Frequently Asked Questions (FAQs)

SciVal Spotlight

- **Are self citations included in the co-citation analysis?**
  Yes, self-citations are included when generating the co-citation model.

  These are important (as are all citations) for determining the structure of science because most self-citations do establish a research flow and topic grouping for an individual or group. The argument that some people have against self-citations is their perception of self-inflation when counting citations for impact, however, we are merely using them to establish the structure of the article clusters.

- **What is meant by the most 'highly-cited' articles that are used for the co-citation analysis?**
  The co-citation analysis is done on the cited references in current papers (those published in the year the map is being generated for). We use a threshold to limit the co-citation analysis to the most highly cited papers. To give you some approximate numbers, if we take one year's current papers (e.g., those published in 2009), they will reference over 15 million different papers. We do not want to use all 15M in the model; we only want to use those that really determine the current structure of science. Therefore, we use a threshold to use only those cited references that really contribute to this structure, and we end up with around 2M cited references.

  The threshold is:

  1. Fewer than 100 citing papers AND
  2. Depending on the age of the publication, one of the following:
     - Age < 2 years, citations ≥ 2
     - Age = 2 years, citations ≥ 3
     - Age = 3 years, citations ≥ 4
     - Age > 3 years, citations ≥ 5

  Note that we only count citations received in the current year. This is not a cumulative number of citations received over many but only those received in the current year (e.g., 2009). This limits the set of cited papers to those that are important to current science - not past science.

- **What does the fractionalized count indicate, and how is it calculated?**
  Current articles are assigned to article clusters based on the location of their reference articles. Therefore, an article's references dictate how it is assigned to more than one article cluster, and each article is given fractional counts that sum to 1.0. These counts can be summed by article cluster (paradigm), competency (EC or DC), author, institution, or market.

  Example:
  If there are 1,000 articles in DC7, but every article is fractionally assigned to one other competency (with 50% of the references residing in DC7 and 50% in another competency), then the fractionalized article count for DC2 would be 500. If the average article uses 25 references, hypothetically, this paper could be split across 25 different article clusters (and even competencies) with a total of 0.04 fractionalized articles in each. Naturally, there is some relationship between the references, so this is the most extreme scenario.

- **What is Publication Leadership?**
  The institution with the greatest Relative Article Share (RAS) exhibits Publication Leadership. RAS is a relative measure of the volume or quantity of publications an institution (or author) has produced in the past 5 years, compared to other institutions (or authors) within a particular competency. To calculate RAS we determine the institution's publication level (over a 5-year period) and divide it by the largest publication level of any other university.

Example:
If University A has published the most articles (e.g., 30) in the competency, then all other institutions contributing to the competency will be compared against University A (the Publication Leader). If we assume that University B has published 10 articles in this competency and University C has published 15 (the second highest number of articles published after University A), then we can calculate the RAS for each of the three universities as follows:

- University A = 30/15 = 2.00
- University B = 10/30 = 0.33
- University C = 15/30 = 0.50

With an RAS of 2.00, we can say that University A published two papers for each one that the next most active institution in the group (University C) is publishing.

- **What is Reference Leadership?**
  The institution with the greatest Relative Reference Share (RRS) exhibits Reference Leadership (RL). RRS is a relative measure of how influential an institution or author (i.e., number of highly-cited references written by an institution or author) is within a particular competency.

  Recall that each co-citation cluster contains:

  - Articles published in the past 5 years
  - Highly-cited references

  Relative Reference Share is the number of highly-cited reference articles that are authored by that institution divided by that of the closest competitor.

- **What is State-of-the-art (SotA)?**
  SotA helps to measure *thought leadership* by indicating the recentness of articles cited by the institution's articles within an EC or DC. The State-of-the-Art measurement varies around zero (average median reference year); positive values indicate that the institution is citing more recent work within the competency than the world as a whole, while negative values indicate that the institution is citing older work than the world as a whole.

  Example:
  Average median reference year = 5

  - Author A has a SotA of 2.0
  - Author B has a SotA of -0.2
  - Author C has a SotA of 0.8

  Author A is, on average, citing more recent publications (references are 3.0 years old) than the average median reference year of the group, and with the most positive SotA, she referenced the newest work of all authors in the group. Author B is referencing slightly older papers (references are 5.2 years old) than the average of the group, and he has cited the least recent papers of all authors in the group.

