**SYNTHESYS+ VA Call 1 Review**

**Review Board:** Arturo H. Ariño, Joseph T. Miller, Pamela S. Soltis

## CHARGE

The SYNTHESYS+ Plus Virtual Access project, hereafter referred to as VA, has completed its first round of funding calls, and the VA team seeks from the Review Board (RB) an analysis of the work and recommendations to take forward to the next calls.

The Review Board was tasked with evaluating the process used in Call 1 for SYNTHESYS+ Virtual Access in advance of opening Call 2. The Review Board considered the following information in its evaluation:

- Summary from the VA Board, the SYNTHESYS+ VA website (https://www.synthesys.info/access/virtual-access.html),
- the RIO paper describing the VA process[1], and
- Appendix 4 (SYNTHESYS+ Virtual Access scoresheet Call 1).

The Review Board was specifically asked to address the following questions and offer suggestions for improvement, if any:

1. *Is the process as it worked in Call 1 fair and unbiased?*
2. *Does the process we followed fulfil the brief in the SYNTHESYS+ grant agreement?*
5. *Do we have adequate reach to the requesters, if not how could this be improved?*
6. *Does the balance of requests show evidence of reaching new communities of users?*
7. *Is the balance of requesters wide enough geographically? Suggestions for ways in which we might improve/expand this?*
8. *Any comments on the VA Coordinator role? Single vs. shared role, scope of work?*
9. *Any other feedback you might want to give us to help improve VA Call 2 in early 2021.*

Here, we provide an overall review of the VA program and then address the specific questions. Further analysis that was used in our assessment is provided in Appendix 1.

## CONSIDERATIONS ABOUT THE REVIEW

### *Overview of VA as described*

SYNTHESYS+ Virtual Access seems primarily a prioritization exercise: on-demand digitizing of the most wanted or useful collections for immediate public availability as Digitally Accessible Knowledge (DAK)[2].

The targets are collections held within the SYNTHESYS+' Taxonomic Access Facilities (TAFs) that collectively represent at least 1/3 of the natural history collections holdings worldwide.

The prioritization is intended to fulfill demand and maximize scientific returns through use of the DAK assets.

---

[1] Hardy, H., Knapp, S., Allan, L., Berger, F., Dixey, K., Döme, B., ... & Kiel, S. (2020). SYNTHESYS+ Virtual Access-Report on the Ideas Call (October to November 2019). http://dx.doi.org/10.3897/rio.6.e50354

[2] Sousa-Baena, M. S., Garcia, L. C., & Peterson, A. T. (2014). Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions*, *20*(4), 369-381. https://doi.org/10.1111/ddi.12136

Through two public calls described as open to anyone, VA intends to allocate 1M€ funds to DAK-gaining projects in the most efficient way. Funds are to be spent at the DAK-supplying institutions, which are a subset of the SYNTHESYS+ consortium.

Five projects were approved during the first call, using 50% of resources.

***Review approach***

A general description of the VA process and a list of candidates for the initial call has been made available. Also, the summary descriptions of the approved projects have been made public. Therefore, the outcome of the call can be mapped against the selection criteria and post-fact drivers can be examined.

As VA appears to hinge on a prioritization exercise intending to maximize returns and efficiency, it seems logical to seek some metrics evaluating this performance. Digitization involves a cost and expects to produce DAK of varying value, but the process also requires the existence of assets upon which digitization investment can be made according to demand. Therefore, three tensors must be compounded in order to maximize the scientific returns:

- Existing assets that can eventually be digitized (e.g. unpublished collections),
- Objective knowledge gaps that need to be fulfilled (e.g. taxonomic, geographic voids),
- Demand for such DAK to be produced (e.g. user needs as recorded in the literature),

The resulting metrics must then be balanced against the available resources (i.e. funding), and the goal of the prioritization exercise should be to allocate such resources so as to maximize those metrics within the resource constraints. Therefore, the review should attempt to discover whether the process followed in VA aligns with that general direction.

***Issues with the review***

1. "Total cost" is stated and apparently these figures are to be financed. However, it is not clear whether these costs are the approved financing or the required financing (in the proposals). This might prevent evaluating the efficiency of the resource allocation.
2. ELViS is accessible and currently shows most information about each submission. However it marks many of them as "approved" which may be misleading as only five have (apparently) been actually approved for funding. Differentiating this "approval" for submission to panel from the actual approval of the funding should be desirable for streamlined analysis.
3. Bryozoa incorrectly identified as molluscs. No consequence for the review.
4. Names of submitters were "removed from the requests"—however they are fully accessible through ELViS—is this on purpose? As the names were available they were used in the review.
5. The concept "specimen" does not seem to be unambiguously defined. As this is a digitizing project, it should be clear which constitutes the digitizing unit (individual specimen, museum accession, collection unit) in order to estimate return efficiency.

