

Results from the 2014 survey of VIVO implementation sites

Paul Albert, Alex Viggio, Jon Corson-Rikert, & Kristi Holmes



August 8, 2014



Who participated?

Section 1

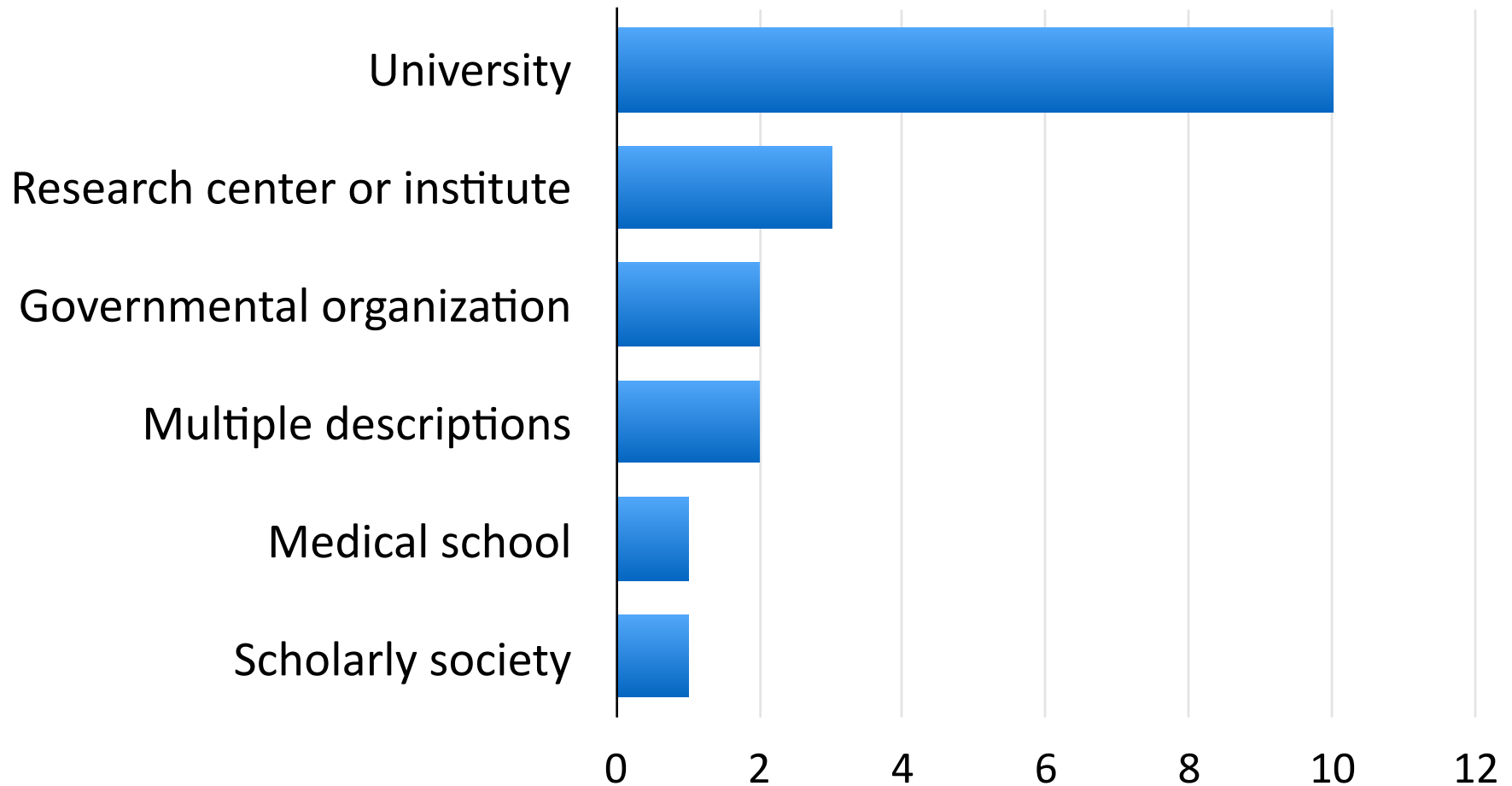


Seeking clarity from the community

- How do sites solve problems?
- How important is improving performance to the community?
- Is VIVO-ISF too complicated?
- How do VIVO sites really feel about documentation?
- To what extent are we meeting the needs of the international community?
- Which code improvements are most important to the community?



What types of implementation sites participated?



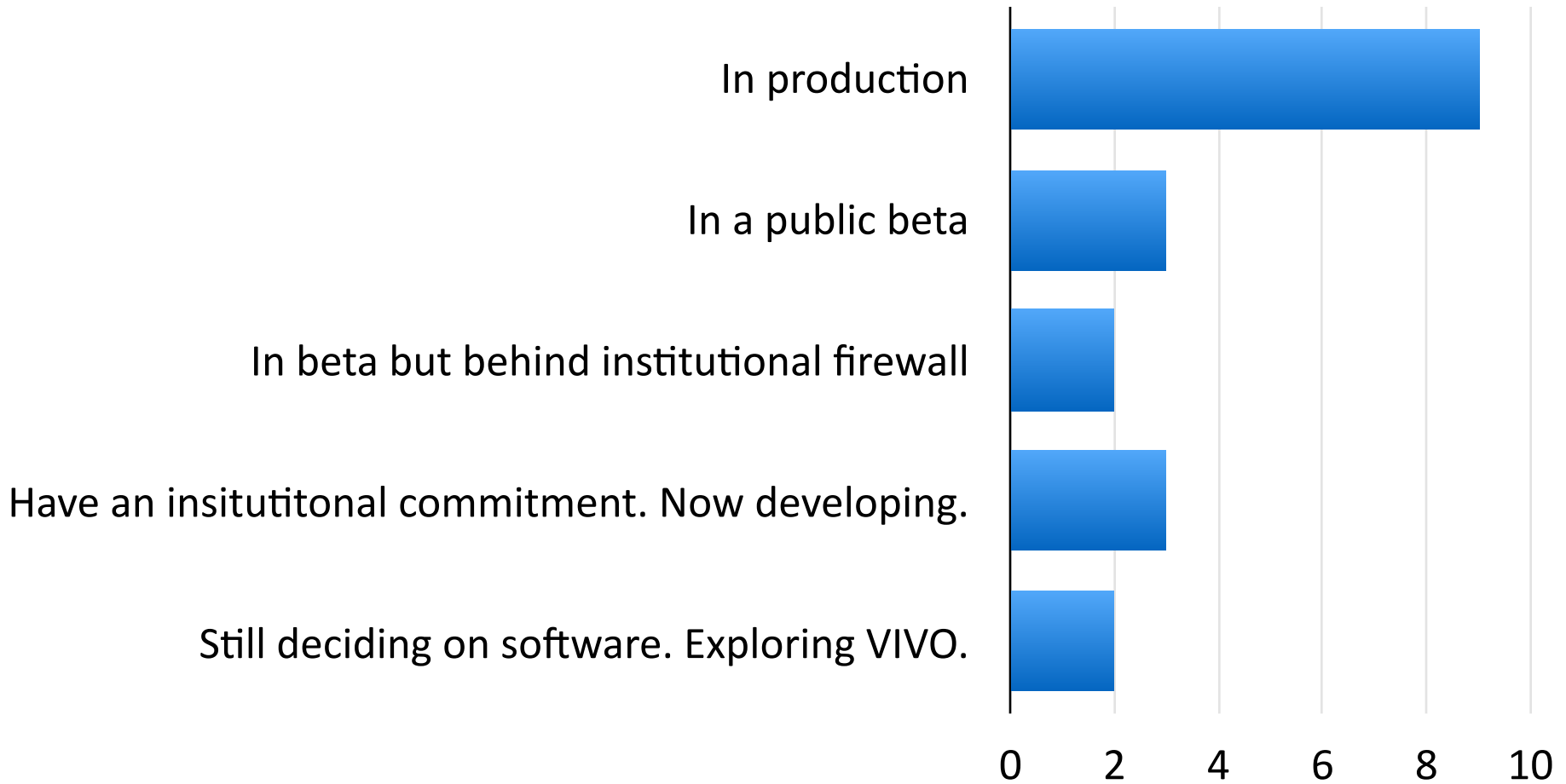
19 sites* participated

Name	Location	Existing/Expected Profiles
American Psychological Association	Washington DC, USA	200,000
Brown University	Providence, Rhode Island, USA	3,500
Cornell University	Ithaca, New York, USA	12,273
Duke University	Durham, North Carolina, USA	8,000
Eindhoven University of Technology	The Netherlands	4,157
Fundação Getulio Vargas	Rio de Janeiro, Brazil	1,000
Laboratory for Atmospheric and Space Physics, CU Boulder	Boulder, Colorado, USA	
Smithsonian Institution	Washington, DC, USA	1,000
Tetherless World Constellation at Rensselaer Polytechnic Institute	Troy, New York, USA	2,000
The Scripps Research Institute	La Jolla, California, USA	350
University of Colorado Boulder	Boulder, Colorado, USA	1,792
University of Florida	Gainesville, Florida, USA	>20,000
University of Melbourne	Melbourne, Australia	7,000
US Environmental Protection Agency	US wide	1,900
Virginia Tech	Blacksburg, Virginia, USA	3,000
Weill Cornell Medical College	New York, New York, USA	5,900

* as well as three anonymous sites



What is your stage of development?



What do the results mean?



See raw results here: bit.ly/vivosurvey



Personnel, hardware, ontology, and performance

Section 2



Personnel

- If in production, how many full-time equivalent (FTE)-years were needed to get VIVO ready for production?

Average	Min	Max	Respondents
3.44	0.5	12	13

- How many FTEs in the past year?

1.66	0.1	4.5	18
------	-----	-----	----

- If in production, how many FTEs are required to maintain it?

1.57	0.1	4	12
------	-----	---	----



Experience

- Before starting on VIVO, did any of your developers have experience with semantic web technologies?
 - Yes: 2
 - No: 16

