Base data information document (v4)

The personnel, financial and bibliometric base data is provided centrally to ensure consistent data across all programs and departments. This document provides information on how and when the data was collected and how it will be used.

How will this data be used?

At the Program-level

- To support the programs' reflections on how they are working and how they can improve.
- To enable everyone in the program to understand how their own program is financed (internal research and external research), organized (in terms of personnel and their funding), and the resulting bibliometrics.
- To enable everyone in the program to see their program in the context of their department, section, and faculty, and reflect on the diversity of ways programs operate and how their program is similar and different.

At the Department-level

- To support the departments' reflections on how they are working and they can improve.
- To enable the department to see the diversity of how their programs work.
- To enable the department to see their own department and programs in the context of the section and faculty, and reflect on how they are similar and different.

At the Faculty-level

- To support the faculty's reflections on how we are working and how we can improve.
- To enable the faculty to see the diversity across all programs and departments.
- To enable the faculty to see supporting evidence for the future proposals programs and departments make through ÖB.
- To enable the faculty to identify excellent examples and places where assistance may be needed.
- To enable the KoF panel members to see the programs and sections they are evaluating in context.

What are we not trying to do with this data?

- This data is not intended to be used to rank programs or departments for directly assigning resources.
- This data is not intended to push programs and departments towards a one-size-fits all way of working.
- This data is not intended to identify a "best" way of working.

Personnel data

Purpose: To understand the program's personnel distribution by career stage and gender and to understand how funding is used across different employment categories and genders.

Source: GLIS/Primula

Time period: Whole 2022, no particular month chosen, everything is given in terms of FTE:s.

How data was collected: The data was extracted from GLIS by looking for all employees within the faculty during the year and extracting which accounts their salary came from and what type of accounts they were (VHS). An initial program mapping was done using the account groups, and this was updated with input from the departments.

Employment categories (e.g., Prof, Assistant, PhD, etc.) and account type categories (e.g., Internal, External, Teaching) were then grouped into the KoF groups using the mapping shown in the table below. Total employment (FTE) was then added up for each employee category for each program and department, and the percent employment by account type category was calculated.

Scope: All employees within the faculty during the year. The personnel funding data will include the average % funding for each employee category for each financial category for 2022. Internal includes FFFs, SFOs and other 210. External includes both 220 and 230.

Sources of error: Manual salary changes made in Raindance will not be reflected here. For 2022, this corresponds to an average of 8% of salary costs. (Max 12% for ICM and 11% for IFA). Stipend postdocs are not included as their salaries are not paid through Uppsala. They should be manually added by the programs to the total FTEs, but they are not included in the statistics on how employees are funded. Staff who are partially funded outside the program (e.g., faculty assignments) may show up as lower percentages in the research program as a result.

Definitions:

- Internationalizations
    - Associate Professor = Universitetslektor (UL)
    - Assistant Professor = Biträdande universitetslektor (BUL)
    - Faculty = Professor, Associate Professor, Assistant Professor
    - Tenured = Professor and Associate Professor
    - Tenure-Track = Assistant Professor
- Other Research includes senior and guest professors, adjuncts, research engineers, etc.
- Other includes communications, direct research support administrators (not general or shared HR/financial support), project leaders, etc.
- MSc and BSc thesis students are not included.

Mapping of employment categories to KoF employment categories:

Befattning	**KoF Category**
Universitetslektor, biträdande	Assistant
Universitetslektor	Associate
Universitetslektor/Överläkare	Associate
Seniorprofessor	Other Research
Universitetsadjunkt	Other Research
Universitetsadjunkt, adjungerad	Other Research
Universitetslektor, bef adjunkt	Other Research
Gästprofessor	Other Research
Forskningsassistent	Other Research
Bioinformatiker	Other Research
Universitetslektor, adjungerad	Other Research
Forskningsingenjör	Other Research
Forskningsrådgivare	Other Research
Tillämpningsexpert	Other Research
Forskningsingenjör, 1:e	Other Research
Instrumentmakare	Other Research
Instrumentmakare, 1:e	Other Research
Verkstadsingenjör	Other Research
Forskare, MSCA	Other Research
Tekniker	Other Research
Forskarassistent	Other Research
Professor, adjungerad	Other Research
Biomedicinsk analytiker	Other Research
Försökstekniker, 1:e	Other Research
Ingenjör	Other Research
Doktorand	PhD
Assistent med doktoranduppgifter	PhD
Doktorand, MSCA	PhD
Postdoktor	Postdoc
Postdoktor, MSCA	Postdoc
Professor bef universitetslektor	Professor
Professor, anst UU	Professor
Professor, utn av reg	Professor
Professor, kallad	Professor
Forskare	Researcher
Other	Other

PhD student data

Purpose: To understand the departments' distribution of PhD students.

Source: GLIS/Uppdok

Time period: 2018-2022

Scope: Number of admitted, graduated, and active PhD students per department per year, along with percent female and net study time at graduation.