***ASSESSMENT***

### *A. General Comments*

The RB appreciates that the VA is a novel way to structure data digitization and therefore had to create many new procedures and materials. Additionally the VA call occurred during the Covid pandemic which altered the landscape for VA but also provided a real-life example of why virtual access to natural history museums is necessary.

The VA team anticipated several risks to the project and preemptively planned the first VA call to ameliorate them. This included communication and administrative risks. This risk assessment was successful in identifying potential issues and the responses were successful. This is indicated by the large number of proposals and a good distribution across collection types. One component not covered in the document but should be considered by the VA team is the diversity of the requesters beyond data type. Were the requests distributed across institutional types and geographic areas of interest to SYNTHESYS+? If not, how can the communication plan be updated to improve this component?

The VA team also developed and implemented several types of surveys. These are valuable feedback for future funding calls. The VA team should be encouraged that most of the survey feedback was about the ELViS system which was in development. The VA team has fed these ideas back to the ELVis team and the system appears to be more robust for the second VA call.

### *B. Responses to Specific Questions*

*1. Is the process as it worked in Call 1 fair and unbiased?*

It certainly seems so. The workflow seems sound, and equal opportunities have been given to anyone. The call and the review were not biased, but the final outcome was not an equal distribution of funds because the program needed to strike a balance in the amount of available resources vs the need to maximize the returns. This seems to have been achieved at least from the quantitative point of view:

- three-quarters of the expected digitization mass could be achieved using 1/3 of the resources requested (although not all digitizations are equal!),
- sufficient diversity of stakeholders have been served.

That said, some concentration of resources has been observed at certain partners. This, in turn, might be a natural consequence of the different weights, of the nature of the collections held, and of the needs of users materialized through the variety of requests.

Future calls might consider prioritizing certain taxa and certain institutions. In addition, although the RB recognizes that funds are not available for hardware, it might be of value to consider expenditures for equipment to permit digitization at additional institutions that could provide access to important collections.

*2. Does the process we followed fulfil the brief in the SYNTHESYS+ grant agreement?*

The process appears to be consistent with the grant agreement.

*5. Do we have adequate reach to the requesters, if not how could this be improved?*

Probably: 38 countries were represented through the proposals. However, it could be expected that in most or all cases a starfish model was followed. More than half (17 out of 30) of the bids had members who were associated with at least one of the VA members, and all of the winning projects showed such inbreeding. Therefore, a sizable proportion of the bids actually originated within the consortium. It is thus reasonable to think that the variety of bidding persons, institutions, and countries originated in no small part from the personal contacts of the members of VA, encouraging and/or suggesting participation. But the fact that there were indeed proposals unrelated to any VA partner (even though all of them were unsuccessful) shows that the call was known.

One single list for disseminating the information, however, may seem scant even though other venues such as facebook, twitter, the RIO publication, and a presentation at BiodiversityNext were used. For Call 2, it seems reasonable to try to use additional mailing lists. BiodiversityNext compiled a large list through merging several mailing lists, potentially reaching about 4,000 people, and could perhaps become helpful if no privacy issues arise. In addition, perhaps greater focus on social media could attract a different community. These efforts may be particularly effective if coupled with 'success stories' from Call 1.

*6. Does the balance of requests show evidence of reaching new communities of users?*

Although the vast majority of proposals were "classical" digitization projects (typically 2D of plants, insects, and invertebrates), there was some representation of other communities that are more often served by other venues, such as molecular or 3D data. At least some of those communities were reached. The fact that there were not many bids might as well represent, rather than lack of awareness, a lack of realization of the opportunities that this infrastructure may represent. This may change. The overall emphasis remains taxonomy, and it would be valuable to reach the ecological community, perhaps through direct advertising to various professional ecological societies, organizations, and research consortia and their listservs.

