Sustainable Community Investment in Action - A Look at Some of the R Consortium Funded Grant Projects and Working Groups

John Mertic
Director of Program Management
Linux Foundation
R is seeing remarkable growth

Source: https://stackoverflow.blog/2017/10/10/impressive-growth-r/
But with growth, come challenges

• ‘The official documentation is written as though for people already R experts and is almost completely useless for beginners. It took me at least a couple of years to figure out how to read help pages!’

• ‘Why n number of packages are there to do same task? Decision tree is produced nicely by cutter command but still apart exists. We need to take out the not so good packages and retain the best in a newly launched list of packages.’

• ‘Not all packages are maintained in a consistent way, lack of backward compatibility, poor readability, and rather infrequent updates.’

• ‘Working seriously on internationalization.’

• ‘Supporting participation of underrepresented minorities in the community.’

• ‘Helping to ensure that the development and improvement of R itself is sustainable in the future. I'm concerned that R Core members and CRAN volunteers are being asked to do too much work with no compensation, and that this is unsustainable.’

• ‘Trying to align how packages work a bit more. Not to say a heavy hand should apply here. Maybe packages that meet certain standards get a certified sticker’

Source: R Consortium Community Survey 2017
Increasing enterprise use of R brings new needs

- ‘Simplifying enterprise level implementation that is cost effective for budget conscience public health facilities would be extremely well received.’
- ‘Users all want updated and more packages every day, and I have to get them certified before they can be used (corporate security policy)’
- ‘The main problem I’ve had is external dependencies. I work in a corporate environment and am not an administrator on my PC, so having to install external software means a multi-week engagement with the IT department.’
- ‘I love open source and minimal corporate meddling, but there are cases where positive corporate influence (good influence, not meddling) can be beneficial. I'm sure there are some who will disagree, but that's my stance.’
- ‘Corporate (Microsoft, IBM) visibility to our upper-level IT managers giving R credibility.’
- ‘I haven’t successfully navigated our company’s politics to make R a central "corporate" analytical solution. To date, it remains a tool that's isolated to my particular workstation.’

Source: R Consortium Community Survey 2017
The future for R

- Organizations are increasingly using R as an essential part of their business processes.

- Increased involvement from and alignment with R users and R stakeholders would benefit everyone.

- The big challenge: How can the R Community scale to address the needs of the next n million users?
R Consortium’s mission is to work with and provide support to key organizations and groups developing, maintaining, distributing and using R.
R Consortium Momentum

3 years since launch

$650k+
Grant dollars awarded

30+
Funded Projects

10+
Hosted Working Groups

150+
Sponsored R Meetups

15+
Supported R Community Events

50k+
R users members of supported meetups

3
useR! events sponsored
Foundations enable ecosystems by creating trust

Successful Projects depend on members, developers, infrastructure to develop technology, which is turned into products that the market will adopt.
R Consortium community investments are guided by the Infrastructure Steering Committee

<table>
<thead>
<tr>
<th>Top Level Projects</th>
<th>Grant Program</th>
<th>Working Groups</th>
</tr>
</thead>
</table>
| • Work considered critical for long term ecosystem sustainability  
  • Projects include R-Hub, R-Ladies, and R User Group Support Program | • Enables community leaders to apply for grant funding to solve key community needs or bootstrap new efforts  
  • Successful grantees include DBI, RL10N, and R package builders  
  • 30+ grants awarded and $650k+ funding to date. | • Vendor-neutral collaboration groups for building consensus amongst the community  
  • Groups include certification, code coverage, and internationalization |
Examples of ISC Funded Projects
DBI Project - Kirill Müller

- DBI, R’s database interface, is a set of methods declared in the DBI R package.
- Communication with the database is implemented by DBI backends, packages that import DBI and implement its methods.
- Several rounds of funding have built out packages and specific backends
  - DBI
  - DBItest test suite
  - Three backends to open-source databases
    - RSQLite
    - RMariaDB
    - Rpostgres
- Learn more at [http://r-dbi.github.io/DBI/](http://r-dbi.github.io/DBI/)
This project seeks to improve and maintain tools for providing binaries

- On Windows: upgrade the Rtools compiler toolchain, and provide up-to-date Windows builds for the many external C/C++ libraries used by CRAN packages.
- For MacOS: expand the r-hub "homebrew-cran" tap with formulas that are needed by CRAN packages but not available from upstream homebrew-core.
- Eventually: lay the foundation for a reproducible build system that is low maintenance, automated as much as possible.
Simple Features (officially Simple Feature Access) is both an Open Geospatial Consortium (OGC) and International Organization for Standardization (ISO) standard ISO 19125 that specifies a common storage and access model of mostly two-dimensional geometries (point, line, polygon, multi-point, multi-line, etc.) used by geographic information systems.

R did not have direct support for simple features – this package was funded to add this.

2,300+ commit and 40+ committers over the past two years.

