Interactive Visualization of Provenance Graphs for Reproducible Biomedical Research

Master's Thesis [Stefan Luger 2015]

Drug development and cancer treatments

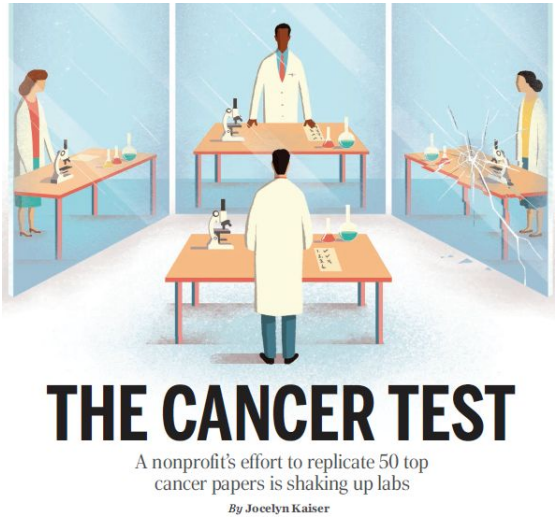
Advances in sequencing and hardware

Large and complex heterogeneous studies that contain external scripts, programs, manual results

No standardized study protocols

http://cancergenome.nih.gov/

Reproducibility crisis



Repeat failures

6 of 53
Cancer papers that Amgen could reproduce

14 of 67
Biomedical papers that Bayer completely reproduced

55%
MD Anderson researchers who could not reproduce a published study

[Kaiser 2015]

How to ensure reproducible research?

Scientific workflow management systems *store*, *analyze*, and ***visualize*** studies

Provenance
track study and its changes over time

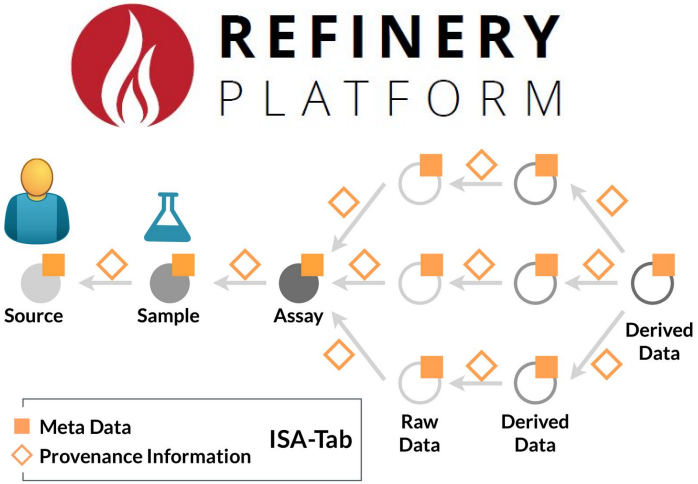
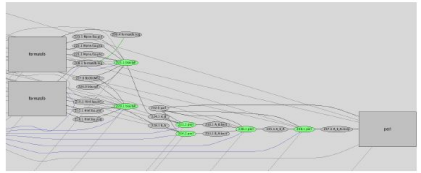
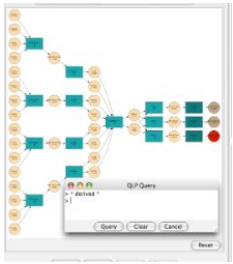


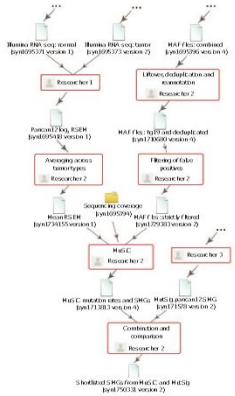
Figure courtesy of Nils Gehlenborg.