*7. Is the balance of requesters wide enough geographically? Suggestions for ways in which we might improve/expand this?*

It is concentrated around the "usual suspects". As a European project, it seems logical that it is Eurocentric. To note, US and BR institutions were amply represented, as were a number of researchers from ZA (although concentrated in a couple of projects). Aside from these, no significant representation came from Asia or Africa apart from the odd CN, JP or IN representatives, who might in turn be actually and scientifically associated with VA partners or PIs.

As mentioned in #3, a wider dissemination of the call might very well result in an increased number of bids (although this may also bring in random bids).

*8. Any comments on the VA Coordinator role? Single vs. shared role, scope of work?*

The work seems well and transparently done, but the RB wonders how the VA Coordinator role scales with additional requests.

*9. Any other feedback you might want to give us to help improve VA Call 2 in early 2021.*

Please see Recommendations section below.

### *C. Recommendations*

Recommendations, given what was learned during the first call:

- Reevaluate the cost book template structure to see if it could be improved.
- Incorporate success stories from the first VA call into the advertisements of the following calls:
    - Examples on how to develop proposals;
    - Examples of what is possible/feasible in a VA call;
    - Broaden communication plan with knowledge for who did and who did not submit proposals;
    - Identify gaps in first call proposals to target in the second call.
- Revisit the application process instructions taking into account the advancements in the ELViS platform.
- Consider normalizing reviewer scores to account for scoring tendency of reviewers to vary.
- Consider funding for investment over the longer term and how VA fits into the future, for example, as the foundation for DiSSCo.

**Appendix 1.**

Many thanks to AHAP for this analysis.

***Post-fact analysis of the call results***

The call elicited 39 bids of which 30 were approved for further review. An additional 9 were duplicates or tests. The 30 approved bids involved 155 scientists from 38 countries, 12% listing German and 10% British institutions (Fig. 1). The median number of proposers per proposal was 3, from 2 countries (max. 30 / 17). Collectively, these 30 proposals estimated digitizing 264K museum units (ranging from 4 to over 90 000) across a range of at least 9 groups, at a cost of over 1 M€. Entomology collections were the most represented group (>= 7 proposals).