Enabled other new packages to form and funded
  – stars - Scalable, spatiotemporal tidy arrays for R
  – mapview - Interactive viewing of spatial data in R
  – Quantity Calculus for R vectors - https://github.com/r-quantities/

Impact has been an increased usage of R for working with spatial data.
Enabling community interactions

- R Consortium is the leader in supporting regional events and user groups in the R Community.
- Aligning and supporting broad efforts ensures the R community is vibrant and sustainable.
- R Consortium Diversity and Inclusion working group compliments as a community driven effort to furtherly define standards and best practices for healthy, diverse, and inclusive R community events.
Strategic Investments for the future of the R Community
R Consortium focus – adoption and support of R in organizations

- Provide a home for vendor-neutral collaboration.
- Enable the community to drive packaging standards to grow organizational use of R.
Organic driven collaboration builds the base for sustainability

<table>
<thead>
<tr>
<th>Regional User Groups, Meetups, and Events</th>
<th>Industry vertical groups</th>
<th>Working groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unifying the regional communities to help support R users and enable collaboration.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Consortium R User Group and Small Conference Support Program</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion of R usage, barriers and challenges, successes in deployment, and tools used</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bringing together a diverse set of constituents to drive consensus on key technology and community initiatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code Coverage Diversity and Inclusion Certification</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R/Medicine 2018
R Packages are a huge part of the R user experience

• The ecosystem is a key asset to the user
  – “I suspect that it’s always got what I need! (My needs are fairly diverse too!!)”
  – “Almost everything I want to do I can find a package that does it for me. I put very satisfied because most of the time there is a professional or near-professional level package that does what I want.”
  – “If you want to perform an analysis that is not available in other statistical software, you can always turn to R. The contributed packages are significant.”

• With growth comes increased user expectations
  – ‘Why n number of packages are there to do same task? Decision tree is produced nicely by cutter command but still apart exists. We need to take out the not so good packages and retain the best in a newly launched list of packages.’
  – ‘Lack of quality oversight = lack of trust. Many packages are not tested adequately for wider use than the author’s thesis or book chapter, for example unusual data situations, larger datasets (inefficient programming), or both. Several packages overwrite base R functions, often without warning - this is sloppy programming.’
  – ‘Not all packages are maintained in a consistent way, lack of backward compatibility, poor readability, and rather infrequent updates.’
Addressing trust and expectations in R packages

Package maintenance
- Development processes
- Licensing/providence
- Security management
- Open source best practices

Package product quality
- Platform/architecture support matrix
- Performance
- Quality standards

Industry/use-case alignment
- Common packages being used
- Regulatory/stability requirements
## Strategic program alignment

<table>
<thead>
<tr>
<th></th>
<th>Package Maintenance</th>
<th>Package Product Quality</th>
<th>Industry/use-case alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal</strong></td>
<td>Create a standard metric for R packages wanting to be used in business critical environments</td>
<td>Set community quality stands and tools to aid package developers</td>
<td>Vendor-neutral, community driven consensus on recommended packages for problem domains</td>
</tr>
<tr>
<td><strong>Efforts</strong></td>
<td><img src="image" alt="Open source wide program focused on driving OSS best practices in projects" /></td>
<td><img src="image" alt="Standard, vendor-neutral, and open service for R package building and validation" /></td>
<td><img src="image" alt="Starting the conversation on comparing usage patterns to find opportunities for vendor-neutral collaboration" /></td>
</tr>
<tr>
<td><strong>Alignment</strong></td>
<td>Open source wide program focused on driving OSS best practices in projects</td>
<td>Standard, vendor-neutral, and open service for R package building and validation</td>
<td>R/Medicine 2018</td>
</tr>
</tbody>
</table>
CII Best Practices Badge makes it easier for users of open source projects to see which projects take security seriously. Introduced in May 2016, over 1,000+ FOSS projects have registered for the badge. While only 10% of the projects successfully passed, every one of them made an improvement to achieve a badge.

Learn more about the program and start the process with your package at

<table>
<thead>
<tr>
<th>R-package</th>
<th>Description</th>
<th>URL</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>pandas</td>
<td>Flexible and powerful data analysis / manipulation library for Python, providing labeled data structures similar to R data.frame objects, statistical...</td>
<td><a href="http://pandas.pydata.org">http://pandas.pydata.org</a></td>
<td></td>
</tr>
</tbody>
</table>
Continued development on R-hub to automate package quality

- R-hub (https://builder.r-hub.io/) is a service for developing, building, testing and validating R packages, which simplifies the R package development process.
- R-hub was the initial funded project by R Consortium
  - Goal to fill the void of having a service for ensuring R package build successfully and are properly packaged for inclusion in CRAN.
  - Has made a significant impact on package quality submitted to CRAN.
  - Open Source project, and a free service for all members of the community (https://github.com/r-hub/)
- Future functionality
  - Add additional license and validation checks, along with code scanning for common package issues.
  - Integration with package repository(s).
  - Better integration with R package developer workflow.
Creating spaces for industry to drive consensus and tools for collaboration

- Hosting R/Pharma and R/Medicine working groups which are building vertical industry consensus.
  - PSI AIMS SIG focusing on Pharma industry regulatory compliance
- All industry efforts benefit from R Consortium community wide efforts
  - CII Best Practices badging program and R-Hub for R package standards
  - R Consortium Working Groups program for collaboration tools and space
  - Future R Certification Program to ensure R users have a common baseline of knowledge
How to get involved?

1. Visit R Consortium website
   https://www.r-consortium.org/

2. Learn about ISC funded and sponsored work
   https://www.r-consortium.org/projects/

3. Learn about the RUGS program
   https://www.r-consortium.org/projects/r-user-group-support-program

4. Have your organization support R Consortium with membership
   https://www.r-consortium.org/about/join
Questions?