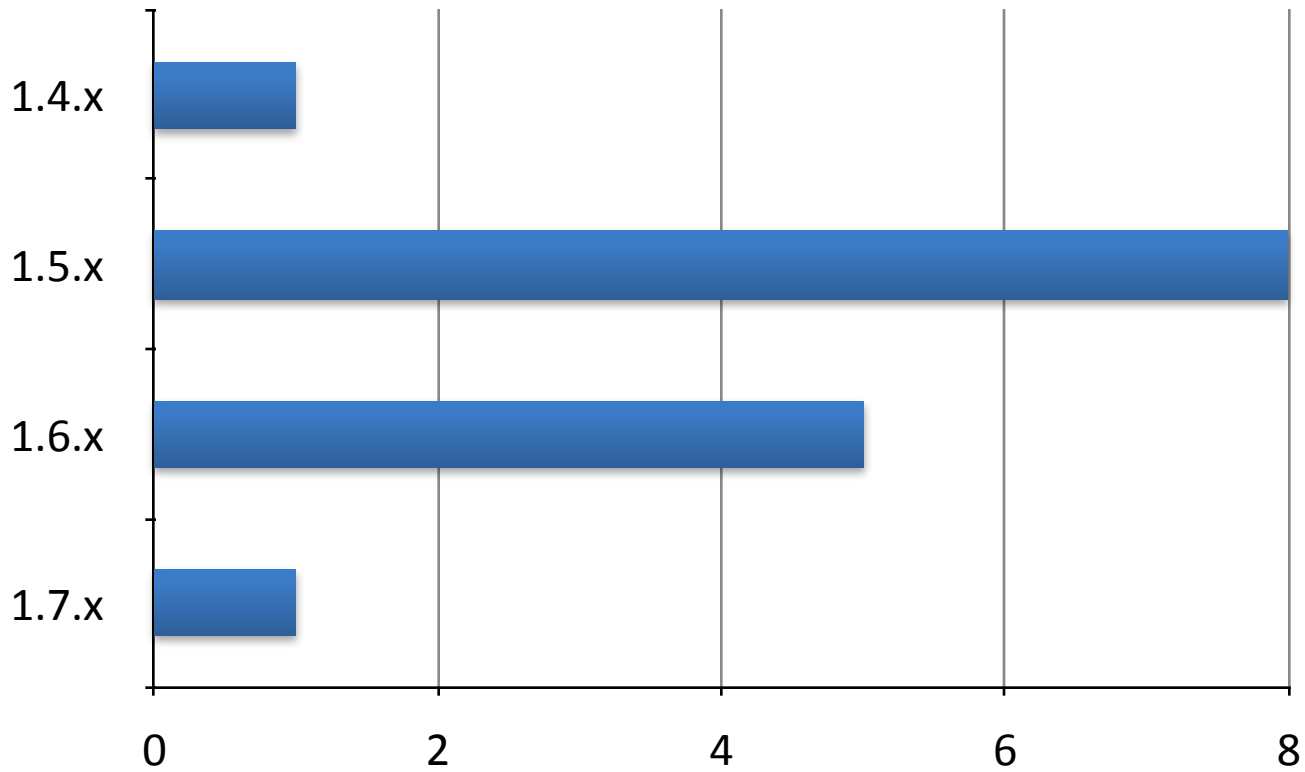


Role distribution

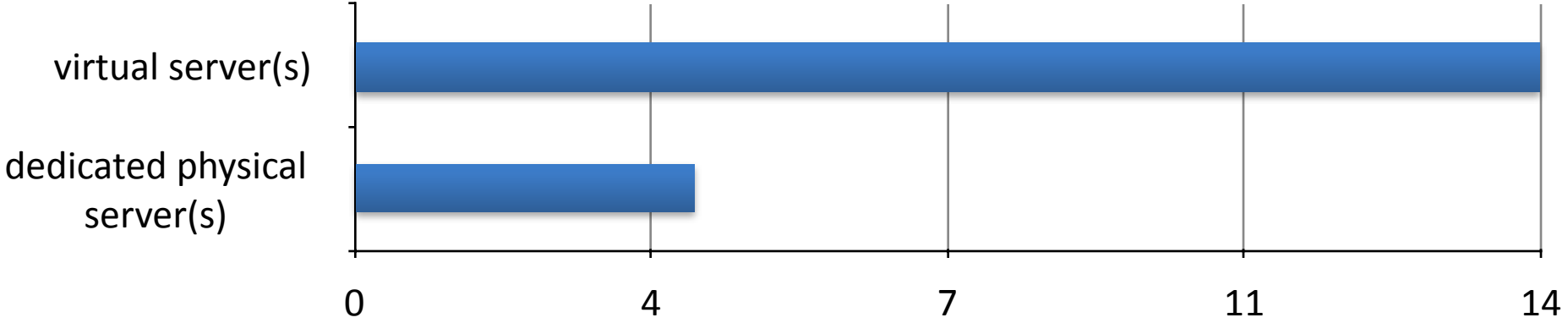
Role (expressed as FTE)	Avg	Min	Max	Comment
Programmer/ETL specialist	0.52	0	2.1 – 3.0	Writing scripts for ingest and update
Project manager or	0.30	0	1.1 – 2.0	
IT/systems engineer	0.19	0	0.6 – 1.0	Installation, configuration, tuning
Content coordinator	0.16	0	0.6 – 1.0	
Ontologist	0.15	0	0.6 – 1.0	Local ontology extensions
Project executive or	0.13	0	0.1 – 0.5	
Web developer	0.13	0	0.6 – 1.0	Branding and customization
Marketing/outreach person	0.12	0	0.6 – 1.0	
Database administrator	0.09	0	0.1 – 0.5	
Business analyst	0.09	0	0.1 – 0.5	
Manual entry clerk	0.07	0	0.1 – 0.5	



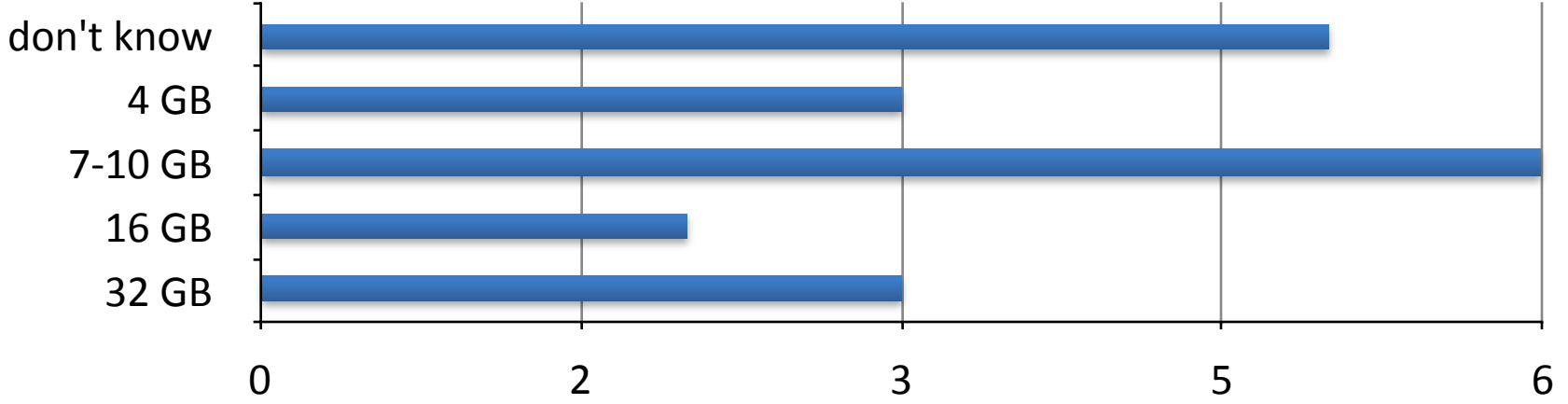
VIVO version



Hardware



Memory

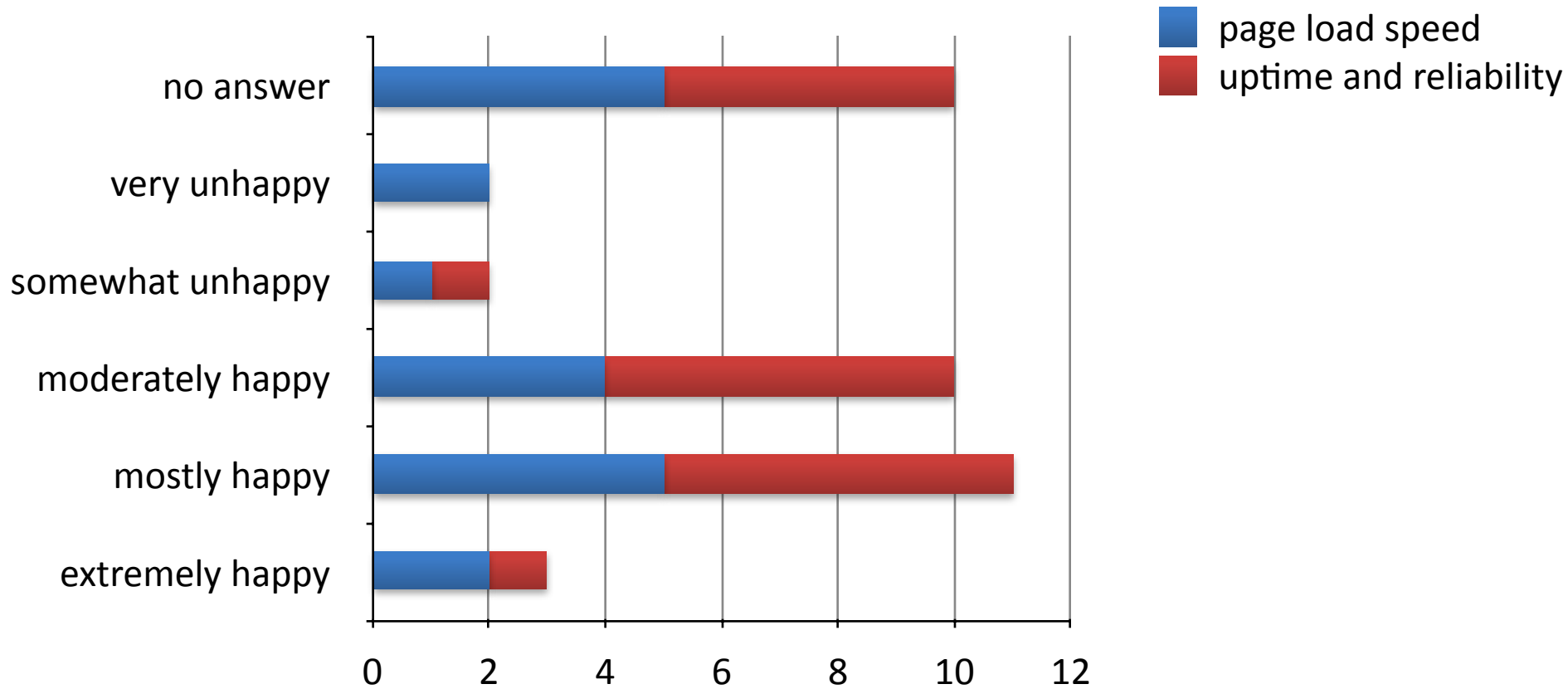


Environment

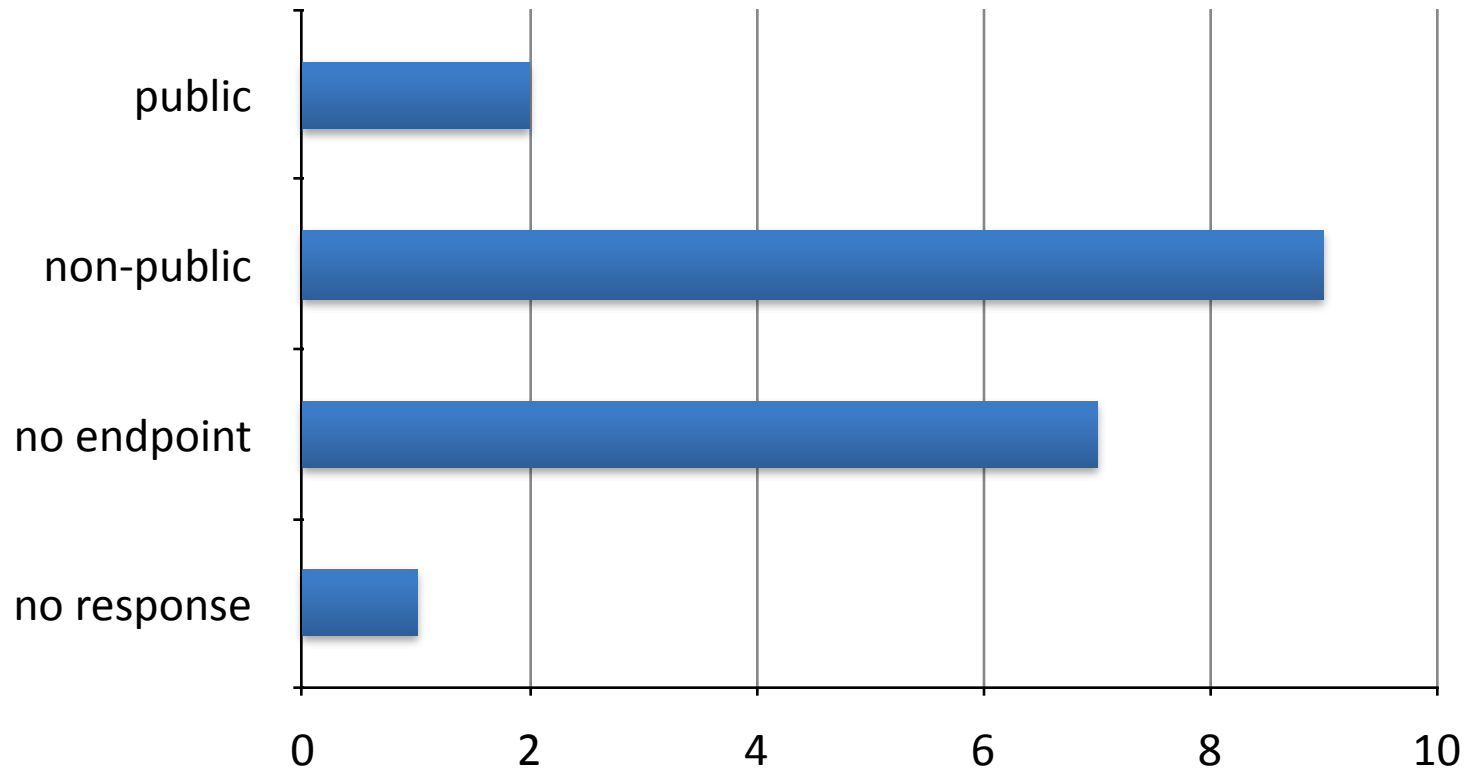
- MySQL database (1 AllegroGraph)
- Tomcat (1 Glassfish)
- Apache web server
 - 2 sites using Apache mod_cache, 1 Squid
- 1 site using Amazon EWS
- 1 site doing load balancing
- 2 use other technologies (Ruby, .Net) in their stack to support editing