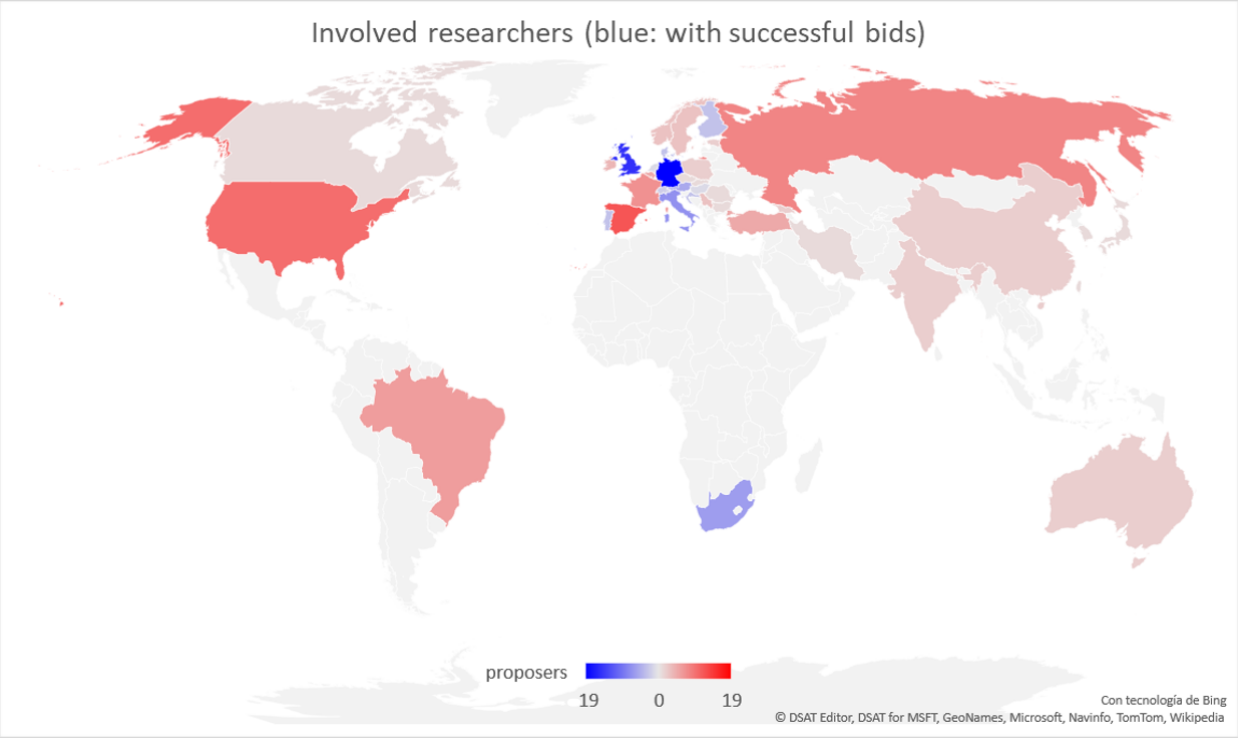


Fig. 1. Provenances of the persons included in the proposals. Shaded according to total number. Successful bids endorsed by researchers hailing from a country turn that country blue.

The five successful proposals were endorsed by 22 scientists from 12 countries: ZA, GB, HR, DE, AT, CH, IT, PT, DK, FI, HU and NL, and involved very specific taxonomic groups. 214K specimens (81%) will be digitized with the allocated funds (39% of the total request). The cost per digitized specimen becomes 3.3€, while the costs per digitized specimen of the remaining bids would have been 58.7€ (but note that the type of digitization and therefore the associated costs vary greatly among projects), a ratio of 1 : 18. The average cost per funded project triples that of the ones not funded, though—accounted for mainly

by the number of expected digitizations. Other differences among the two groups did not seem relevant (table I, fig 2).

| funded | Funded | Not funded Ratio |
|---|---|---|
| Average amount of specimens per project | 42787 | 2043<br>21 : 1 |
| Average cost per project (€) | 86864 | 28883<br>3 : 1 |
| Average digitizing cost per specimen (€) | 3.27 | 58.67<br>1 : 18 |
| Mean (Median) number of proposers | 4.8 (2) | 5.4 (3) |
| Mean (Median) number of requesting countries | 2.6 (2) | 3.2 (2) |
| Average no. of papers per PI | 48.2 | 62.4 |

Table I. Some summary statistics of the proposals.

