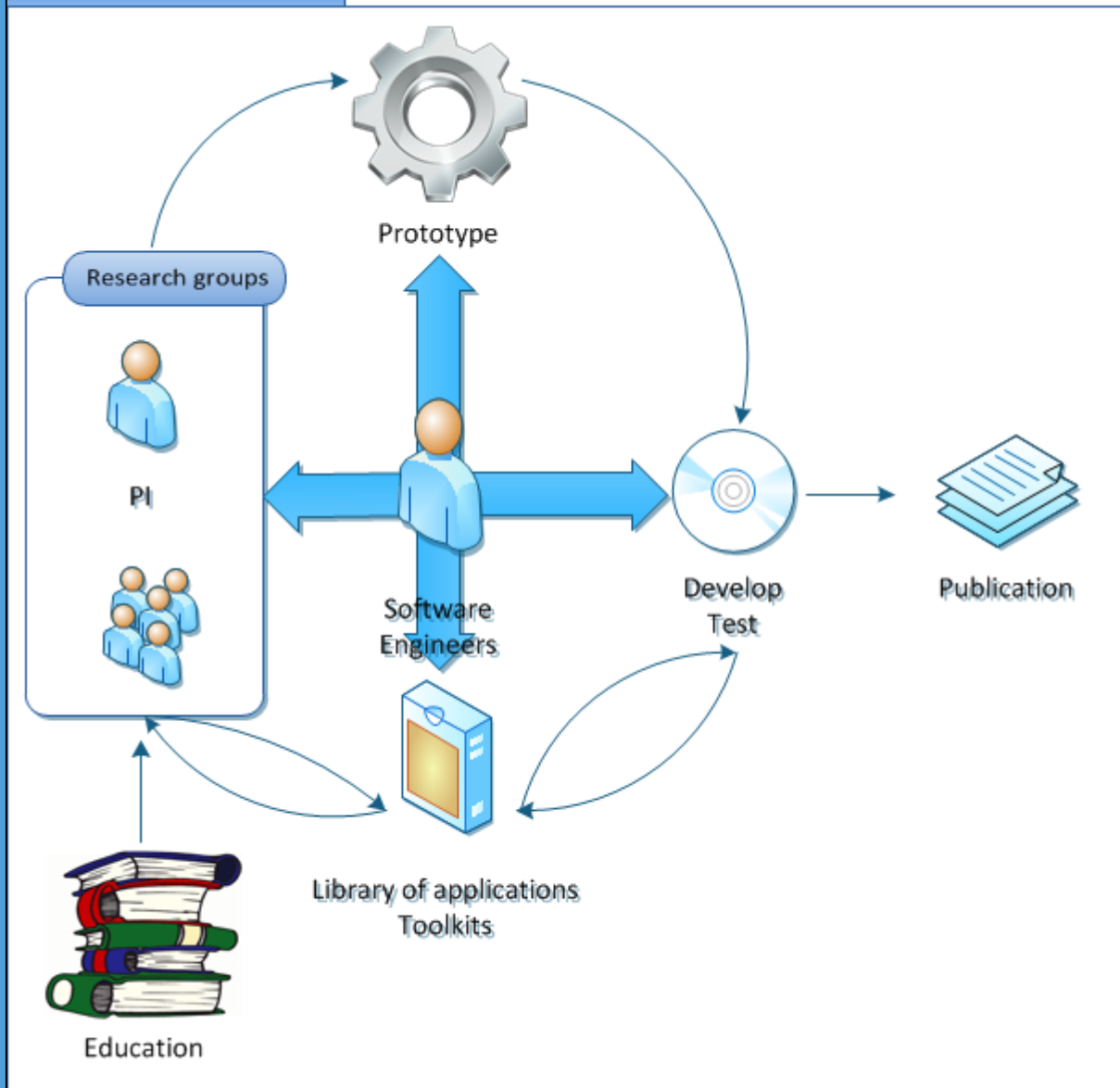
Smart use – Big Data research theme

- Academic software has a bad reputation
 - Hard to read
 - Even harder to understand
 - Impossible to maintain
 - Often does not scale-up
- First steps in making better quality tools
- New collaborations on software development and effective data processing
- Training researchers on data processing, reproducible science and basic