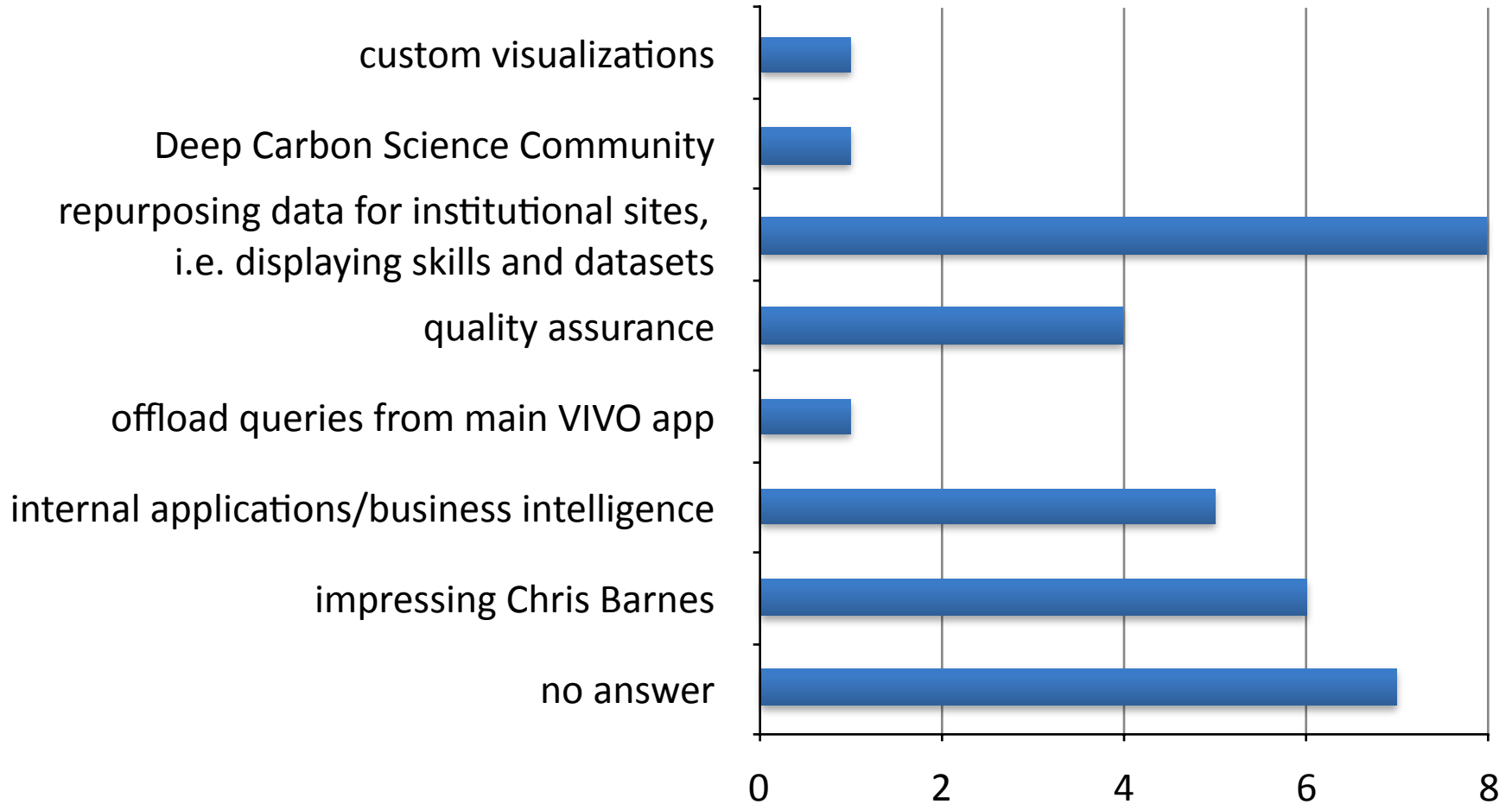
Stakeholder satisfaction: performance and uptime/reliability



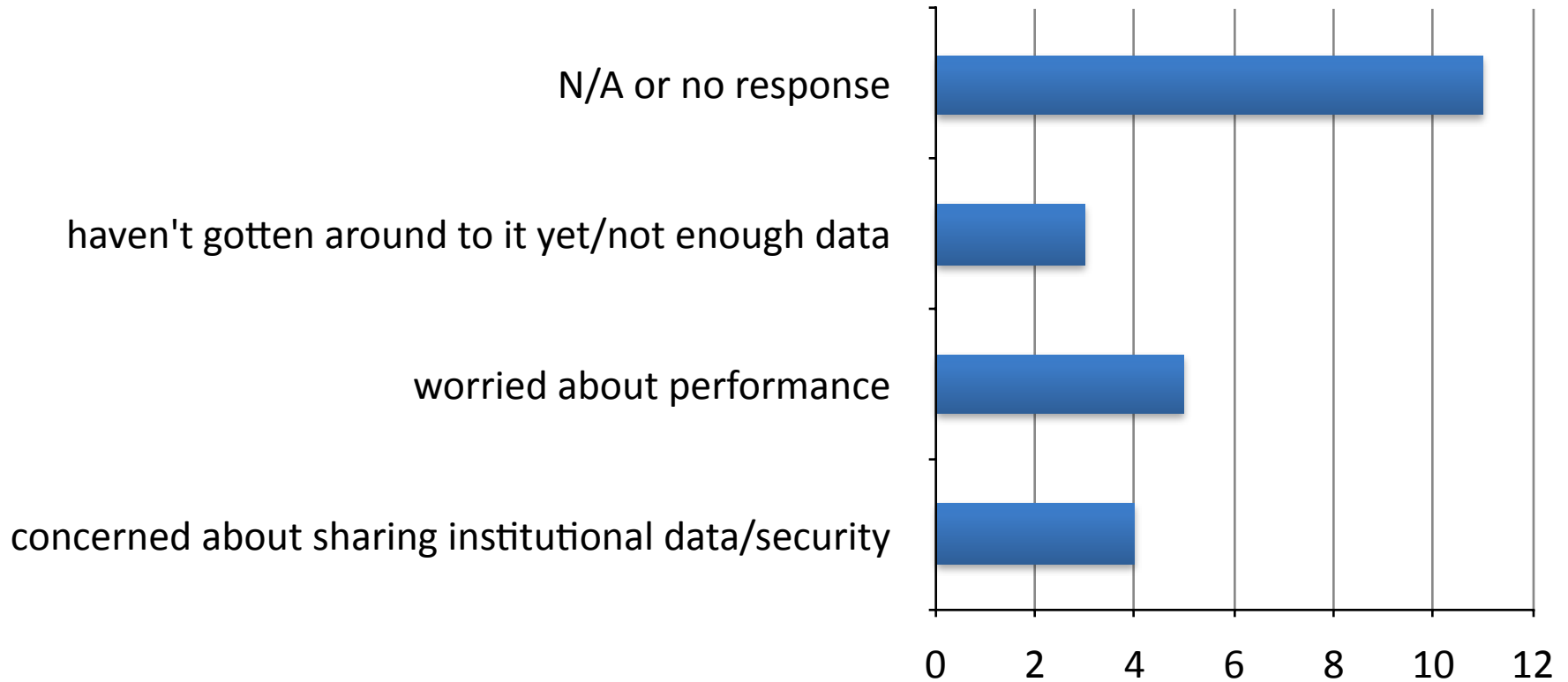
SPARQL endpoints



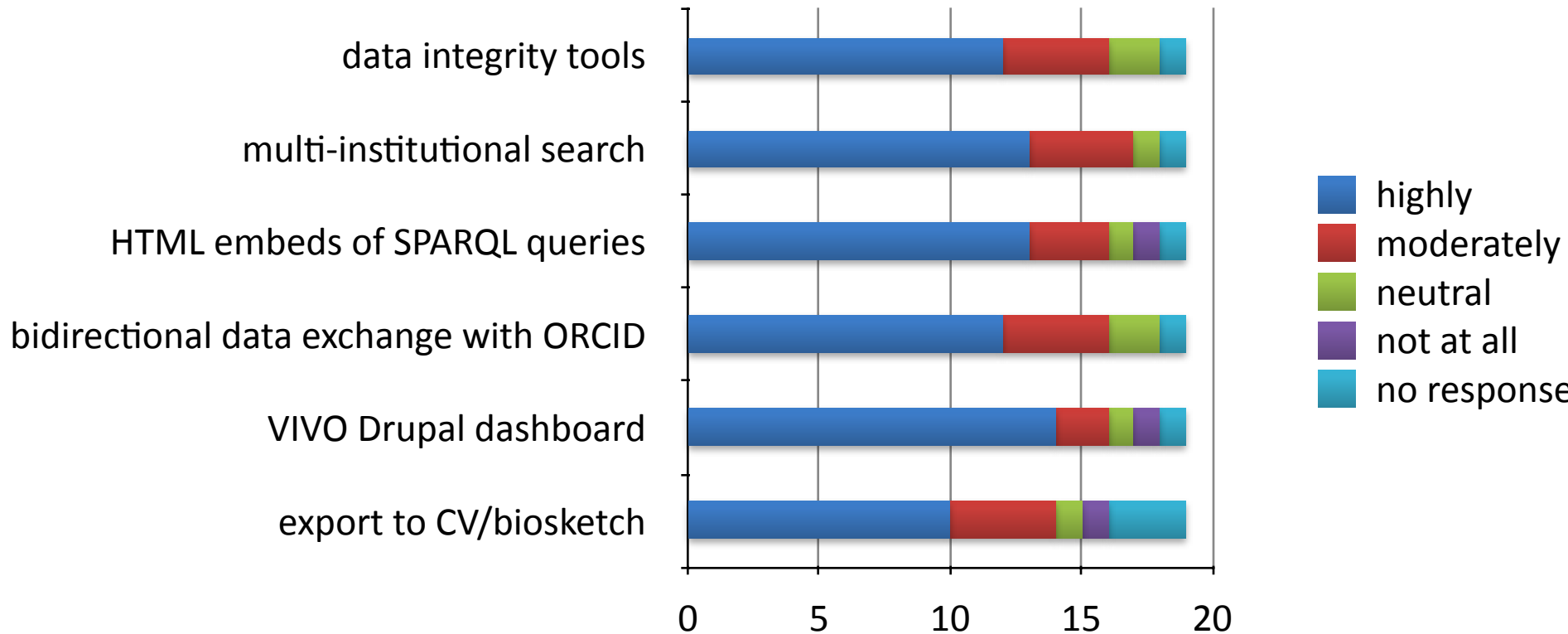
How is your SPARQL endpoint used?



If not, why not?

