APPENDICES
SciK-Health: An Open-Data Dashboard for the Multidimensional Evaluation of Italian Academic Health Science Centers

Massimo Aria[1], Corrado Cuccurullo[2], Luca D'Aniello[3], Maria Spano[4*]

*Corresponding author(s). E-mail(s): maria.spano@unina.it;
Contributing authors: massimo.aria@unina.it;
corrado.cuccurullo@unicampania.it; luca.daniello@unina.it;

The following appendices provide detailed technical documentation supporting the main text. Appendix A documents all 142 performance indicators implemented in the SciK-Health platform, organized by dimension and category. Appendix B presents detailed walkthroughs of the platform interfaces through concrete user scenarios, demonstrating how each stakeholder group interacts with the system. Appendix C provides technical implementation details covering the deployment architecture, graph database schema, frontend design, and the natural-language concept-mapping module.

A Performance Indicators

This appendix provides comprehensive documentation for all 142 performance indicators implemented in the SciK-Health platform. Indicators are organized across five principal dimensions, each addressing distinct facets of AHSC research productivity, quality, and impact. All indicators are computed annually by institution (AHSC) and publication year (PY) to enable temporal trend analysis and cross-institutional comparisons. The majority of indicators are normalized or standardized to facilitate meaningful comparisons across institutions with different sizes and organizational configurations.

A.1 Bibliometric Dimension (51 indicators)

The bibliometric dimension describes scientific production volume, authorship patterns, collaboration dynamics, and research quality. These indicators are primarily computed using publication metadata from OpenAlex and enriched with journal quality metrics from SCOPUS.

A.1.1 Productivity Indicators (9)

Table 1 Productivity Indicators

Code	Description
PP1	Total number of articles published in the year
PP2	Average number of authors per article
PP3	Average number of internal authors per article
PP4	Average number of articles per internal author as first author
PP5	Average number of articles per internal author as last author
PP6	Average number of articles per internal author as corresponding author
PP7	Average number of authored articles per internal author
PP8	Percentage of Open Access articles
PP9	Number of single-author articles

A.1.2 Impact Indicators (11)

Table 2 Impact Indicators

Code	Description
PI1	Number of articles published in Q1 journals (quartile 1) according to SJR[a]
PI2	Number of articles published in Q2 journals according to SJR
PI3	Number of articles published in Q3 journals according to SJR
PI4	Number of articles published in Q4 journals according to SJR
PI5	Percentage of articles published in Q1 journals
PI6	Percentage of articles published in Q2 journals
PI7	Percentage of articles published in Q3 journals
PI8	Percentage of articles published in Q4 journals
PI9	Average number of citations per article
PI10	Average Normalized Total Citations (NTC) per institution and year[b]
PI11	Average citation trend over time[c]

[a] Indicator based on data from SCOPUS. Journals assigned to multiple subject categories were selected based on the best percentile. Field-weighted citation impact (PI10) is the recommended citation-based measure for cross-institutional comparisons.
[b] Measures the average citation performance of an institution's publications in a given year,

normalized to the global average citation rate for that year, enabling cross-institution comparison.
[c] Citation trend is a metric assigned to articles by OpenAlex starting from 2012.

A.1.3 Collaboration Indicators (2)

Table 3 Collaboration Indicators

Code	Description
PC1	Percentage of articles with international co-authors
PC2	Percentage of articles with co-authors affiliated with different institutions

A.1.4 Talent Indicators (14)

These indicators focus on the concentration and distribution of scientific contributions across internal and external authors. They provide insights into the presence of high-performing individuals, institutional equity, and authorship leadership roles.

