ITCR SIP WG / Genentech PHC Data Science & Analytics – Sustainability Discussion 12/16/19

Lee’s notes

Ryan Copping – group head

Franco Davi – leading academic collaborations

Xiao Li – digital path

Diego Saldana (Basel) – multimodal, patient representations

Tao Xu (Basel) – r packages

Capability building

IT Infrastructure

Analytics platform – R packages

Academics collaborations – just started this year – build stronger links with academia and industry. Current collaborations with Stanford, Berkeley and the Turing Institute.

Criteria for evaluation open-source software

Diego – reproducibility and having a good example application. Starting with papers that have an open-source implementation or pipeline to reproduce results. Availability of data.

Ryan – in the past rarely used anything open-source. Heavy emphasis on SAS. Has changed a lot in recent years, leveraging open-source resources. Thinking about how to put out their own tools.

Arvind – tool quality – is this tool ready to adopt or to take forward? Is there something that Genentech can provide in terms of standards or mentoring on software development?

Ryan – they have an IT team that helps with adoption. It’s a collaborative process.

Ryan – academic collaborations are focused on either research projects of joint tool development.

Ryan – code readability, extensibility, modular design are all considerations.

Tao – better documentation or manuals beyond original publication helps with pre-evaluation of software.

Licensing

Some considerations in building predictive models and things with imaging where there is potential product development.

Academic collaborations

Juli – are these collaborations formalized or informal?

Ryan – building relationships to develop tools and to help with talent development as well.

1. Joint research projects using Roche datasets (w/ Stanford) – can we build out a tool that will help decide on inclusion / exclusion criteria for clinical trials. New biomarkers to outcomes.

New data analysis capabilities. These involve 1-2 data scientists on Genentech side and 1-2 professors + graduate students on the other side. These projects run 1-2 years.

1. Industry-alliance programs – sponsoring a department and sponsoring career days. Sending Genentech scientists for training. 10 running with 5 different institutions.
2. Key advisors – academic experts join an advisory panel in key areas to provide Genentech with monthly advice on strategy with key data areas.
3. Internship program.

Challenges in sharing Roche data – patient consent and IP issues.

Ryan – brainstorming with investigators, approach with completed tools, reaching out with early ideas.

IP ownership is a challenge. If Roche data is being used then IP ownership claims tend to be stronger.

Gordon – examples of successful sustainment and long-term maintenance and how is that funded.

Diego – project or codebase that is competitive where the entire industry can benefit. In this case there may be some cooperation among industry on open-source projects. A single company may lack a critical mass.

Ryan – typically rely on people’s passion to maintain these things.

FDA regulations

Ryan – when we do a submission to FDA we have used a validated software environment and tool (SAS). In recent years as moved to open-source, there are concerns if these tools would be accepted by the FDA. There has been some progress here with FDA accepting .

Two-different R environments – validated, formal FDA environment, exploratory environment where people can use their own packages. Software testing is essential.

What open-source tools are they using?

Ryan – R, Python, TensorFlow

Diego – Tidyverse (R).

Ryan – People believe SAS is a tried and true validated system.

Ryan – vagueness.

Juli – do you do any work in commercial cloud environments?

Ryan - they use AWS for some projects (50/50). There are other projects where data is restricted and cannot be stored in the cloud.

Michael Reich – what is your outreach process?

Ryan – thinking about how to scale academic collaborations and governance. Opportunistic, through existing connections.

Guergana – how can we get ahold of you?

Franco Davi ([davi.franco@gene.com](mailto:davi.franco@gene.com)) – they will reach out to the group.