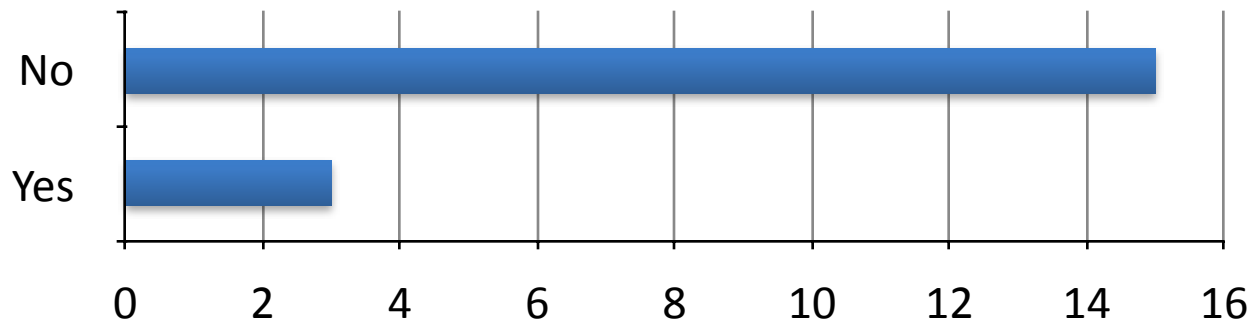


Desired features

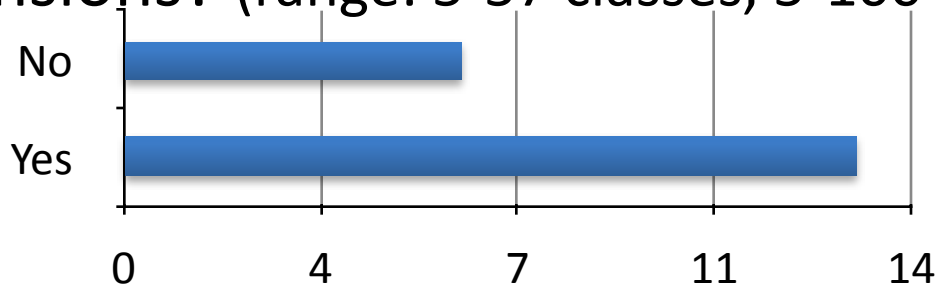


Ontology

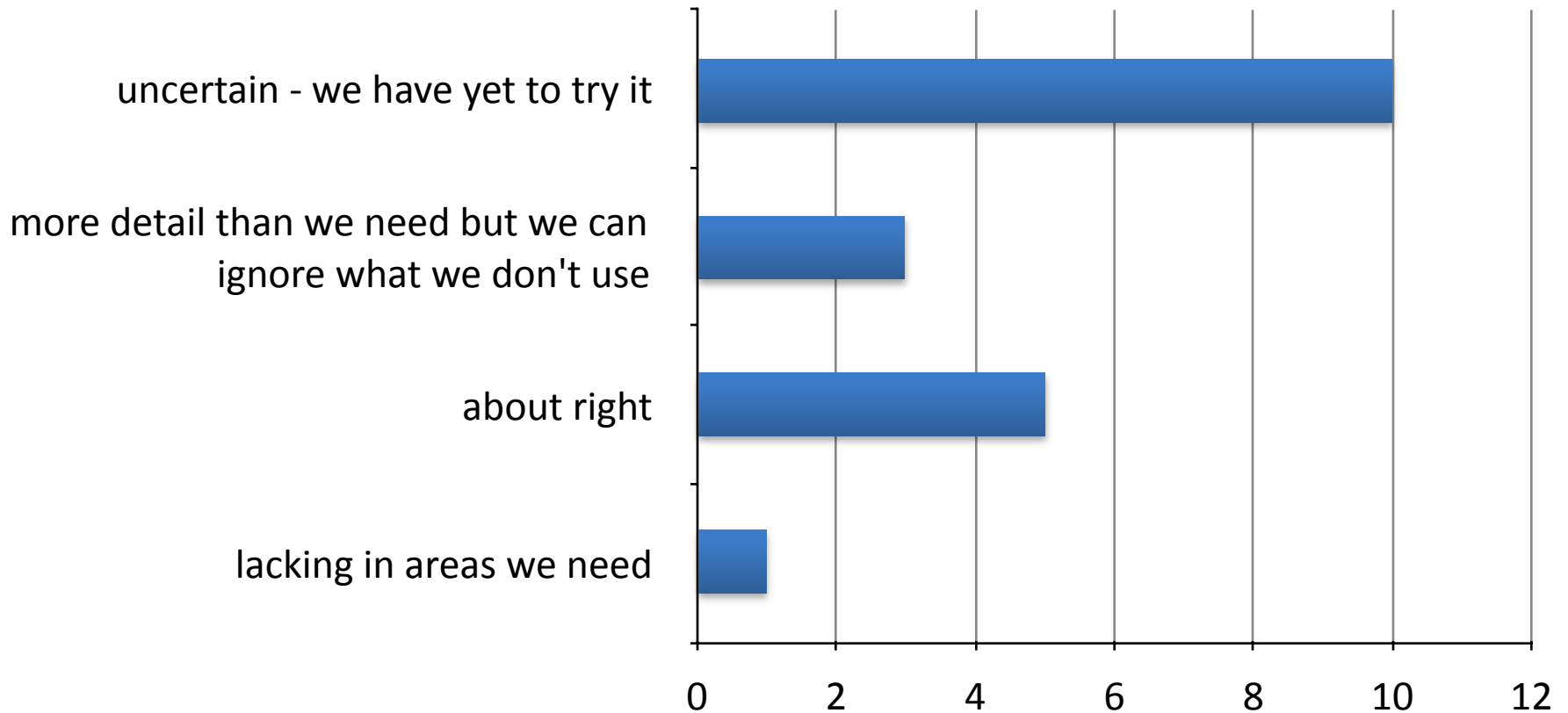
- Have you included other established ontologies?



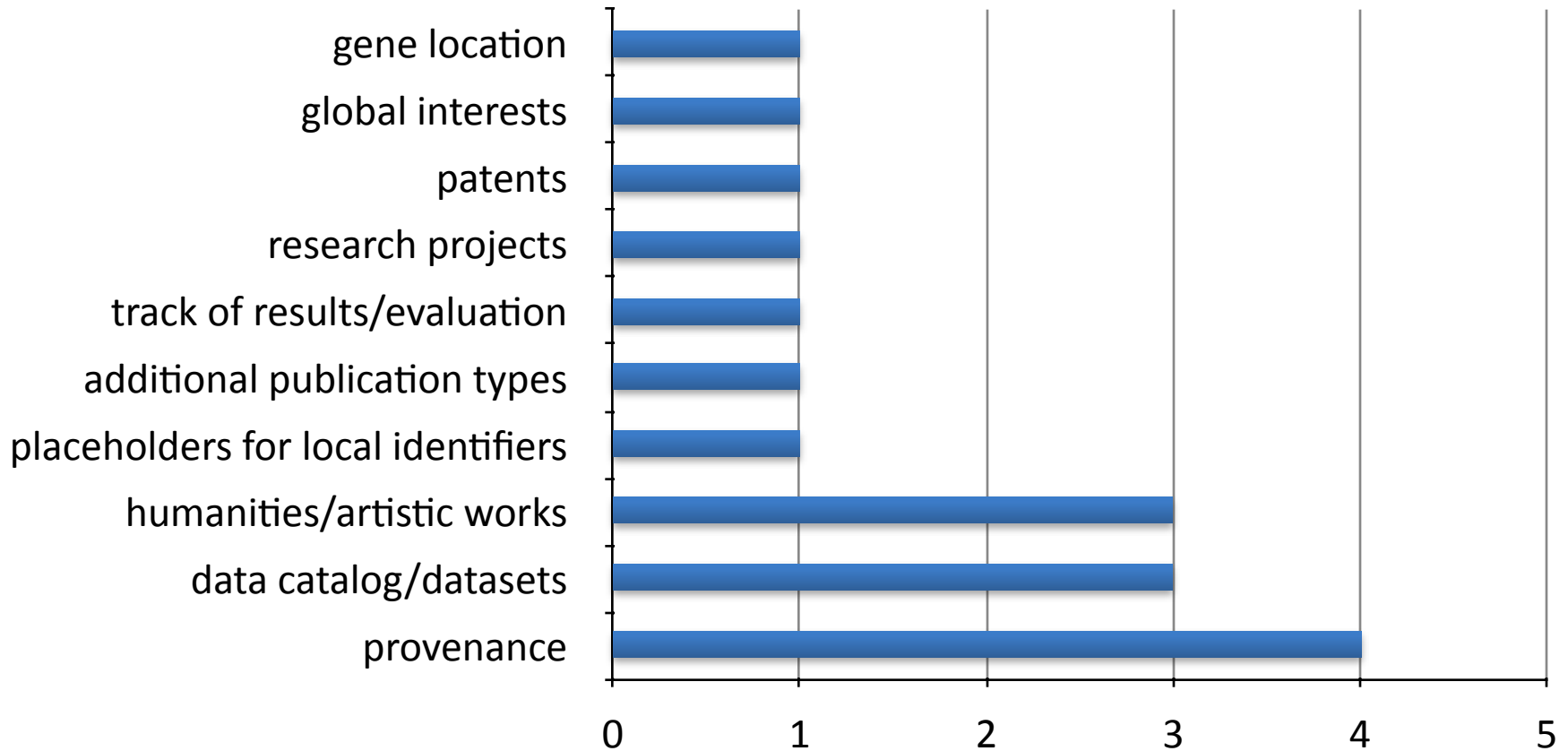
- Have you made any local customizations or extensions? (range: 5-57 classes, 5-100+ properties)



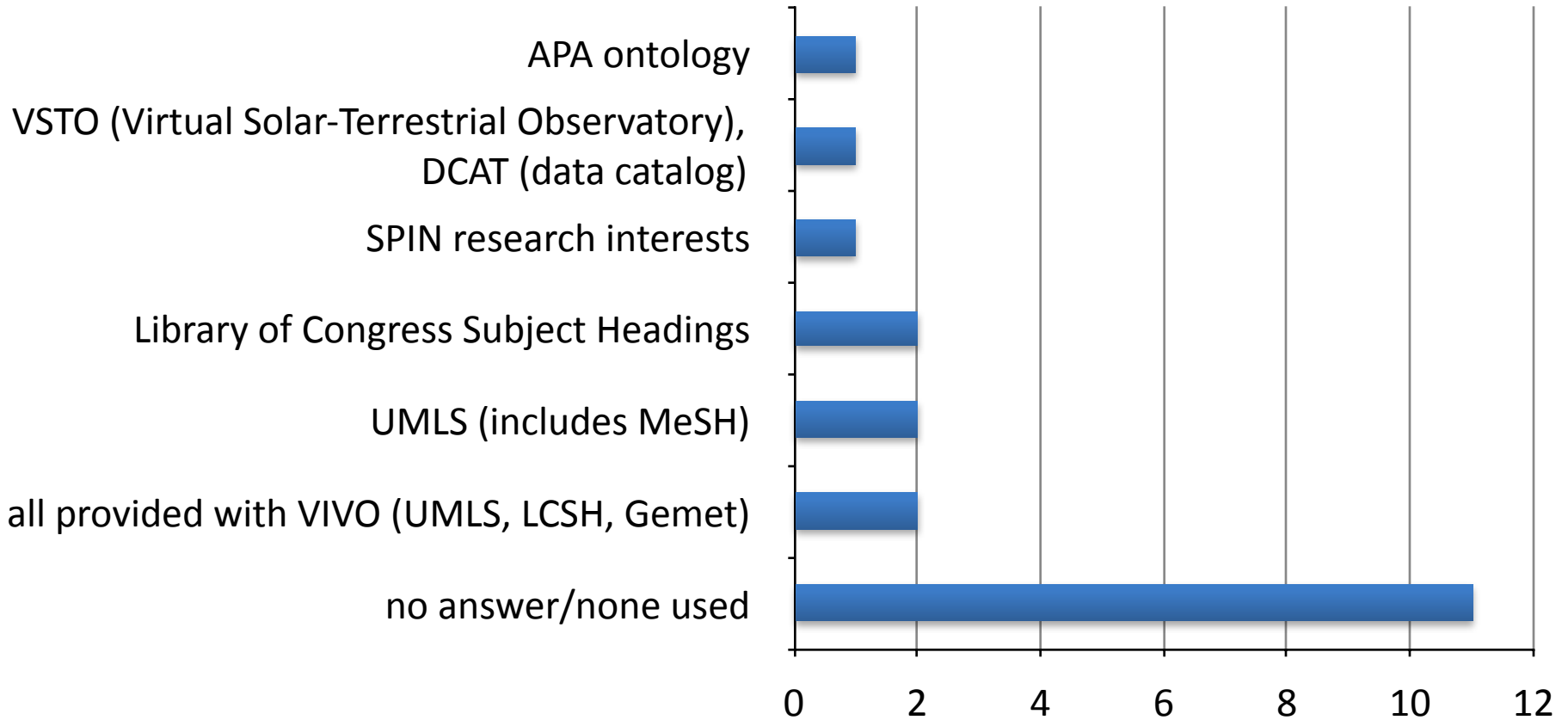
Rating the VIVO-ISF 1.6 ontology



Desired areas for ontology extensions and/or inclusion of other existing ontologies



External vocabularies linked to VIVO



What content do sites care about?