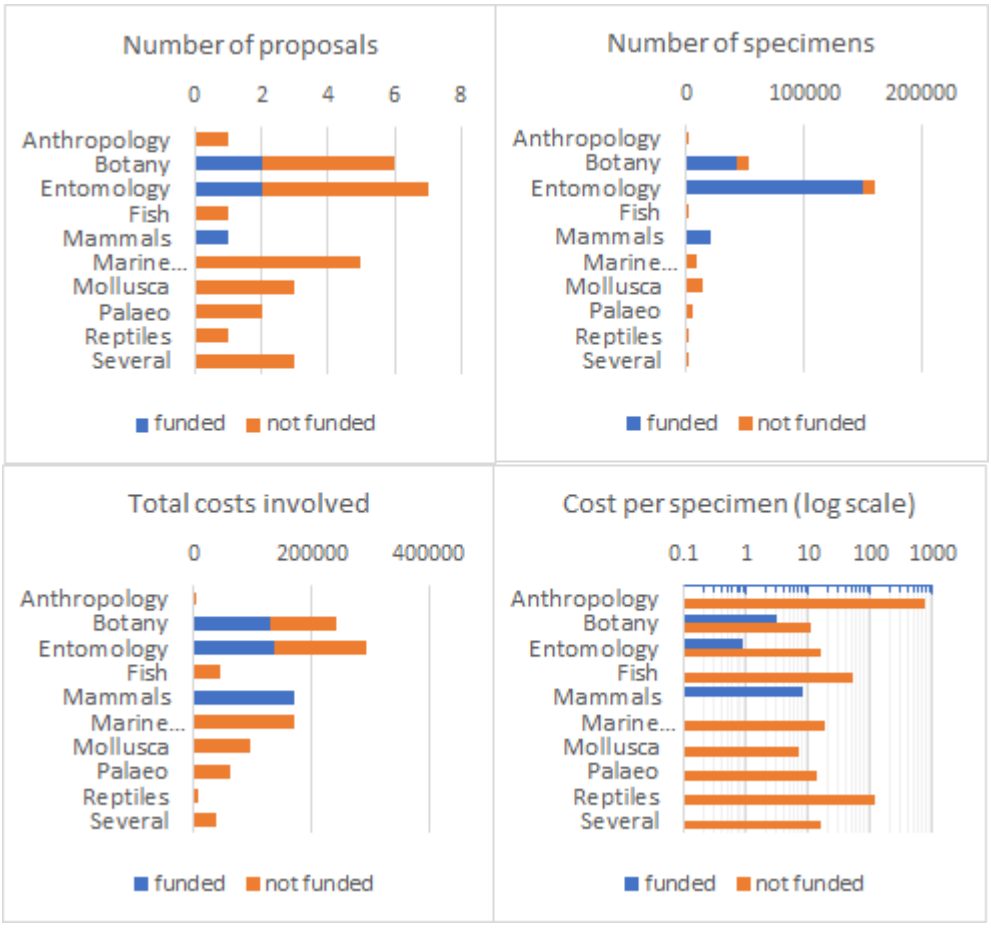


Fig. 2. Distribution of proposals and costs across taxonomic groups.

Of the 19 partners in VA, all but one (RBINS) were targeted in at least one bid. MNHN and MfN were targeted in the most proposals (13 each), while NHM and NHMW were included in 11, with the latter being present in 3 of the 5 funded bids. CSIC, HUJI and NRM were not in any funded bids. The rate of commitment for funded projects is thus 79% (15/19).

| queta | NHMW | NHM | MfN | MNHN | SGN | HNHM | CSIC | NMP | BGM | RMCA | BGBM | HUJI | LUOMU | UCPH | RBGE | RBGK | SMNS | NRM | Total ge |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BE | 1 | 8 | | 4 | | 4 | 5 | 4 | 5 | | 5 | 4 | 4 | 4 | 4 | 4 | 4 | | 60 |
| DE | 12 | 5 | 10 | 5 | 3 | 2 | 1 | 2 | 2 | | 2 | 2 | | | 2 | 2 | | | 50 |
| ZA | | 8 | 4 | 2 | | 2 | 2 | 2 | 2 | 4 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | | 40 |
| GB | 3 | 7 | 5 | 7 | | | 1 | | | 3 | | | | 2 | | | | | 28 |
| US | 8 | 5 | 3 | 3 | | | 3 | | | | | | | | | | | | 22 |
| HR | 3 | | | | | 3 | | 3 | 3 | | 3 | | | | 3 | 3 | | | 21 |
| PT | 1 | 3 | 2 | 3 | 2 | 3 | 1 | 2 | | 1 | | | 1 | | | | 1 | 1 | 21 |
| SE | | 3 | 3 | 3 | 3 | 3 | | | | | | 3 | 3 | | | | | | 21 |
| RU | 4 | 4 | 1 | 4 | 4 | 1 | 1 | | | | | | | | | | | | 19 |
| IT | 3 | 2 | 2 | 2 | 1 | 2 | | 1 | 1 | 2 | 1 | | | | 1 | 1 | | | 19 |
| ES | 4 | 9 | 2 | 1 | 1 | | 1 | | | | | | | 1 | | | | | 19 |
| FR | 1 | 4 | 1 | 4 | | | 4 | | | | | | | | | | | | 14 |
| NO | 2 | 2 | 2 | 2 | 1 | 1 | 1 | | | | | 1 | 1 | | | | | | 13 |
| FI | 1 | 2 | 2 | 2 | | 1 | 1 | 1 | | | | | 2 | | | | | 1 | 13 |
| AT | 1 | 2 | 3 | 1 | 2 | 1 | | | | 2 | | | | | | | | | 12 |
| CH | 1 | 1 | 1 | | | 1 | | 1 | 1 | 1 | 1 | | | | 1 | 1 | | | 10 |
| HU | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | | 1 | | | | | | | 1 | | 10 |
| TR | 5 | | | | 4 | | | | | | | | | | | | | | 9 |
| BR | 6 | 1 | 1 | 1 | | | | | | | | | | | | | | | 9 |
| IL | | 1 | 4 | 1 | | | | | | 1 | | 1 | | | | | | | 8 |
| SK | | 1 | 1 | 1 | 1 | 1 | | | | | | 1 | 1 | | | | | | 7 |
| EE | 2 | 1 | 1 | 1 | | | 1 | | | | | | | | | | | | 6 |
| PL | 1 | 1 | 1 | 2 | | | 1 | | | | | | | | | | | | 6 |
| DK | 2 | | | | | | | | 1 | | | | | 2 | | | | | 5 |
| GE | 2 | | | | 2 | | | | | | | | | | | | | | 4 |
| RS | 1 | | 2 | | 1 | | | | | | | | | | | | | | 4 |
| NL | 1 | | | | | | | | 1 | | | | | 1 | | | | | 3 |
| JP | 1 | | 1 | | | | | | | | | | | 1 | | | | | 3 |
| IR | 1 | | | | 1 | | | | | | | | | | | | | | 2 |
| AU | 2 | | | | | | | | | | | | | | | | | | 2 |
| CN | 2 | | | | | | | | | | | | | | | | | | 2 |
| IN | 2 | | | | | | | | | | | | | | | | | | 2 |
| BG | 1 | | | | 1 | | | | | | | | | | | | | | 2 |
| CZ | 1 | | | | | | | | | | | | | | | | | | 1 |
| TW | | | | | | 1 | | | | | | | | | | | | | 1 |
| RO | 1 | | | | | | | | | | | | | | | | | | 1 |
| KR | 1 | | | | | | | | | | | | | | | | | | 1 |
| CA | 1 | | | | | | | | | | | | | | | | | | 1 |
| general | 79 | 71 | 53 | 50 | 28 | 28 | 23 | 17 | 16 | 15 | 14 | 14 | 14 | 13 | 13 | 13 | 8 | 2 | 471 |