Frequently Asked Questions (FAQs)

SciVal Spotlight

- **When determining whether an institution has a Distinctive Competency, why is a sliding scale of Relative Article Share used?**

  The sliding scale relates to the size of the institution. A small institution with less research intensity would only need to have a Relative Article Share of 0.2 (*i.e.,* compared to the institution with the most papers in this group of article clusters, they would need to publish 1 for every 5 that the leader is publishing, so 1/5=0.2). For a large, research-intensive institution, they would need a Relative Article Share of 0.6 (*i.e.,* compared to the institution publishing the most papers included in this competency (Publication Leader), they would need to publish 3 for every 5 that the leader has published, so 3/5=0.6). This sliding scale allows us to take into account the varying size of institutions, so that even smaller ones will be able to visualize their research competencies.

- **How is 'Growth' defined?**

  Growth is the average annual change in the number of articles (using fractional counts) by an institution in a competency over the 5-year publication window and is calculated over a sliding 2-year window.

  Example:
  When Y is annual market size, growth = ((Y4+Y5)-(Y1+Y2))/6*Y3

- **How are the competencies (DCs and ECs) ordered?**

  The numbering of the competencies is related to the market size; however, there are actually two types of market size.

  - **Region size:** The market size you see in SciVal Spotlight is called *region size* (i.e., "Country", (geographic) "Region", or "Global"). The *region size* is the sum of all articles from the institutions in the specified region (de-duplicated for collaborative articles), but not all institutions' papers in the *region size* are included (e.g., we exclude institutions publishing less than 50 papers per annum).
  - **Raw market size:** The *raw market size* is simply the sum of all fractionalized articles in a competency, regardless of the institution publishing them. We do not display this value because it is not sensitive to the region. Nevertheless, it is the value used to order the competencies from largest market size (EC1 or DC1) to smallest (the last competency on the list with the highest number – e.g., EC112). It is also the value used when calculating the market size threshold for becoming either an EC or DC.

Frequently Asked Questions (FAQs)

SciVal Spotlight

- **Are papers associated with authors or institutions?**
  Although we would like the papers to travel with the author, currently, the papers are matched to the institution. Eventually, we will be able to assign papers to authors, so their intellectual property will move with them to their new university.

- **Why do some authors show "0 documents in Scopus" in Spotlight?**
  Spotlight in links to the Scopus Author-ID that we have listed in the dataset used to generate the map. Meanwhile, Scopus is continuously updating and improving the profiles which means new Author-IDs may be created for some researchers. Currently, when an author's profile changes, the old Author-ID reverts to "0 documents", which is what one sees in Spotlight. These changes will be fixed each time the data in Spotlight is refreshed.

- **What if an article belongs to multiple institutes?**
  If a publication is a collaborative effort between, for example, three institutions then each institution will receive credit for that paper. This does result in a duplication of papers and is an accepted method for attributing papers.

- **How can an institution have none of the Top Authors in their own competency?**
  It can be related to the way we use author profiles in combination with author movements (i.e., when an author moves from one institution to another or is on sabbatical). Occasionally there is a mismatch between the institutional coding on the papers and the institutional coding of the authors, and we are currently working on a solution.

- **Why would an institution be ranked #2 in their own DC?**
  Rank is based on article output, so the institution may be either the Reference Leader (RL) or have an SotA higher than the Publication Leader (and RAS>0.8).

- **How are competency keywords selected?**
  It is important to note that the keyword phrases indicated for each competency do not affect the competency's formation. The competencies are formed by looking at a university's Relative Article Share (RAS) within individual article clusters (paradigms) that were created during the co-citation analyses (i.e., looking at the frequency at which reference articles occur together within all global publications of the map year). The keyword phrases are used to describe the area of research and are derived from the articles that contribute to the competency. The keyword phrases are not author keywords or controlled vocabulary; we determine which 100 two-word phrases ("keywords") have the highest entropy across the titles and abstracts of the articles that comprise the competency.