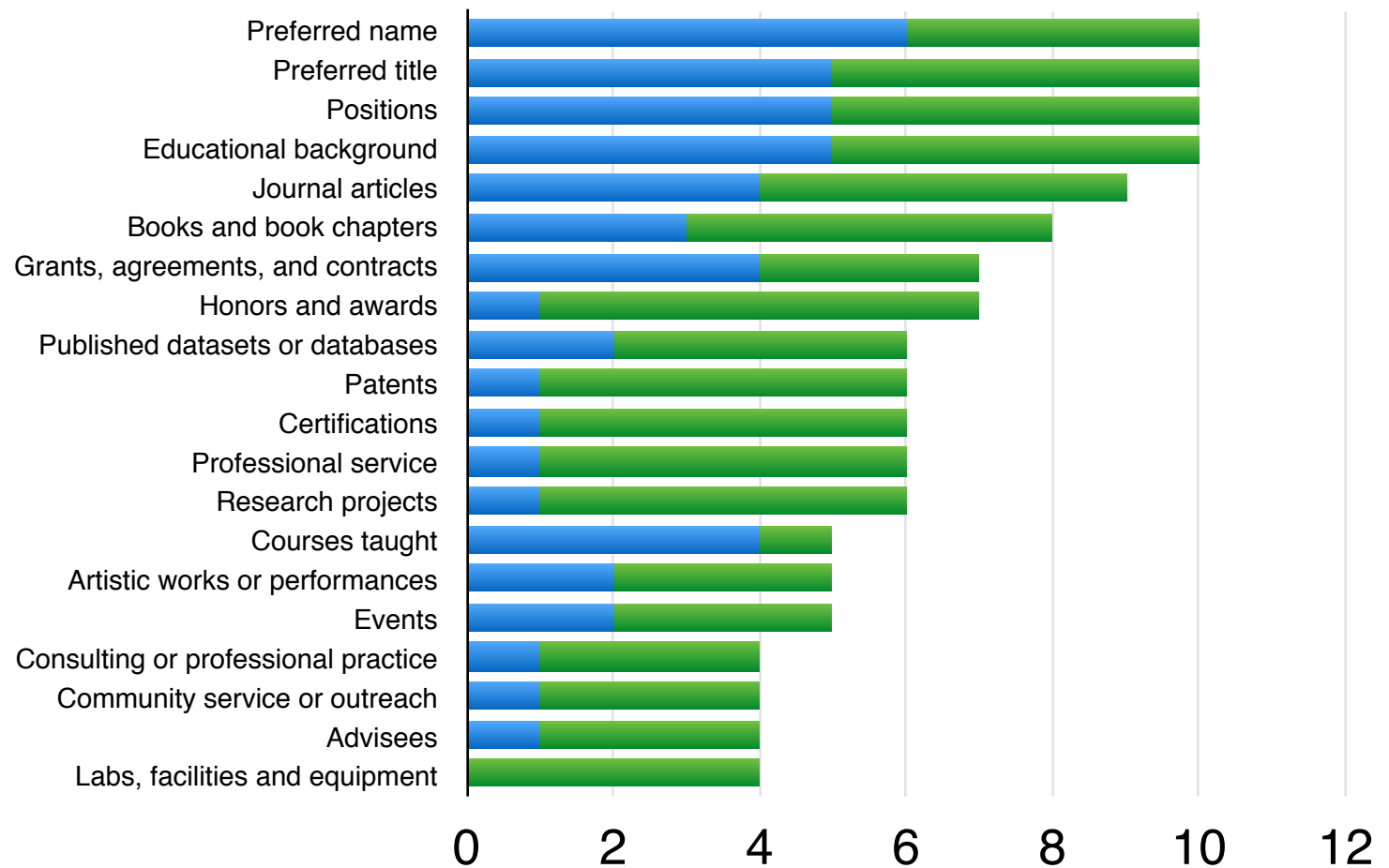
Section 3



Where can end users manually add data?

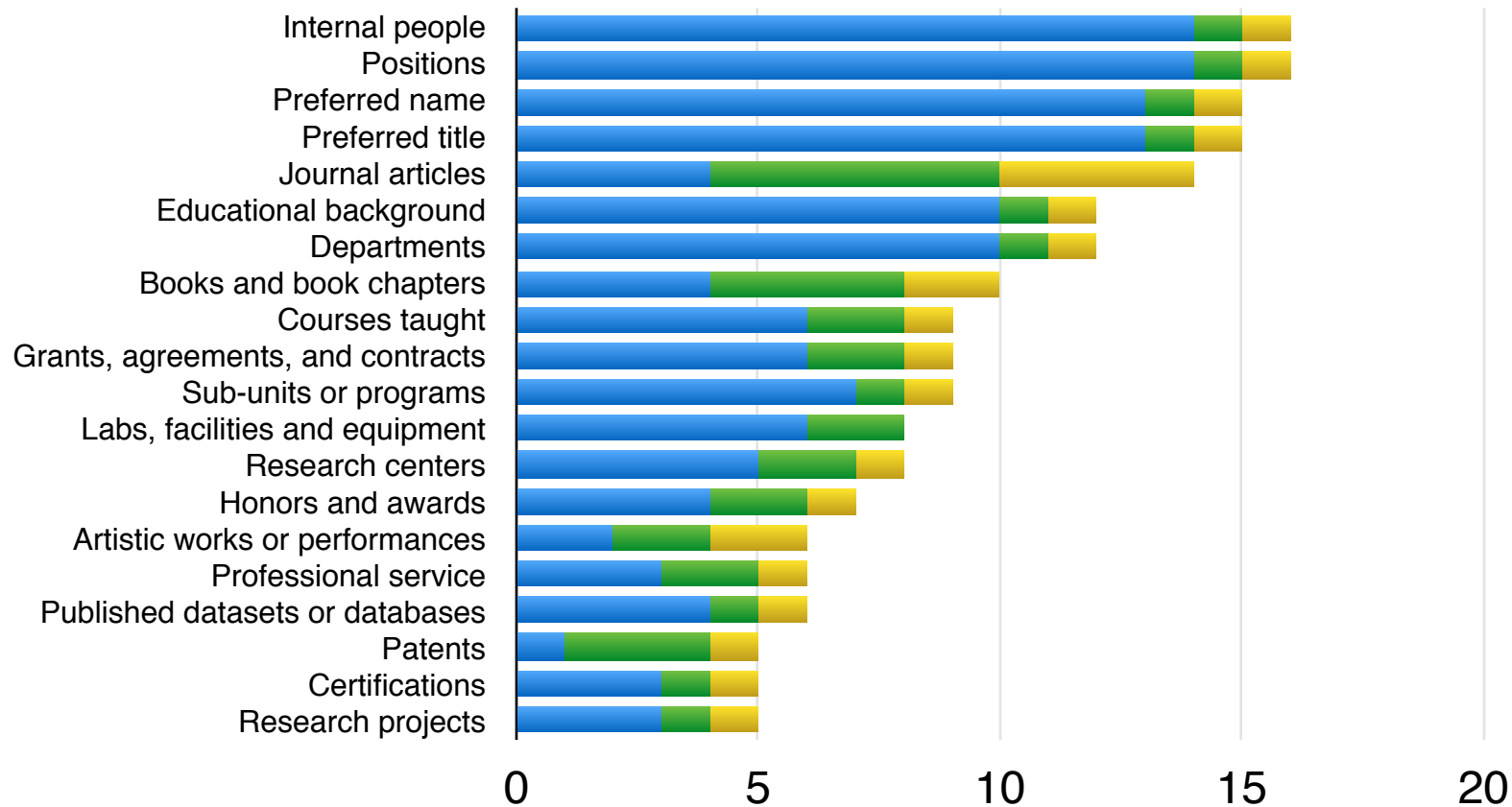
■ Upstream source

■ Directly within VIVO



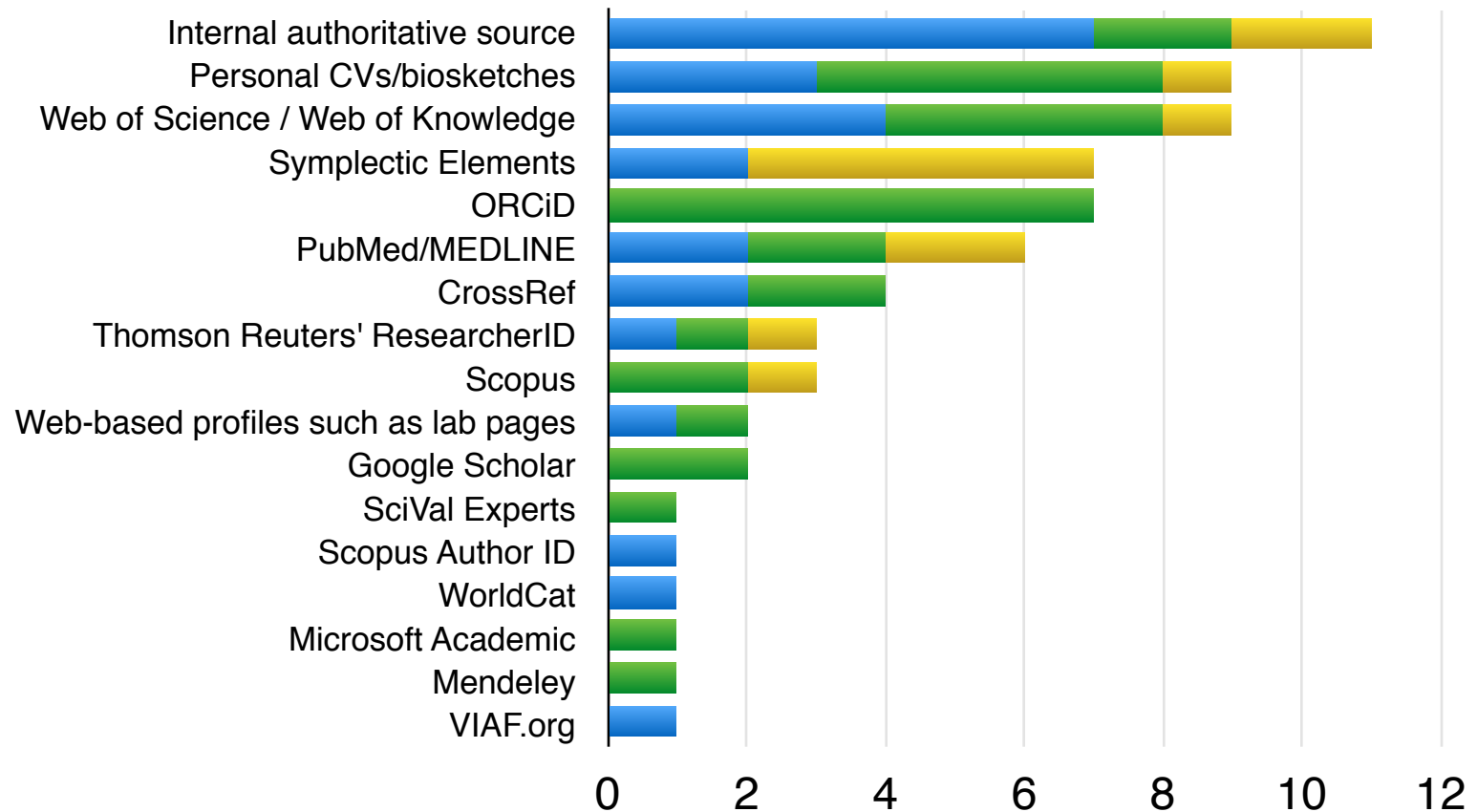
Which data are currently in your VIVO and what's the source type?