Table 4 Talent Indicators

Code	Description
PT1	Gini concentration index (fractional) on internal authors[a]
PT2	Gini concentration index on internal authors in first position[b]
PT3	Gini concentration index on internal authors in last position[b]
PT4	Gini concentration index on internal corresponding authors[b]
PT5	Most frequent authors (internal and external) - names and frequencies
PT6	Most frequent authors (internal and external) by fractional frequency
PT7	Fractional frequencies associated with each author in PT6
PT8	Most frequent internal authors - names and frequencies
PT9	Most frequent internal authors by fractional frequency
PT10	Fractional frequencies associated with each internal author in PT9
PT11	Most cited (internal and external) authors - names and total citations
PT12	Most cited internal authors - names and total citations
PT13	Most cited internal authors by normalized citations per year[c]
PT14	Normalized citations associated with most cited internal authors in PT13

[a] The weight assigned to each author is fractional (considering both publication count and co-authorship).
[b] Gini concentration computed using the number of published articles per author as weight.
[c] Top cited authors, where citation counts are normalized by dividing each article's citation count by the average citations of all articles published in the same year.

A.1.5 Topic Indicators (15)

This group of indicators summarizes the conceptual and thematic content of the scientific output produced by AHSCs. It includes hierarchical concept levels (from general disciplines to specific biomedical entities), MeSH terms, authors' keywords, and text-mined bigrams from titles and abstracts.

Table 5 Topic Indicators

Code	Description
PTC1	Most frequent level 0 concepts (general disciplines, e.g., MEDICINE, BIOLOGY) and frequencies
PTC2	Most frequent level 1 concepts (subdisciplines or specializations, e.g., CARDIOLOGY, SURGERY) and frequencies
PTC3	Most frequent level 2 concepts (more specific topics) and frequencies
PTC4	Most frequent level 3 concepts (more detailed, e.g., related to diseases or methods) and frequencies
PTC5	Most frequent level 4 concepts (highest specificity, e.g., biological entities, anatomical structures) and frequencies
PTC6	Most frequent MeSH terms associated with the articles - names and frequencies
PTC7	Most frequent primary topics - names and frequencies
PTC8	Most frequent authors' keywords - names and frequencies
PTC9	Most frequent bigrams extracted from titles - names and frequencies
PTC10	Most frequent bigrams extracted from abstracts - names and frequencies

A.2 Online Attention and Digital Engagement (19 indicators)

This section includes indicators that capture the online visibility, digital reach, and media attention of scientific publications. Metrics are collected via the Dimensions API of the *dimensionsR* R package and provide aggregated and disaggregated views of attention across platforms such as social media, blogs, news outlets, and policy documents.

Methodological note: These indicators measure *online attention and digital engagement*, not research impact. The correlation between Altmetric Attention Scores and citation counts is generally weak and highly discipline-dependent. Coverage is subject to disciplinary and linguistic biases, and Twitter/X data contain a structural break beginning in 2018 following API restructuring, making pre- and post-2018 comparisons inadvisable. The Altmetric Attention Score's proprietary weighting scheme introduces additional opacity. These indicators should be interpreted as measures of communication reach and societal visibility rather than proxies for research quality or impact.

[a] For each indicator, the 75th percentile (top 25%) was computed. This choice reflects the zero-inflated distribution of online attention metrics: a large proportion of publications attract no measurable digital engagement, and the mean is therefore a poor summary statistic. The 75th percentile captures the level of engagement characteristic of actively disseminated research outputs while remaining robust to extreme outliers. Caution is warranted when comparing values across pre- and post-2018 periods due to the Twitter/X API restructuring (see Section 6 of the main text).

Table 6 Altmetric Indicators

Code	Description
AL1	Number of articles in the year with at least one Altmetric Score (Altmetric Score > 0)
AL2	Percentage of articles with Altmetric Score > 0 out of total articles published
The following indicators represent the 75th percentile (top 25%)[a]:	
AL3	Altmetric Score value (75th percentile)
AL4	Number of unique accounts mentioning the article (75th percentile)
AL5	Number of Facebook wall mentions (75th percentile)
AL6	Number of mentions in RSS feeds and blogs (75th percentile)
AL7	Number of mentions on Google+ (75th percentile)
AL8	Number of mentions in mainstream media (newspapers, magazines, etc.) (75th percentile)
AL9	Number of mentions in patents (75th percentile)
AL10	Number of mentions on peer review sites (e.g., PubPeer) (75th percentile)
AL11	Number of citations in policy documents (75th percentile)
AL12	Total mentions in posts (including blogs, forums, etc.) (75th percentile)
AL13	Number of mentions in Reddit threads (75th percentile)
AL14	Number of mentions in Research Highlights (75th percentile)
AL15	Number of Twitter/X users who mentioned the article (75th percentile)
AL16	Number of mentions in video content (e.g., YouTube) (75th percentile)
AL17	Number of citations on Wikipedia (75th percentile)
AL18	Number of general public users interacting with the article (75th percentile)
AL19	Number of interactions by professionals (clinicians, operators, policy makers) (75th percentile)