Sources of error: Data provided only for departments, not programs.

Financial data

Purpose: Understand the financing (income from internal and external research funds) and spending (use of internal funds by category) for programs and departments. For departments, teaching is also included.

Source: GLIS/Raindance and VP22

Time period: 2022 (2023 will come in March)

How the data was collected: Initial financial data is provided for 2022, with a later update for 2023. The data is extracted based the standard practice of each program having its own project group in the financial system. Departments that had not followed this standard were asked to provide updated personnel-and-account to program mappings, which were used for the final computation. Please note that there are accuracy and practicality tradeoffs in financial data and that our goal is to achieve about 90% accuracy. If there are particular inaccuracies that meaningfully affect the evaluation, these should be described in the form.

For the Top-10 finance sources, the data was collected from VSH 220 (bidragsforskning) and VSH 230 (uppdragsforskning) matching the project group with the research program. Top-10 external funders shows the amount spent on each financier during the year.

Scope: The data includes the amounts taken in million SEK. Total internal research funding is all government base research funding (VHS 210, 220 and 230), including funds used for co-funding (VHS 220). Total external research funding is all external grant funding (VHS 210, 220 and 230).

Internal research expenses are all costs for VHS 210 and external research costs are all costs for VHS 220 and 230. Note that spending data is for how internal research funding was spent, as that is directly related to how we fund internally. Department data also includes the total amount spent on education as a separate category.

Other internal research: FFF+SFO is the amount of FFF and SFO resources allocated. Other Internal Research is the difference between total internal funding used by the program/department and the amount allocated.

The FFF+SFO data is allocated from VP22 and total internal funding, together with co-funding, is allocated from GLIS/Raindance. This means that there is no raw data for Other Internal Research. Other Internal Research can be, for example," institutionsresurser"," tidsbegränsade resurser" and" särskilda satsningar" (see appendix 2.4, 2.7 and 2.15 in VP22).

Sources of error: employment figures may be incorrect if staff are paid from project groups in other departments or programs. Similarly, financial (income and expenses) are based on project groups. This means that if staff have moved between departments/programs or if departments/programs have split or merged, and the paying projects have not been moved accordingly, those income and expenses will be accounted to the previous department/program. To address this, departments were asked to provide a mapping of personnel and accounts to programs.

Other Internal Research has limitations in its interpretation. It is the difference between total internal funding used by the program/department and the amount allocated (FFF, SFO, and co-financing).

This amount will be higher/lower if the program has transferred internal funding from/to other programs or universities (changes booked funding), and if the program has received other internal money including department resources, studiestöd, rektor support, etc. (increases booked funding beyond allocated FFF/SFO), it will also be higher.

Bibliometrics

For terminology, including the calculation of indicators, please see the Uppsala University Annual Bibliometric Monitoring 2023 report (UU ABM 2023).

Overview of Bibliometrics

We provide three different types of statistics: *quantity* (publication volumes, full counts and fractionalized by authors), *impact* (measured by two publication-level indicators based on field-normalized citation statistics, and the *Norwegian model*), and *coverage* (of UU-publications in the citation database used for impact calculations). Please note that when we talk of citation impact in KoF, we are talking of it in the context of Web of Science and not in the context of any other citation database.

For the *Coverage* (fractionalised) indicator reported in the Self-Evaluation forms, note that it is defined as *P_frac_wos/P_frac_total* (see the Bibliometrics Data tab in the Base Data and Analysis Graphs sheet) over the years 2017-2021, that is the proportion of DiVA publication fractions used in the calculation of citation statistics. This differs from the indicator WoS coverage reported in UU ABM 2023, which reports the proportion of DiVA publication fractions that is indexed by Web of Science, there over the years 2014-2022. This choice was made in order to more clearly present the proportion of publication fractions that forms the basis for the calculated citation indicators, and to avoid programs to misidentify the set of publications indexed in WoS with the set of publications used by CWTS Leiden to calculate the two impact indicators below.

The two publication-level impact indicators are mean normalized number of citations per publication (MNCS), and proportion of frequently cited publications (top 10%) (PP (top 10%)). Both indicators are used to compare citations to one publication from the research program, with an algorithmically constructed comparable set of publications. These indicators can be useful measure of impact, given that there is a sufficiently large set of publications from the research program covered by the citation database. Please have a look in Base Data and Analysis Graphs and the sheet of *Included Publications*, which lists the set of publications on which the normalized citation statistics is based.

We complement these two impact indicators by giving the proportion of publication fractions in channels of extra high prestige ("level 2 channels") according to the Norwegian model. This is useful for programs with a low level of *Coverage* (see above).

We have also added one sheet of *Most Frequent Publications* in Base Data and Analysis Graphs, from which the program should manually aggregate and extract their top list of publication channels. The data in this sheet has been programmatically harvested, with conservative merging of channel names.