■ Internal ■ External ■ Both

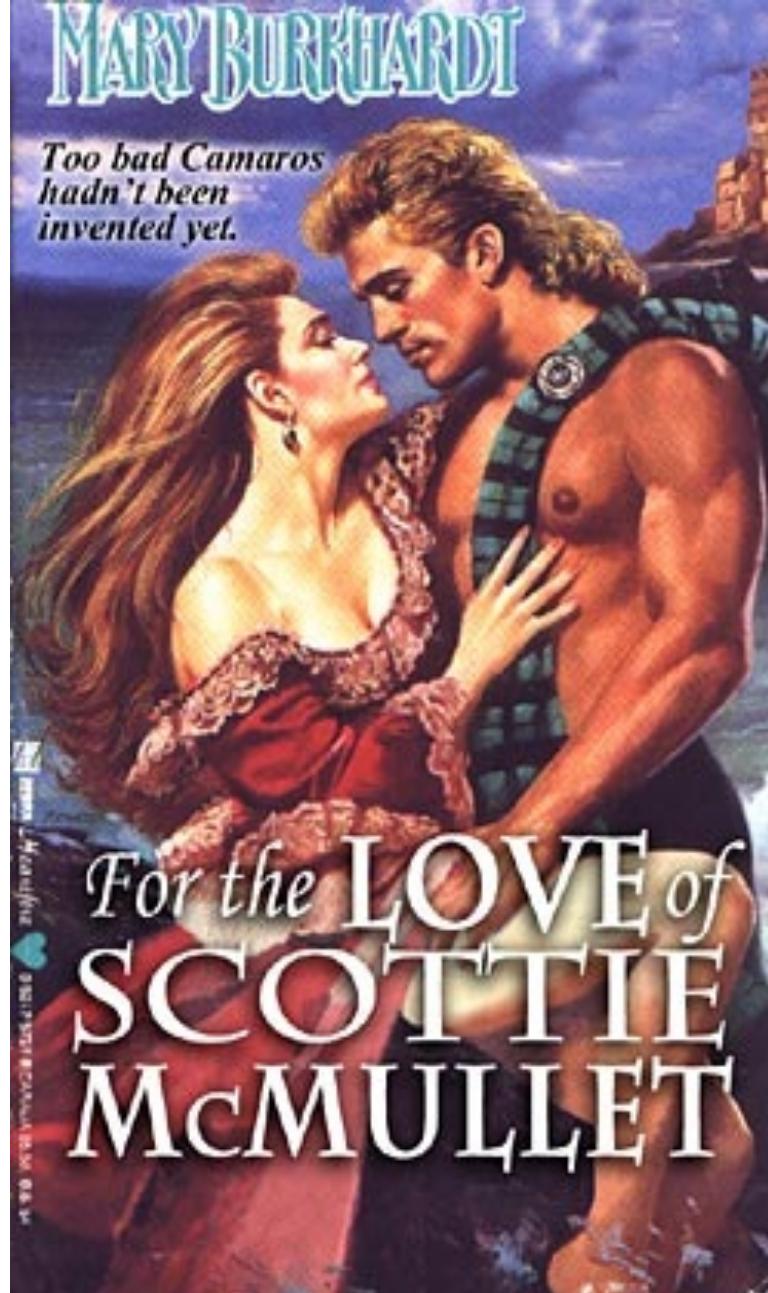


Which sources do VIVO sites use for publications and for what purpose?

■ As a source for ingest ■ To disambiguate authors ■ Both purposes



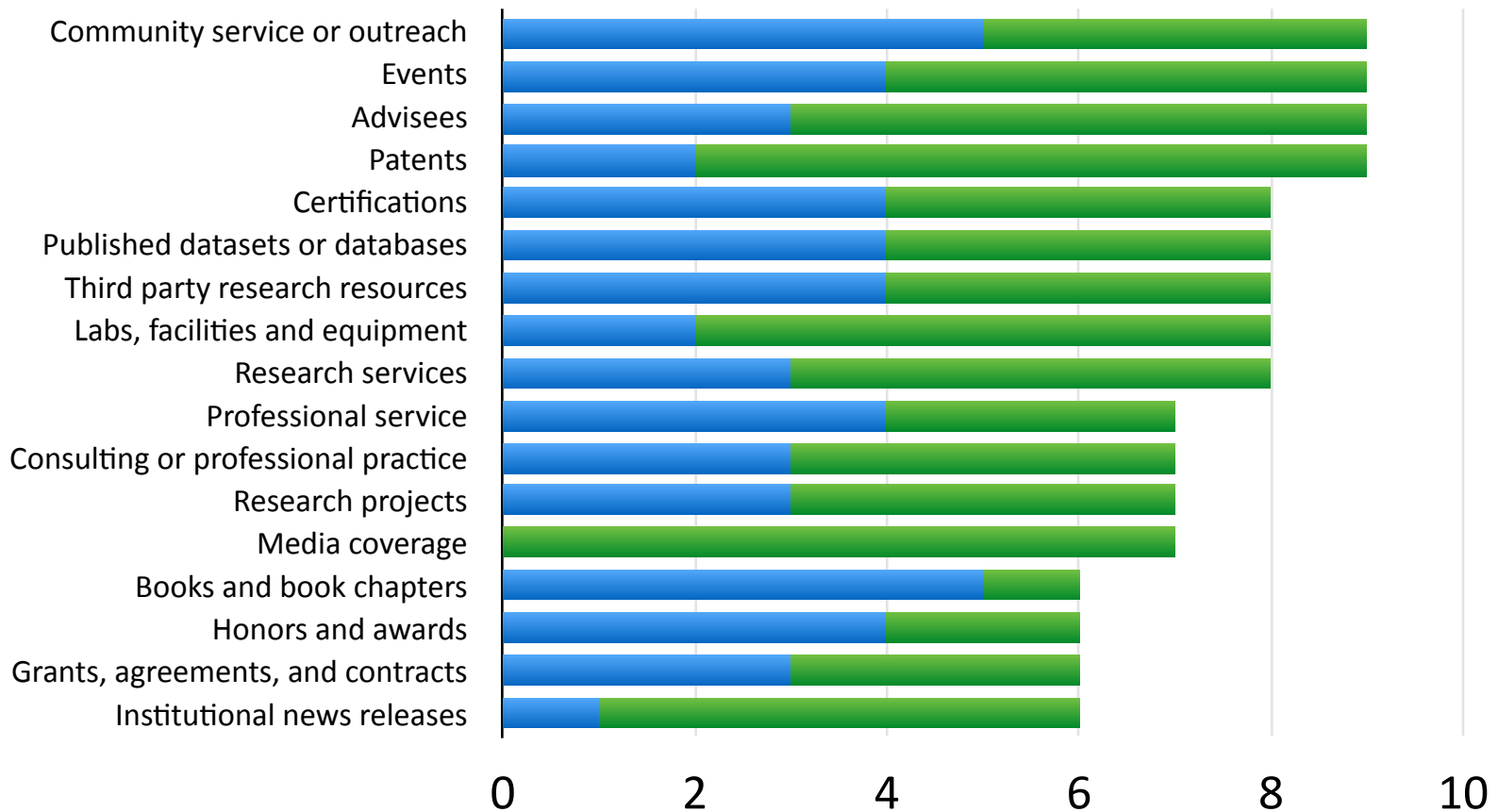
**Institutions'
unfulfilled
desires for data**



Institutions' unfulfilled desires for data

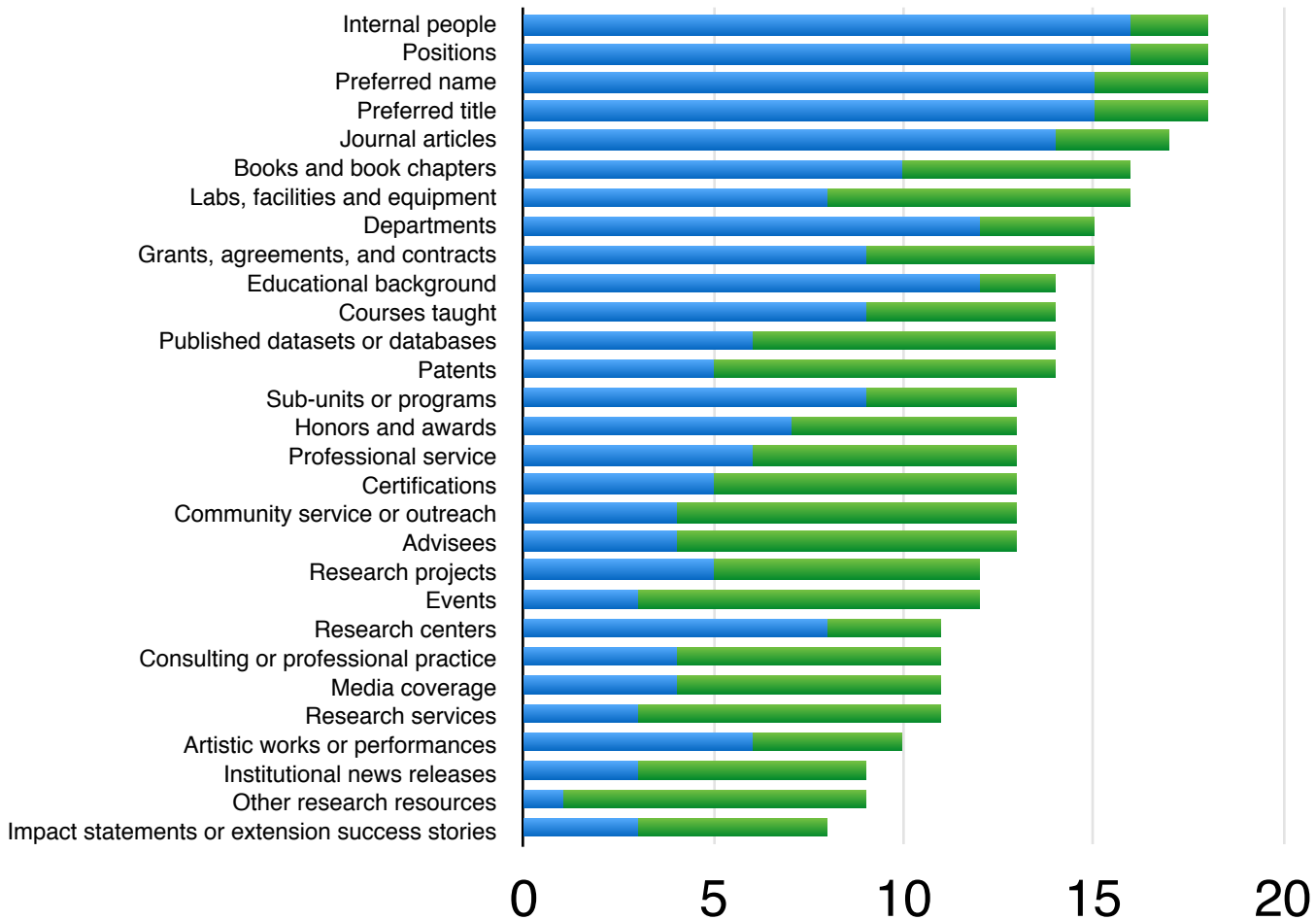
■ No, but we're working on it

■ No, but we're interested



What data do institutions have and want, combined?

■ Have it ■ Want it



State of the VIVO community

Section 4



How was developer time spent during the first year of VIVO implementation?

Learning Semantic Web
37%

Implementation Tasks
48%



Implementation Challenges (cont'd)

Political: getting access to data from institutional sources? **Not at all** / **To a limited extent**

Addressing poor performance of the application?
Not at all / **To a limited extent**

Getting adequate resources from administrative leadership to proceed? **To a limited extent**



Implementation Challenges (cont'd)

Theming and other changes to the templates?

To a large extent

Understanding VIVO's semantic data model?

To a large extent / Extremely

Cleaning up dirty data from institutional sources?

Extremely

Ingesting data using Harvester or other similar tools?

Extremely



How effective have each of the following methods been for solving implementation roadblocks?

Trial and Error? **Highly**

Email to the mailing lists? **Moderately**

Searching mailing list archives? **Moderately**

Contacting developers off-list? **Moderately**

Implementation/development call? **Moderately**

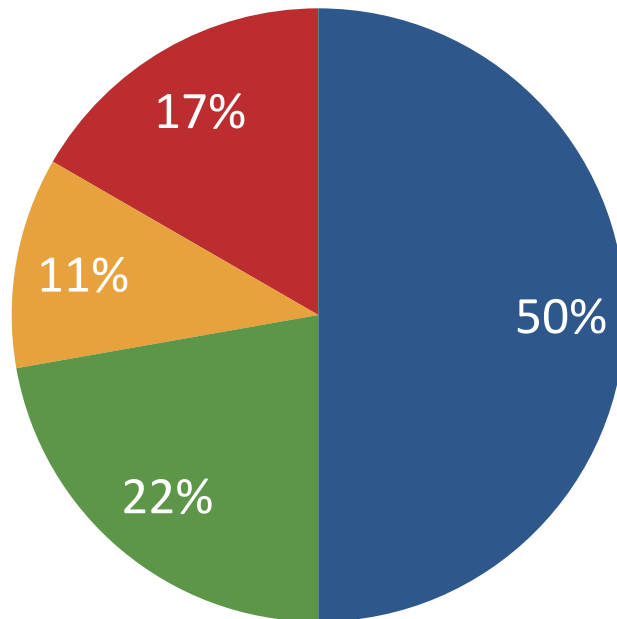
Using the wiki? **Moderately**

Posting to Stack Overflow? **Not at all**



VIVO Documentation

- Decent
- Okay but not great
- Poor
- Has some serious gaps



Inspiring Docs in Open Source?



Q: Should VIVO docs describe a tool, an application, or a platform?