A.3 Industrial Impact & Research Innovation: Patents (21 indicators)

This section includes indicators measured on patents collected from the Dimensions database. These variables describe yearly statistics on patent status, classification, inventorship, and jurisdiction, along with metadata from original documentation.

[a] Kind Code: a standardized alphanumeric code used to classify and categorize patents based on their type and purpose.

A.4 Clinical Research Activity (30 indicators)

This section provides indicators derived from clinical trials and dataset data. These indicators support analysis of research activity using data exported from Dimensions.

A.4.1 Clinical Trials (20 indicators)

A.4.2 Datasets (10 indicators)

A.5 Funding & Grants (22 indicators)

This section presents indicators related to competitive funding obtained by Italian AHSCs, covering financial aspects, international collaborations, outputs, and key actors involved in funded research projects.

Table 7 Patent Indicators

Code	Description
PA1	Number of distinct patents (unique Family ID) per year
PA2	Number of patents by legal status (e.g., Abandoned, Pending, Expired)
PA3	Number of patents of type 'Application'
PA4	Number of patents of type 'Grant'
PA5	Number of patents with unspecified publication type
PA6	Average number of citations received by patents per year
PA7	Average number of inventors per patent per year
PA8	Average number of inventors listed in patents per 100 internal authors
PA9	Average number of institutes involved per patent per year
PA10	Number of patents with a single inventor
PA11	Number of patents with institutional co-affiliations (institutions involved ¿ 1)
PA12	Number of distinct jurisdictions where patents were granted
PA13	Number of patents granted in specific jurisdictions (AU, US, CH, etc.)
PA14	Number of patents by Kind Code (DOCDB standards)[a]
PA15	Most frequent inventors of the patent - names and frequencies
PA16	Most frequent organizations that first owned the patent - names and frequencies
PA17	Most frequent Fields of Research (ANZSRC 2020) - names and frequencies
PA18	Most frequent RCDC Categories - names and frequencies
PA19	Most frequent HRCS RAC Categories - names and frequencies
PA20	Most frequent IPCR categories - names and frequencies
PA21	Most frequent Cooperative Patent Classification (CPC) categories - names and frequencies

A.6 Standardization and Normalization Methods

To enable meaningful cross-institutional comparisons despite substantial heterogeneity in AHSC size, organizational configuration, and mission emphasis, the majority of indicators employ standardization or normalization procedures:

- **Author-normalized indicators**: Metrics such as average articles per internal author (PP7), clinical trials per 100 internal authors (CT4), and inventors per 100 internal authors (PA8) adjust for institutional size by normalizing output volumes by the number of internal researchers.
- **Temporal normalization**: Citation-based indicators including normalized total citations (PI10) and normalized citations per author (PT13-PT14) account for publication year and discipline-specific citation patterns, enabling fair comparisons across time periods and research domains.
- **Percentile-based metrics**: Altmetric indicators (AL3-AL19) report 75th percentile values to capture significant engagement levels while mitigating the impact of zero-inflated distributions common in social media metrics.
- **Efficiency ratios**: Funding efficiency indicators (GR10) measure research productivity per unit investment, revealing return on investment independent of absolute funding levels.