Table II. Provenances of proposers (rows) vs targeted TAFs (columns).

All five funded projects had 100% inbreeding, including proposers belonging to the target institutions. Among the unfunded proposals, only half of the bids had inbreeding while the other half did not include proposers from the target institutions.

Belgium and Germany provided the largest contingent of proposers, while Wien was the most sought-after collection holder (plot: instances of researchers of a country involved in projects targeting specific partners) (Table II).

From the “repatriation” point of view, Belgium had the widest interests while Germany and Great Britain were perceived as the largest and most diverse holders (plot: countries of proposers vs countries of facilities) (Table III).

| | DE | GB | AT | FR | BE | HU | ES | CZ | IL | FI | DK | SE | Tot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BE | 9 | 16 | 1 | 4 | 5 | 4 | 5 | 4 | 4 | 4 | 4 | | 60 |
| DE | 15 | 9 | 12 | 5 | 2 | 2 | 1 | 2 | 2 | | | | 50 |
| ZA | 8 | 12 | | 2 | 6 | 2 | 2 | 2 | 2 | 2 | 2 | | 40 |
| GB | 5 | 7 | 3 | 7 | 3 | | 1 | | | | 2 | | 28 |
| US | 3 | 5 | 8 | 3 | | | 3 | | | | | | 22 |
| HR | 3 | 6 | 3 | | 3 | 3 | | 3 | | | | | 21 |
| PT | 5 | 3 | 1 | 3 | 1 | 3 | 1 | 2 | | 1 | | 1 | 21 |
| SE | 6 | 3 | | 3 | | 3 | | | 3 | 3 | | | 21 |
| RU | 5 | 4 | 4 | 4 | | 1 | 1 | | | | | | 19 |
| IT | 4 | 4 | 3 | 2 | 3 | 2 | | 1 | | | | | 19 |
| ES | 3 | 9 | 4 | 1 | | | 1 | | | | 1 | | 19 |
| FR | 1 | 4 | 1 | 4 | | | 4 | | | | | | 14 |
| NO | 3 | 2 | 2 | 2 | | 1 | 1 | | 1 | 1 | | | 13 |
| FI | 2 | 2 | 1 | 2 | | 1 | 1 | 1 | | 2 | | 1 | 13 |
| AT | 5 | 2 | 1 | 1 | 2 | 1 | | | | | | | 12 |
| CH | 2 | 3 | 1 | | 2 | 1 | | 1 | | | | | 10 |
| HU | 3 | 1 | 1 | 1 | 1 | 2 | | 1 | | | | | 10 |
| TR | 4 | | 5 | | | | | | | | | | 9 |
| BR | 1 | 1 | 6 | 1 | | | | | | | | | 9 |
| IL | 4 | 1 | | 1 | 1 | | | | 1 | | | | 8 |
| SK | 2 | 1 | | 1 | | 1 | | | 1 | 1 | | | 7 |
| EE | 1 | 1 | 2 | 1 | | | 1 | | | | | | 6 |
| PL | 1 | 1 | 1 | 2 | | | 1 | | | | | | 6 |
| DK | | | 2 | | 1 | | | | | | 2 | | 5 |
| GE | 2 | | 2 | | | | | | | | | | 4 |
| RS | 3 | | 1 | | | | | | | | | | 4 |
| NL | | | 1 | | 1 | | | | | | 1 | | 3 |
| JP | 1 | | 1 | | | | | | | | 1 | | 3 |
| IR | 1 | | 1 | | | | | | | | | | 2 |
| AU | | | 2 | | | | | | | | | | 2 |
| CN | | | 2 | | | | | | | | | | 2 |
| IN | | | 2 | | | | | | | | | | 2 |
| BG | 1 | | 1 | | | | | | | | | | 2 |
| CZ | | | 1 | | | | | | | | | | 1 |
| TW | | | | | | 1 | | | | | | | 1 |
| RO | | | 1 | | | | | | | | | | 1 |
| KR | | | 1 | | | | | | | | | | 1 |
| CA | | | 1 | | | | | | | | | | 1 |
| neral | 103 | 97 | 79 | 50 | 31 | 28 | 23 | 17 | 14 | 14 | 13 | 2 | 471 |