Documentation Priorities

Self-guided tutorials for data ingest? **Very important**

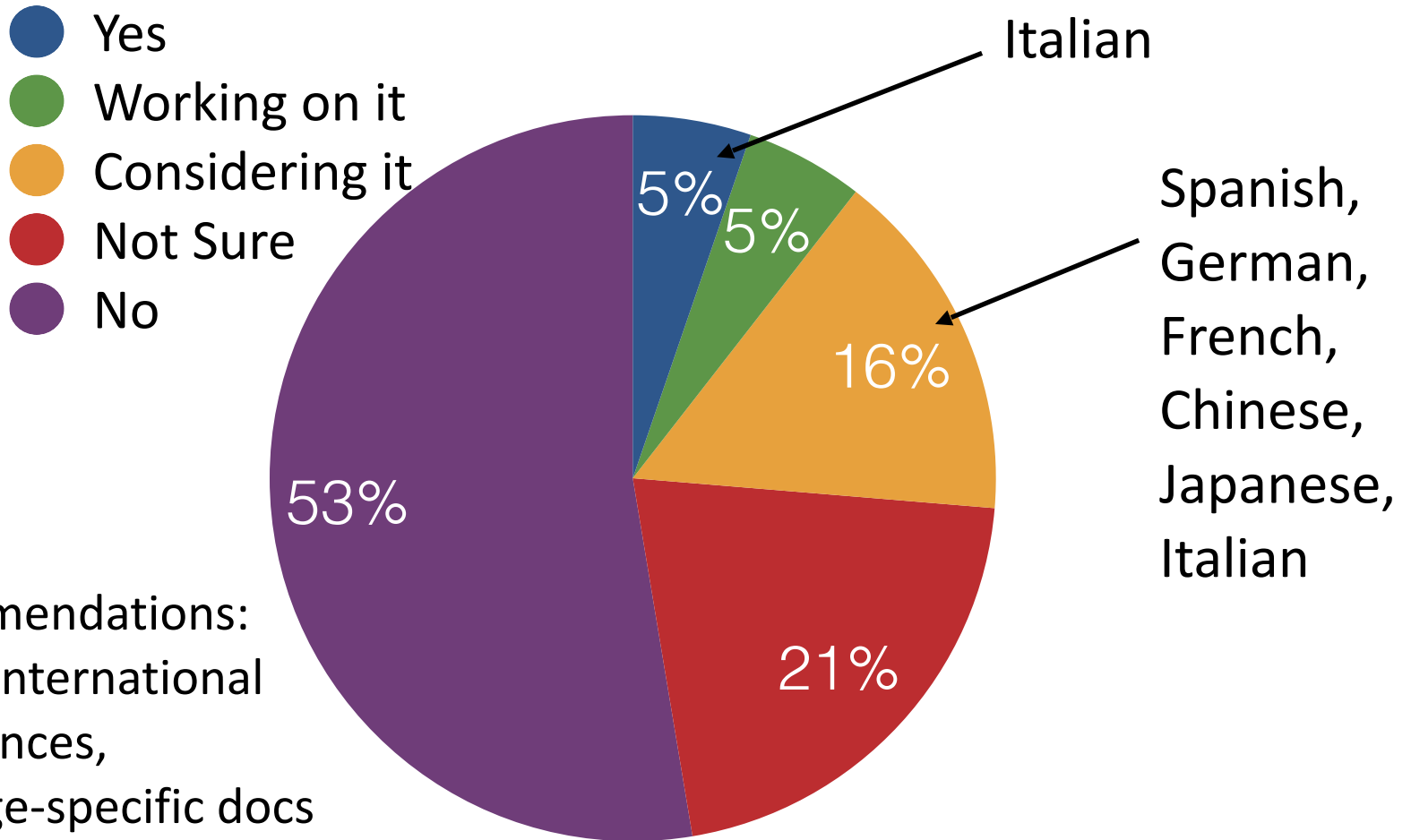
Standard data set to work with? **Very important**

White paper highlighting use cases for how VIVO can benefit an institution? **Very important / Important**

Detailed list of standard application and server configurations? **Important**



Internationalization



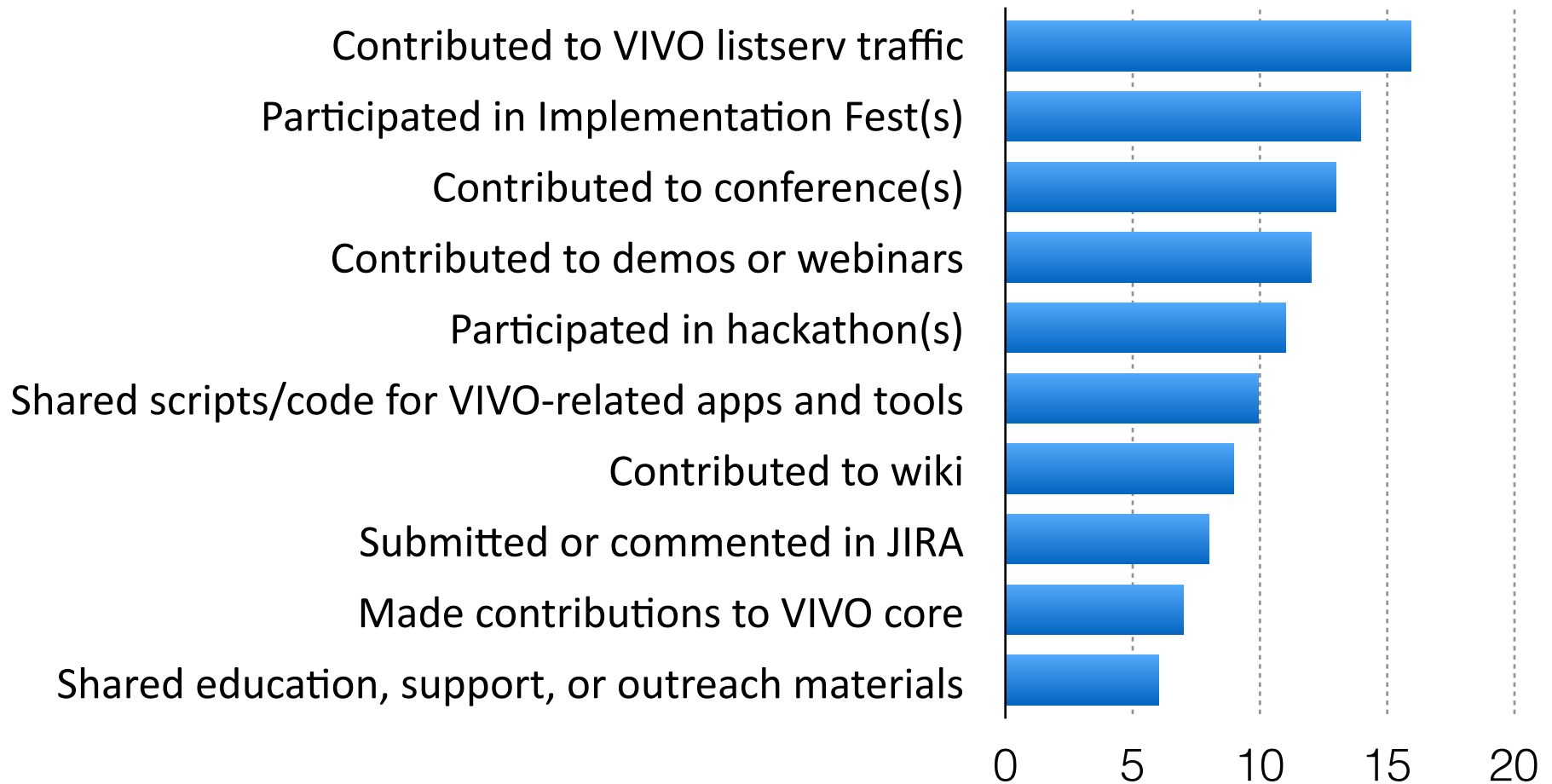
Contributing to VIVO

90% of responders have
contributed to open source
projects

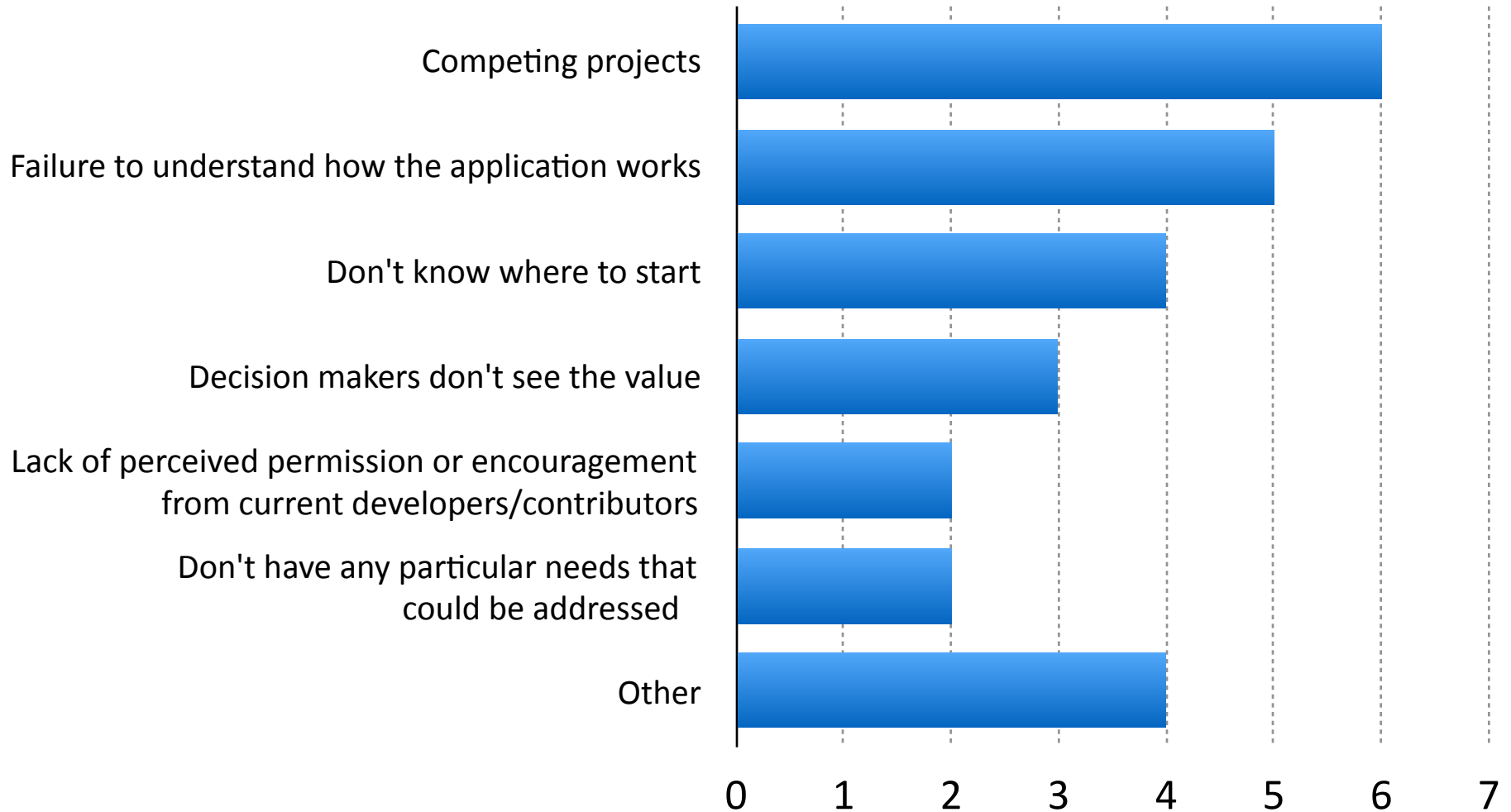
>50% using GitHub



Contributing to VIVO



Why aren't we contributing to VIVO code base?



Outreach / Competing visions for VIVO

Section 4



How frequently do you receive the following feedback about your VIVO?

- My data is wrong
- Key data of mine is missing
- I don't want VIVO to replace my institutional website
- I'm not allowed to edit certain data from external source
- I don't want my data public
- I wish I could easily repurpose my VIVO data (e.g. populate a LinkedIn or ORCID profile, or lab site)

Constantly
Never
Infrequently
Sometimes
Occasionally



For your institution, what is the importance of the following visions of VIVO?