Table 8 Clinical Trial Indicators

Code	Description
CT1	Number of clinical trials started in the given year
CT2	Average duration of clinical trials (in years)
CT3	Average number of participants per trial
CT4	Average number of clinical trials per 100 internal authors
CT5	Number of trials with an Altmetric score
CT6	Percentage of clinical trials with an Altmetric score among all trials started in the year
CT7	Average number of sponsors/collaborators per trial
CT8	Number of clinical trials by Phase (e.g., Phase 1, Phase 1/2 combined)
CT9	Trials with no phase declared (N/A - typically indicates a single-arm trial)
CT10	Number of post-authorisation studies
CT11	Number of clinical trials by study type (e.g., interventional, observational)
CT12	Number of clinical trials by gender of subjects included (All, Male, Female)
CT13	Number of trials registered in the ClinicalTrials.gov registry
CT14	Most frequent funder countries - names and frequencies
CT15	Most frequent fields of research (ANZSRC 2020) - names and frequencies
CT16	Most frequent RCDC categories - names and frequencies
CT17	Most frequent HRCS health categories - names and frequencies
CT18	Most frequent HRCS research activity categories - names and frequencies
CT19	Most frequent cancer types - names and frequencies
CT20	Most frequent CSO categories - names and frequencies

Table 9 Dataset Indicators

Code	Description
DS1	Number of free-available datasets published in the year
DS2	Most frequent dataset repositories - names and frequencies
DS3	Most frequent fields of research (ANZSRC 2020) - names and frequencies
DS4	Most frequent RCDC categories - names and frequencies
DS5	Most frequent HRCS health categories - names and frequencies
DS6	Most frequent HRCS research activity categories - names and frequencies
DS7	Most frequent health research areas - names and frequencies
DS8	Most frequent broad research areas - names and frequencies
DS9	Most frequent cancer types addressed in datasets - names and frequencies
DS10	Most frequently addressed Sustainable Development Goals - names and frequencies

- **Percentage-based metrics**: Indicators such as Open Access percentage (PP8), international collaboration rates (PC1, GR7), and quartile distributions (PI5-PI8) express institutional characteristics as proportions, facilitating size-independent comparisons.

These standardization approaches ensure that performance indicators reveal genuine institutional strengths and improvement trajectories independent of scale differences, supporting

Table 10 Grant and Funding Indicators

Code	Description
GR1	Number of grants
GR2	Average total funding per grant (in EUR)
GR3	Average funding per research organization involved in a grant (in EUR)
GR4	Average duration of grants (in years)
GR5	Average number of research organizations involved per grant
GR6	Average number of distinct countries involved per grant
GR7	Percentage of grants with international collaboration
GR8	Average number of resulting publications per grant
GR9	Total amount of funding (in EUR) available to the institution in a given year, based on grants active in that year
GR10	Number of publications produced per 1,000 EUR of funding available to the institution in a given year, based on grants active in that year
GR11	Most frequent research organizations - names and frequencies
GR12	Most frequent funding bodies - names and frequencies
GR13	Most frequent funder groups - names and frequencies
GR14	Most frequent funding programs - names and frequencies
GR15	Most frequent research fields based on ANZSRC 2020 - names and frequencies
GR16	Most frequent RCDC categories - names and frequencies
GR17	Most frequent HRCS Health Categories - names and frequencies
GR18	Most frequent HRCS Research Activities Categories - names and frequencies
GR19	Most frequent studied cancer types - names and frequencies
GR20	Most frequent Common Scientific Outline (CSO) categories - names and frequencies
GR21	Most frequent Units of Assessment - names and frequencies
GR22	Most frequently addressed Sustainable Development Goals - names and frequencies

fair and meaningful comparative assessment across the heterogeneous landscape of Italian Academic Health Science Centers.

B User Interface Examples and Usage Scenarios

This appendix provides detailed walkthroughs of the SciK-Health platform from the perspective of each primary stakeholder group. Through concrete usage scenarios supported by interface screenshots, we demonstrate how the platform's user-centered design enables diverse audiences to access and utilize multidimensional performance data effectively.

B.1 Citizen Interface: Finding Institutional Expertise

This scenario illustrates how a citizen uses the natural language search functionality to identify Academic Health Science Centers with expertise in a specific medical condition. The scenario follows a patient recently diagnosed with melanoma who wishes to identify Italian AHSCs with demonstrated research expertise in this disease area.