Analysis of the full set of publications

Source: GLIS/DiVA

Time period: 2017-2022

Scope: Publications (journal articles, articles in anthologies, monographs and conference papers) affiliated with the disciplinary domain of Science and Technology. The analysis makes use of the Norwegian model (Section 2.1 in UU ABM 2023).

Purpose: To provide an overview over total publication volumes, to obtain proportion of publications covered by Web of Science, proportion of publications in channels of extra high prestige (level 2 channels in the Norwegian model), and host publication frequencies by research program.

Sources of error: Publications affiliated with departments, but not correctly affiliated with individual UU-authors. Specifically, UU-authors may lack their AKKA-id in publications, which results in their publications fractions not being accounted for on research program level. Conversely, there can be Akka-ids in publication entries added for non-UU-affiliated authors, leading to publication fractions accounted for on program level but not on department level. The departments and programs should be aware of these two discrepancies.

With GLIS/DiVA as basis, we report the following indicators in the Self-Evaluation document: *Number of publications* (full/fractional counts) and *Proportion of publication fractions in the Norwegian model level 2.*

Analysis of the subset of publications covered by Web of Science

Source: The in-house version of Web of Science available at the Centre for Science and Technology Studies (CWTS) at Leiden University, the Netherlands. Let WoS-CWTS be this database.

Time period: 2017-2021

Scope: A proper subset of *core publications*. Core publications are publications in international scientific journals in fields that are suitable for citation analysis (https://www.leidenranking.com/information/indicators).

Purpose: To provide an overview over UU-publications covered by WoS-CWTS, and to obtain field-normalized (i.e. subject-normalized) citation statistics for UU-publications included in WoS-CWTS. (See Section 2.1 in UU ABM 2023.)

Sources of error: Coverage low for some subjects. Citation indicators unstable for small publication volumes. See also the sources of error point for the source GLIS/DiVA.

With the addition of the source WoS-CWTS, we report the following indicators in the Self-Evaluation document: *Proportion of publication fractions used for citation analysis* (the *Coverage* indicator), *Mean normalized number of citations per publication* (MNCS), *Proportion of frequently cited publications (top 10%)* (PP(top 10%)).

Data provided by the departments

Source: Manual distribution by prefects/head of departments of Uppsala authors (more precisely: distribution of AKKA-ids) to research programs.

Time period: 2017-2023

Scope: AKKA-ids found in DiVA or via manual searches.

Purpose: To allocate publication fractions to research programs via AKKA-ids

Sources of error: Manual curation in steps by a large number of people gives a high risk for errors. Researchers no longer affiliated with UU may be missed in the distribution. Specifically, no longer active AKKA-ids may have been missed.

Short description of work process

- A list of all Uppsala affiliated authors with publications registered in the faculty was extracted from GLIS/DiVA on 8-Oct-2023.
- This list, combined with personnel data (see above), was sent to the prefects on 7-Nov-2023.
- The 12 person-to-program lists received were curated in December 2023 at the Office of Science and Technology in communication with department prefects, and combined to one final mapping in January 2024 by the Office of Science and Technology.
- The final person-to-program mappings from the departments were prepared for calculation in the beginning of January 2024. Data were sent to CWTS on 11-Jan-2024. Post-processing of results was then performed in the end of January. This was done by the university Bibliometrician at the Planning Division.
- The extraction of host publication frequencies is based on data extraction from DiVA in November-December 2023, curated/collated by the university Bibliometrician.
- Visualization of results was done at the Office of Science and Technology.

Notes on statistics and indicators

All articles, articles in anthologies, monographs, and conference papers are included in total statistics. Publication statistics are mainly reported as fractionalized counts. (E.g., a publication with 1 of 4 authors from Uppsala will count as 0.25.) For some indicators, full counts are also reported.

Total publication volumes and percent of publications in the Norwegian Model's level 2 channels will be provided for 2017-2022. See Section 2.1 of UU ABM 2023 for details.

We use WoS-CWTS to calculate field-normalized citation statistics. See Section 2.2 of UU ABM 2023 for details. WoS-CWTS does include journal articles and reviews, and is based on a proper subset of the Web of Science database. Programs with low coverage should specifically reflect on the relevance of coverage when discussing bibliometrics.

Field normalized citation counts require publications to be at least 2 years old to have stable citation statistics, that there is a sufficient number of publications to provide valid statistics, and that there is a sufficient coverage of the publications from the evaluated unit to be representative. This means that for many of the 67 research programs evaluated, publication volumes are so small that the indicator values are unstable. Hence, *these statistics must be interpreted with great care*. The two-year citation window limits our statistics to providing aggregates over 2017-2021.

Please note that all statistics for research programs are based on the department-provided lists of people. This means that people who moved between programs or departments during the evaluation period will have their publications counted only in their program(s) and department(s) as of year 2023. If this is of particular importance, it should be noted in the form.