1. A vehicle for building intra-institutional networks of people
2. A vehicle for building inter-institutional networks of people
3. VIVO as a tool for awareness of current activities
4. A loose federation of many lightweight, extensible, and creative apps on a common semantic framework
5. A flagship product that needs to be easier to adopt, populate, and grow
6. A discovery front end for repositories of documents & data
7. VIVO as a tool for archiving scholarly activities over time
8. A vehicle for building networks of research data
9. A tool to help universities and government agencies meet open government mandates
10. A home and integration point for a virtual organization
11. A reference implementation for the VIVO-ISF ontology as a standard for international data exchange



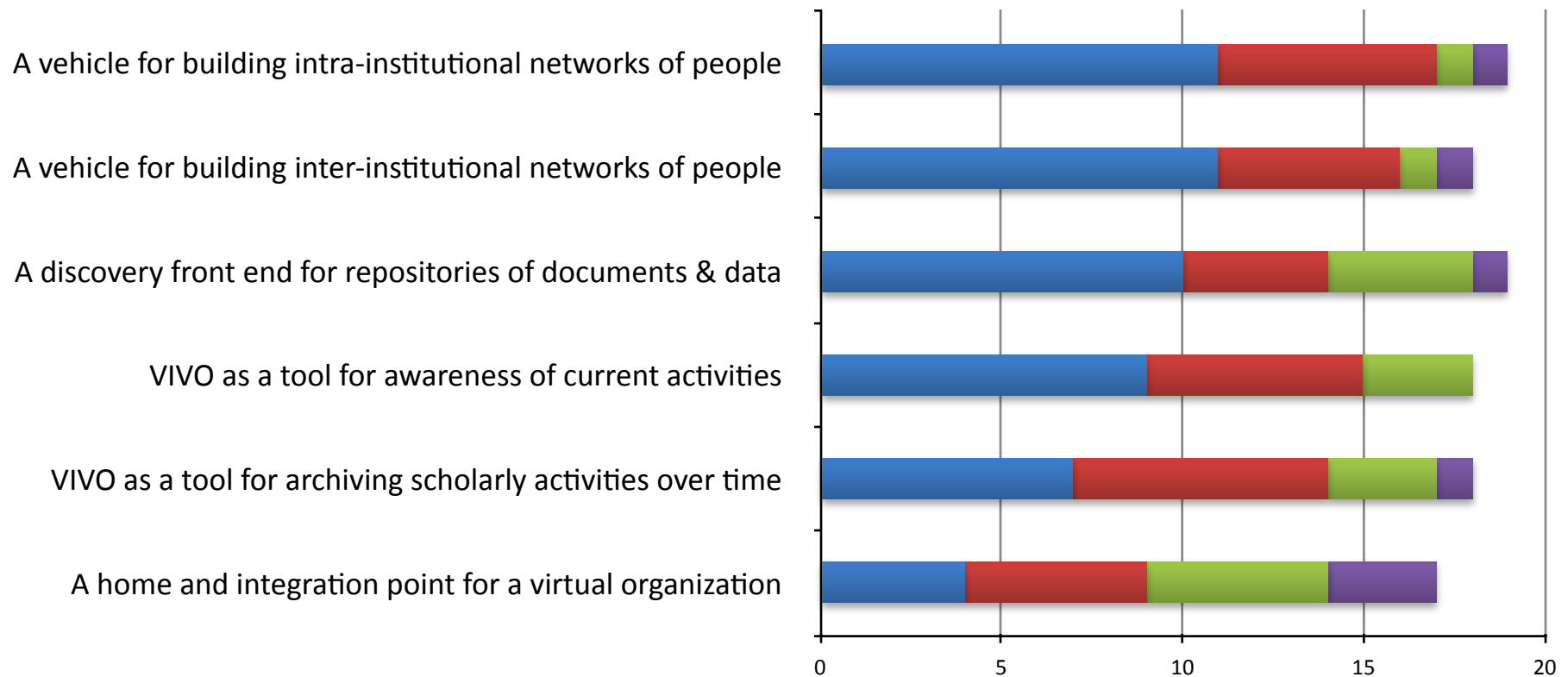
A vehicle for building intra-institutional networks of people

- The metaphor of VIVO as vehicle may be attractive because:
 - Vehicles are generally simple to use.
 - Vehicles do something useful rather than serve its own end.
- VIVO is not an archive or repository.
- VIVO's most popular purpose – current awareness about internal researchers' activities.
- Intra-institutional network > Inter-institutional network
- Researchers > Researcher Data



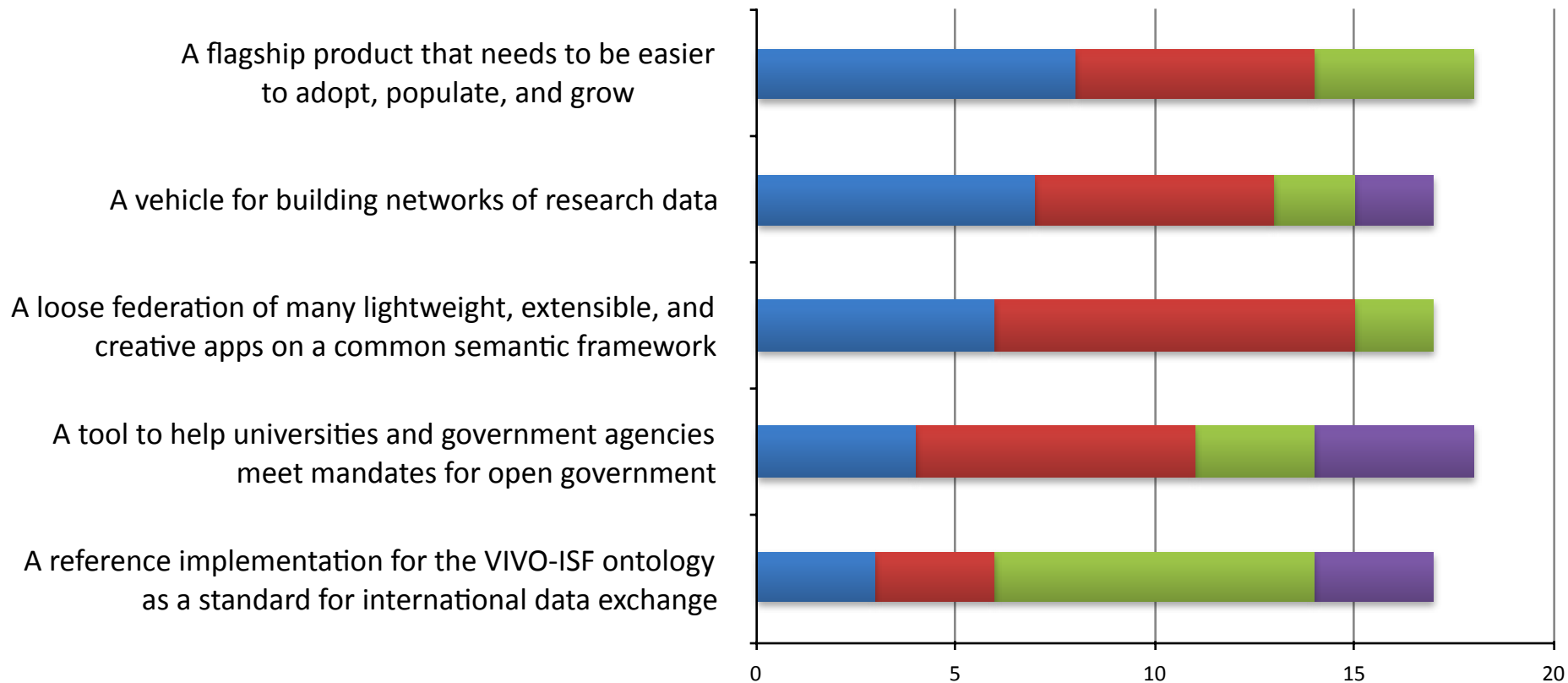
For your institution, what is the importance of the following visions of VIVO? (internal)

■ Indispensable ■ Valuable ■ Moderately useful ■ Not at all

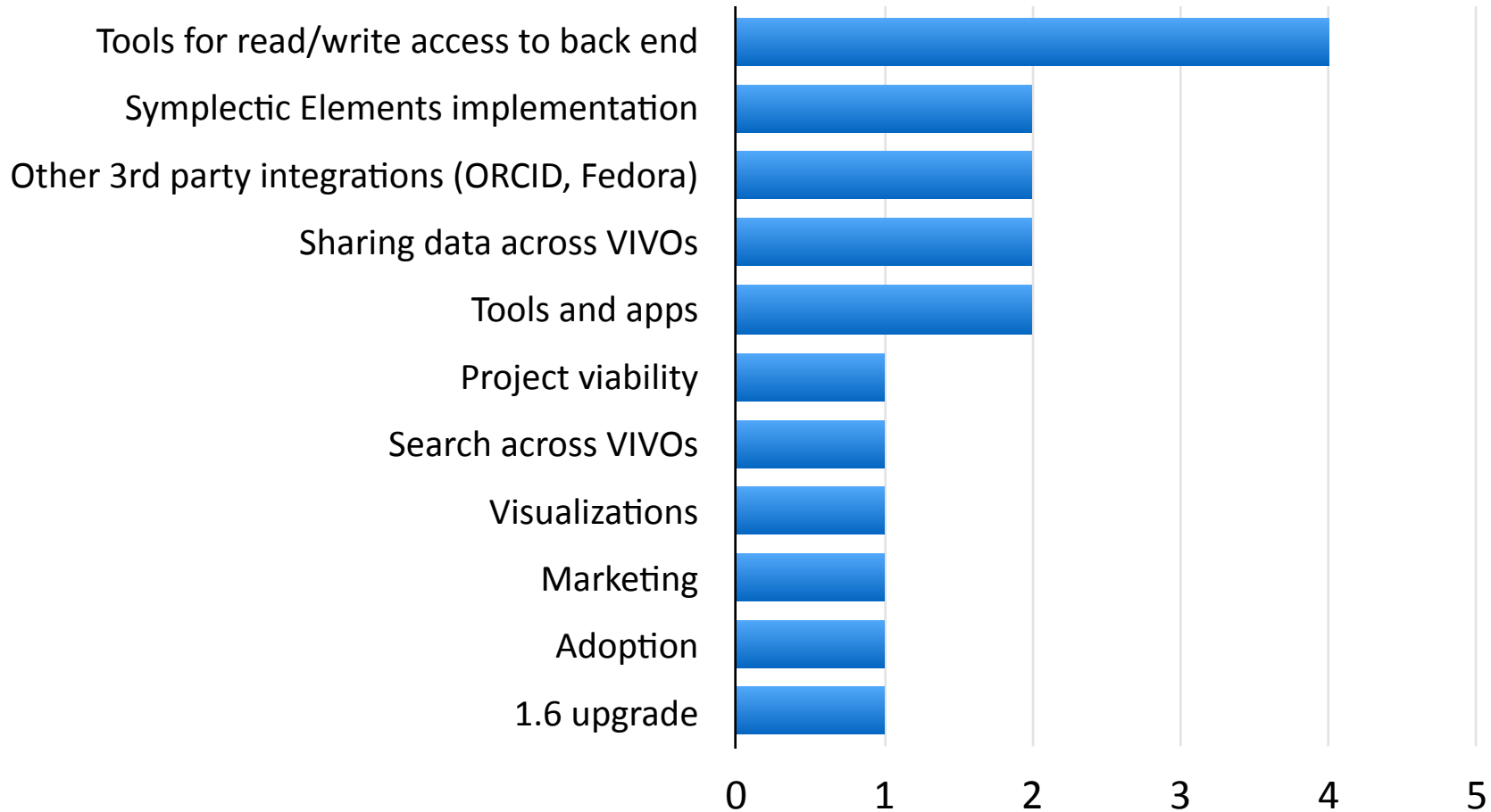


For your institution, what is the importance of the following visions of VIVO? (external)

■ Indispensable ■ Valuable ■ Moderately useful ■ Not at all



Themes in the answers to the question, “What is your site’s priority?”



Would the following messages be successful in marketing VIVO to your profiled users?

- VIVO promotes your work within your discipline
- VIVO promotes your work to granting agencies
- VIVO promotes your work to prospective students and post-docs
- VIVO saves staff and researcher time
- VIVO meets key administrative needs such as business intelligence
- VIVO helps researchers find collaborators
- VIVO promotes your institution's work and makes you more competitive

Strongly Agree

Neutral

Agree

**Thank you to everyone who took time
to share your site's perspective.**