Role of software engineers



Training of researchers

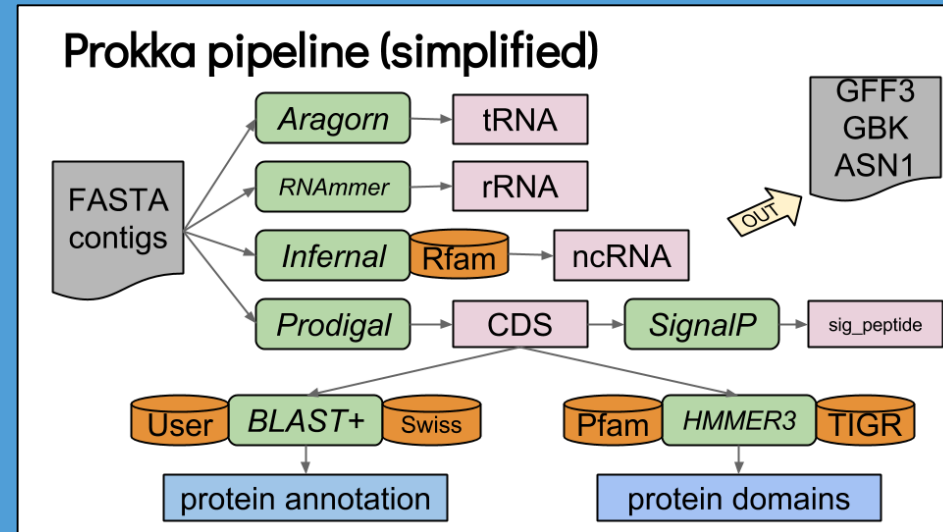
- Enhancing basic technological skills of researchers
- Better communications with engineers
- 2-day workshop principle
 - Software / Data carpentry
 - Open Source
- R & Big Data
- Basic R
- *De novo* assembly using long read technology
 - Genetwister
- Linux, GIT and R
 - DTL / Elixir / WDCC

software carpentry



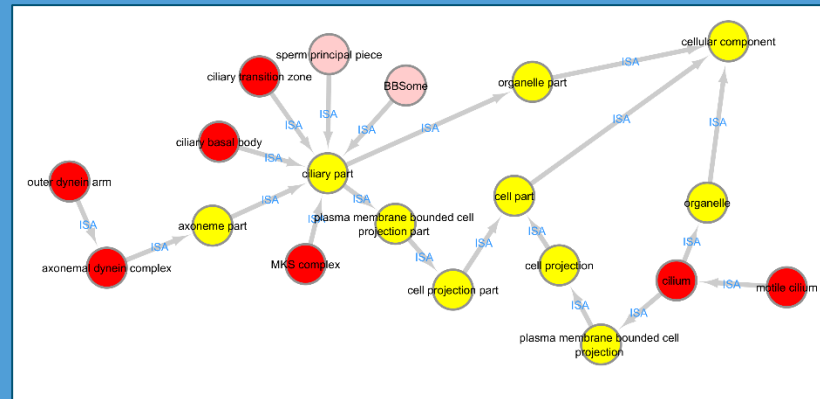
Case study: Genome annotation

- Extensive review of available software
 - Level of maturity
 - Easy of use
 - Does it install
 - Does it run
 - License to use
 - Publications
- Selection: Prokka
- Connected to CLCBio
 - User-friendly interface, already in use
- Evaluated with users



Case study: Cytoscape & Neo4j

- Graph databases & biological data: natural match
- Cytoscape: network visualization tool
 - Protein interaction networks, KEGG, etc
- Developed a plugin to connect Neo4j to Cytoscape
- Over 1000 downloads!




Cytoscape & Neo4j

- New biological insights:
 - 'Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of *Synchytrium endobioticum*' (submitted to PNAS)
- Pre-print on the technology:
 - 'Mining functional annotations across species'
- Plugin available in Cytoscape App Store



Conclusions

- 2-day workshops work
 - Happy participants 
 - In high demand
 - Better understanding of complexity of software
- Worth the effort to upgrade academic software
 - Technological frameworks exists
 - More effective researchers
 - Less loss of know-how, time and effort
 - New insights possible in your field
- First steps!
 - More data, more applications, more researchers

Future work

- Started with project Microbrewery
- Collaboration with Ordina
- Implementing microservices to connect data stores / analytics
 - Small, dedicated software packages
 - Easy to maintain
 - Add and remove functionality
 - Scale-up only those who need it
 - More overhead in (human) communication
- Yeast REACTOME, Neo4j and ADAM (Apache Spark)
- Key technology in a FAIR world



Discussion points

- Where does the responsibility of the researcher end?
- Who is willing to pay for all this:
 - Hard to publish
 - Might increase citations of original work
- Does FAIR data not imply high quality software?
 - Nice that you can reuse the data, but the results are useless when the code contains bugs
- Software engineering is not the holy grail
 - Things still go wrong...



Thank you!

Theo van der Lee

Bart van de
Vossenbergh

Steven Dijkxhoorn

Teije van Sloten

Ron Wehrens

Jan-Peter Nap

Jan van Haarst

Ronald Nieuwenhuis

