

BioBuilds: A Model for Long Term Sustainability of Open Source Bioinformatics

Chris Mueller*, Varshal Davé, Thomas Burnet, and Cheng Lee

Lab7 Systems, Inc., 515 Congress Ave. Suite 2100, Austin, TX 78701

www.lab7.io - www.biobuilds.org

*Please send correspondence to chris.mueller@lab7.io

Open Source Bioinformatics

Open Source software has always played an important role in bioinformatics, providing foundational scientific tools that have enabled many research areas to flourish.

As bioinformatics moves from the academy to the enterprise, it is essential that the community evolve to provide methods for long term support and sustainability.

BioBuilds addresses two important challenges we face when introducing Open Source into production environments:

- The internal costs of supporting Open Source tools with limited human resources
- Long term support and maintenance for Open Source tools, including vendor optimized versions

Free as in ...

Beer? Freedom? **Puppies!** Open Source software has a set of (often overlooked) costs associated with it:

- **Building** is often a non-trivial task that requires tracking down and maintaining multiple dependencies
- Open source tools rarely follow standard **release cycles** and critical updates can be disruptive to production processes
- **QC** and **V&V** are often non-existent, requiring the end user to validate tools independently
- **Interoperability** between tools is as fluid as the tools themselves

Over time, especially when many tools are used, maintenance for a suite of Open Source tools can easily consume an FTE and cost an organization \$100k+/year.

Ready for the Clinic?

Clinical and **applied** markets have very different resourcing and IT support models than the research environments that bred Open Source tools:

- Full-time bioinformaticians (when available) are typically expected to contribute directly to the core mission of the organization
- Tools used are expected to work “out-of-the-box” and require little maintenance
- Tool maintenance can trigger very costly re-validation efforts that impact the productivity of the entire organization

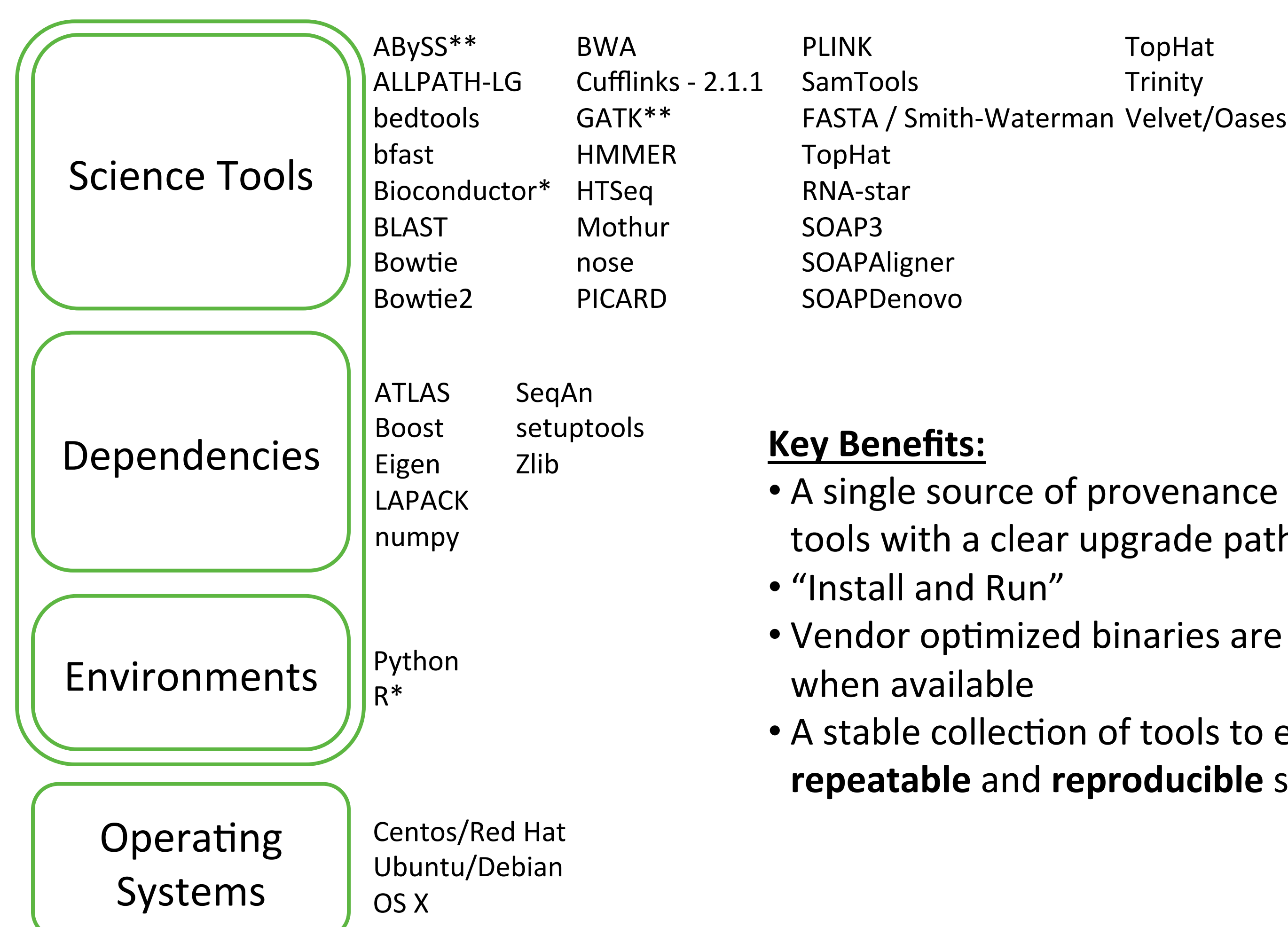
The model used at many academic core facilities of hiring engineering teams to support infrastructure often does not scale to clinical and applied markets.