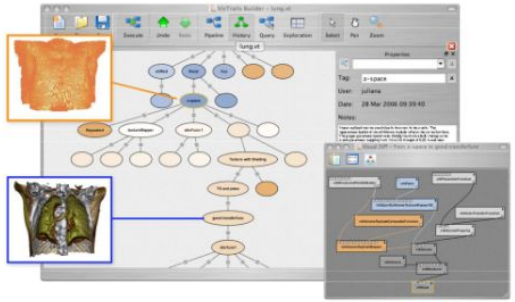
Existing provenance visualizations

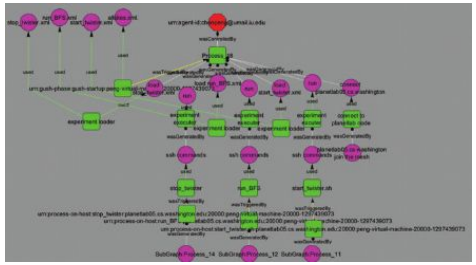
- static node-link approach
- do not scale
- lack visual encoding





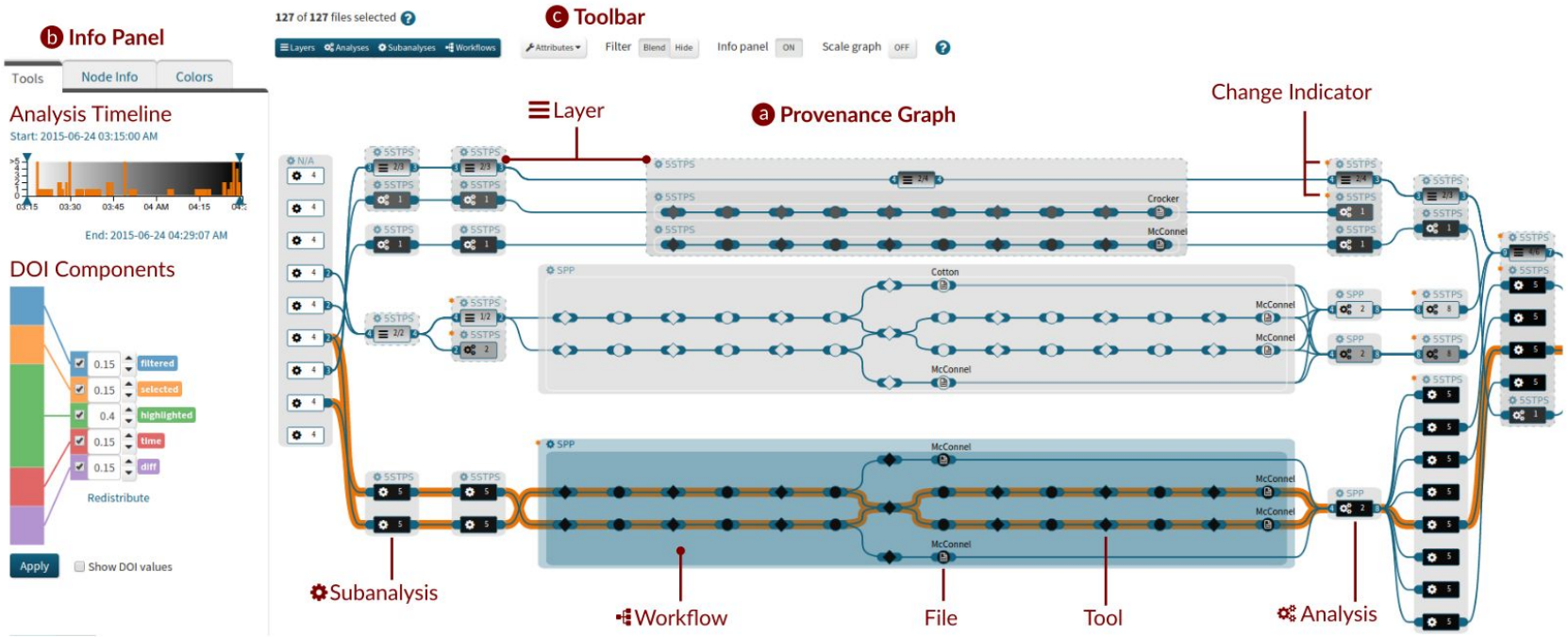






[Altintas2015], [Omberg2013], [Anand2010], [Seltzer2011], [Chen2012]

Provenance Visualization for the Refinery Platform



Best Poster at IEEE VIS'15, USA, Chicago: https://vimeo.com/137118409