Table III. Repatriation table: country of provenance of the proposers (rows) vs country of location of TAF (columns).

As designed, the prioritization workflow was of the hybrid independent-scoring + judgmental type. Six panelists were to evaluate each approved project against a rubric of seven concepts with ponderations from 10% to 30%, and a final score was calculated. The five projects selected for funding were finally selected through discussion in view of the scores and did not follow the raw scoring. The exercise proved generally sound, and all rubrics were taken into account, as demonstrated by the PCA of the results (fig. 3), with almost 90% of the variance accounted for in just two components. The group of selected projects was clearly detached from the rest (red crosses). Of all seven drivers, only Data Availability seemed to gather independent scorings, but its contribution to the overall decision was low. Loadings for the other drivers were rather similar.

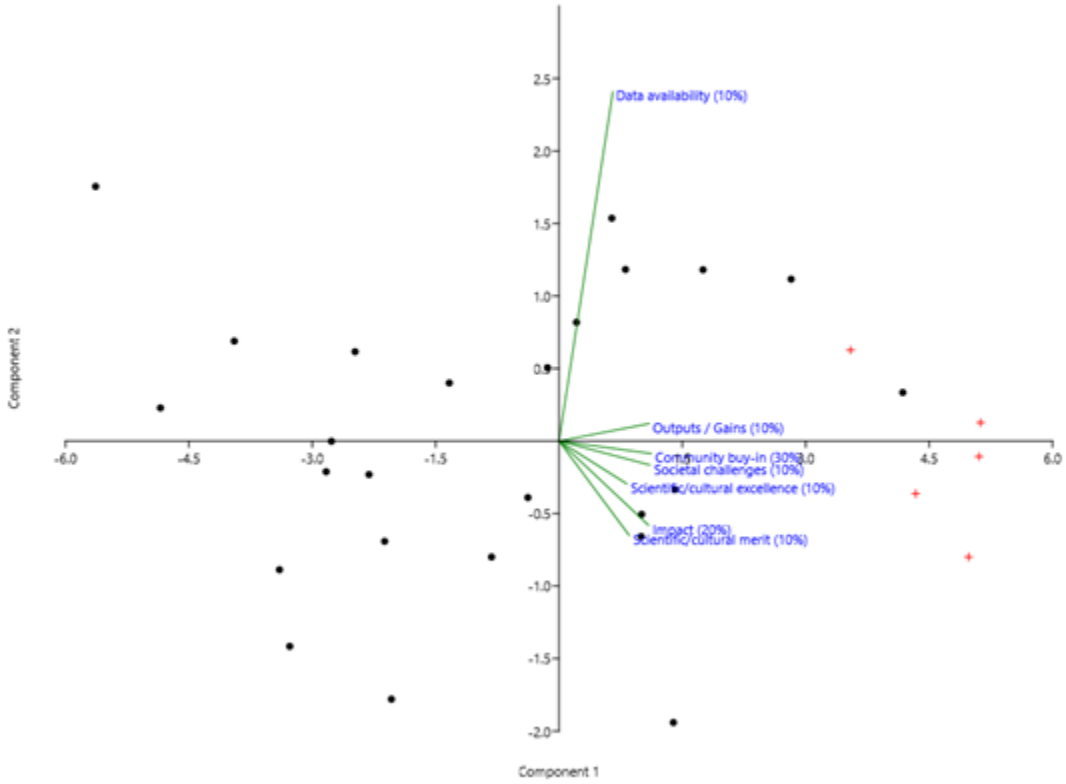


Fig. 3. PCA of scores on proposals. Red crosses are successful bids.