BioBuilds

BioBuilds’ mission is to ensure long-term support and accessibility for Open Source bioinformatics tools across a broad range of research, diagnostic, and industrial applications.

Distribution

BioBuilds is distributed as a collection of prebuilt binaries for multiple operating systems and hardware platforms. All dependencies are included in the distribution



* Scheduled for Q1 2015 ** Non-commercial use only

Key Benefits:

- A single source of provenance for all tools with a clear upgrade path
- “Install and Run”
- Vendor optimized binaries are included when available
- A stable collection of tools to enable **repeatable** and **reproducible** science

Support

- BioBuilds is bootstrapped by Lab7
- **Sponsorship** models allow third-parties to contribute to the support and maintenance of BioBuilds
 - **Hardware** vendors can support a distribution that includes optimized binaries for their platform
 - **Instrument** vendors can support a distribution with tools used by their platform
 - **Grants** can provide additional support for sustainability projects
 - The **community** can contribute by suggesting new additions and, of course, encouraging end users to use BioBuilds

Open Source Funding

Most Open Source tools used in the Life Sciences are **grant funded** and initially developed to address research topics. While this model has proven effective for many tools and databases, there is pressure on funding agencies to support **research** over **development**. The NIH occasionally issues proposals for sustainability grants, but those programs are often not renewed.

Unlike in many other disciplines, **industry** does not currently play a major role in supporting Open Source bioinformatics software, despite gaining significant benefits from using it.

A core mission of BioBuilds is to develop methods for long term sustainability of Open Source bioinformatics tools. By applying models used in other fields, our goal is to find a model that allows both academia and industry to maintain a solid collection of shared tools.

Sustainability Models

Different approaches for long term support of critical tools have been successful in other domains:

- **Volunteers:** Many Open Source projects are sustained simply by volunteers contributing to the code base. This model is acceptable for hobbyist projects, but struggles to meet the needs of applied communities.
- **Support/Consulting:** Red Hat successfully pioneered the model of building a business around providing support for Linux distributions and using revenues to contribute back to the community. In the scientific space, **Enthought, Inc.**, maintains the **SciPy** collection of Python packages by subsidizing its development with consulting revenues.
- **Direct Corporate Investment:** Many corporations contribute to Open Source projects by hiring staff to work directly on the projects. **IBM** has contributed over a billion dollars to Linux development, mostly by hiring engineers to work directly on components of Linux that benefit IBM. In fact, more than 80% of Linux development is funded directly by corporations.
- **Non-Profit/503(c)(3):** Non-profit foundations have emerged to support specific Open Source communities. The **Mozilla Foundation** and its supporting Mozilla Corporation use \$300M in revenues from Google, Inc., to fund a wide range of Web tools. **NumFocus** operates as an umbrella foundation for NumPy and other projects related to numeric and scientific computing.