Interestingly, though, the outcome could have been partially predicted for most successful bids from independent metrics, with one exception. As suggested by the specimens and costs ratios described above, these became good predictors of the efficiency sought. A PCA of not-scored parameters shows that the total cost, the estimated amount of specimens, and the number of institutions targeted serves to detach a group of three successful projects, with another one in their vicinity, with only one single project (digitization of peat moss) falling squarely among the not funded ones (Fig. 4).

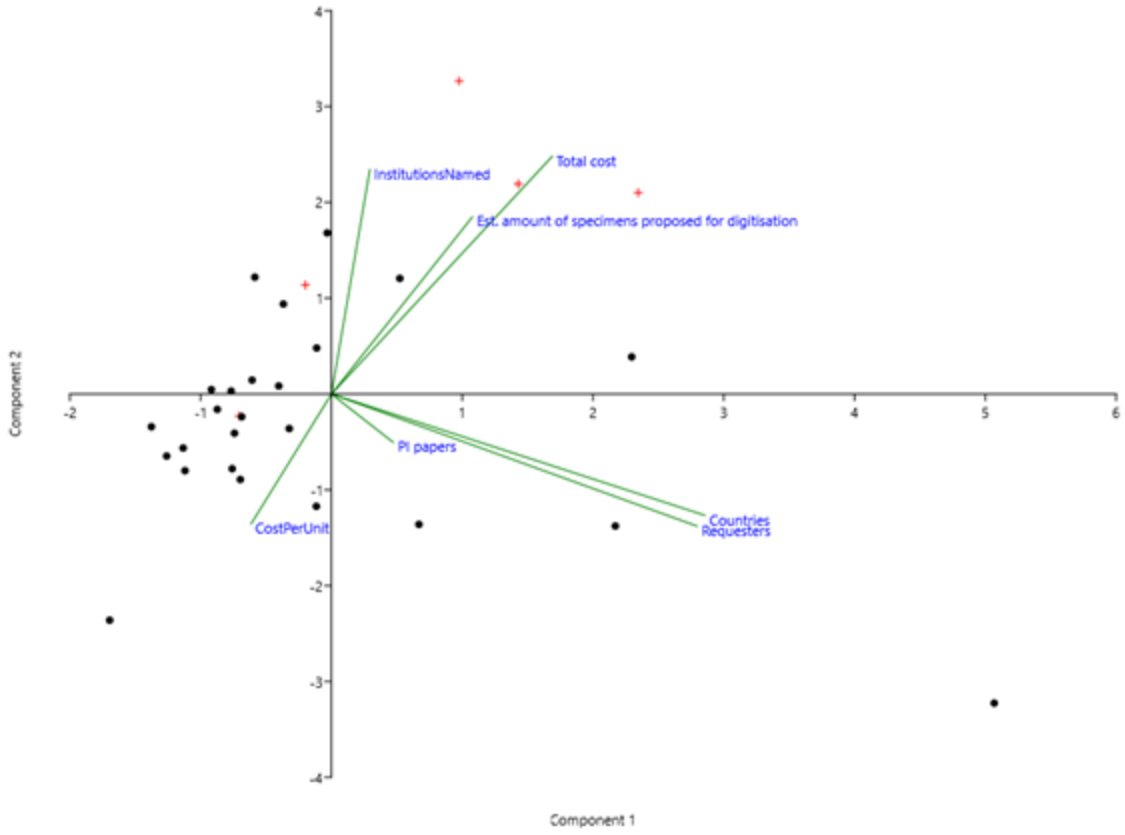


Fig. 4. PCA of non-scored variables on proposals. Red crosses are successful bids.

Thus, all other things being equal, projects with a large number of objects to be digitized at low cost per unit (yet large enough to require lots of resources) seemed to be perceived as most likely fundable. This seems in line with the intent to maximize the project's scientific returns at least quantitatively.

This point should, however, be regarded with some care, as some of these variables (countries and requesters, specimens and costs) may exhibit some degree of correlation. Also, while the first component scores more than the second, the difference is small, and the second component stays above chance expectation (defined by a broken stick model).

Notably, the "a priori" curricular strength of the PI as proxied by the number of peer-reviewed publications was inconsequential.