Platform Entry and Institution Browse

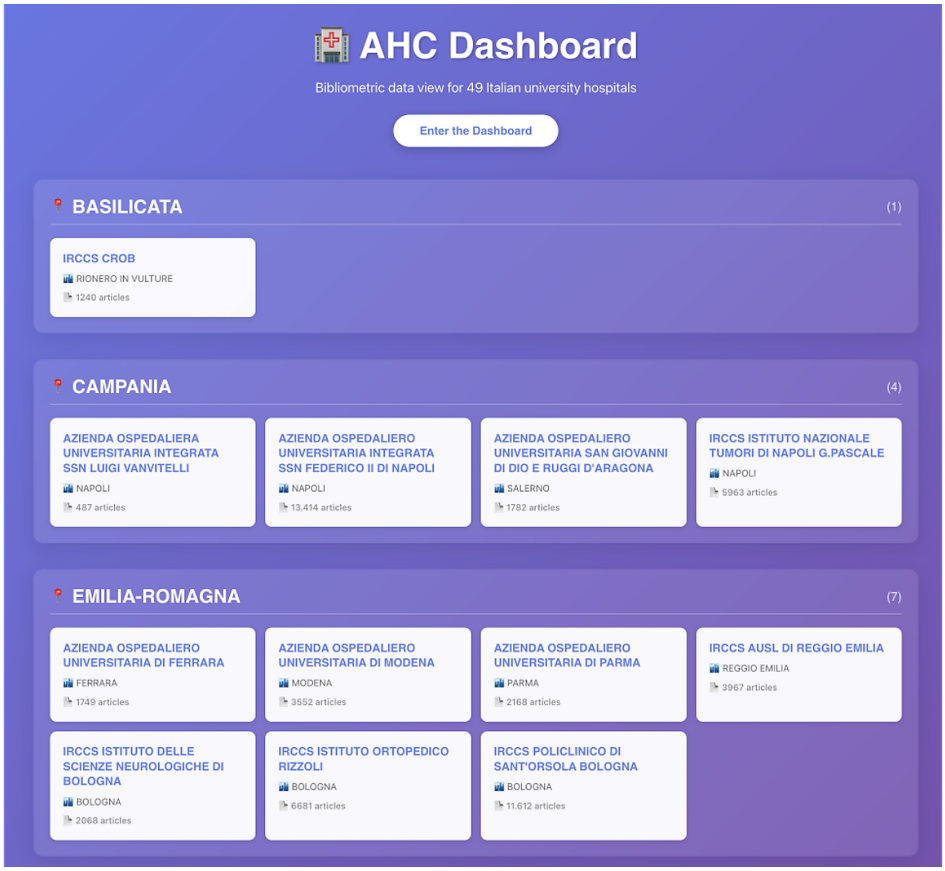

Fig. 1 Citizen Interface: Platform Landing Page. Institutions are grouped by region (BASILICATA, CAMPANIA, EMILIA-ROMAGNA shown). Each card displays the institution name and total indexed publications, enabling direct browse without a search query. The "Enter the Dashboard" button leads to the search and filtering interface.

Upon accessing SciK-Health, the citizen is presented with the main landing page (Figure 1), which organises all 49 Italian public AHSCs by geographic region. Each card displays the institution name, location, and the total number of indexed publications, providing an immediate snapshot of each institution's research volume without requiring a query. The citizen can either browse directly by region or proceed to the search interface.

Natural Language Search and Concept Mapping

To perform a targeted search, the citizen enters the query *"melanoma"* in the AI-assisted search box. The prompt-based LLM concept selector (see Appendix C, Section C.4) processes this input and automatically suggests a ranked list of semantically relevant OpenAlex concepts, in this example 11 concepts are selected from the controlled vocabulary, including *Melanoma*, *Choroidal melanoma*, *Lentigo maligna melanoma*, *Melanoma diagnosis*, *Metastatic melanoma*, and related terms. The citizen can accept the full suggested set or deselect specific concepts. Regional filters allow geographic restriction of results; in this example BASILICATA, CAMPANIA, and LAZIO are selected to limit results to institutions within accessible distance (Figure 2).

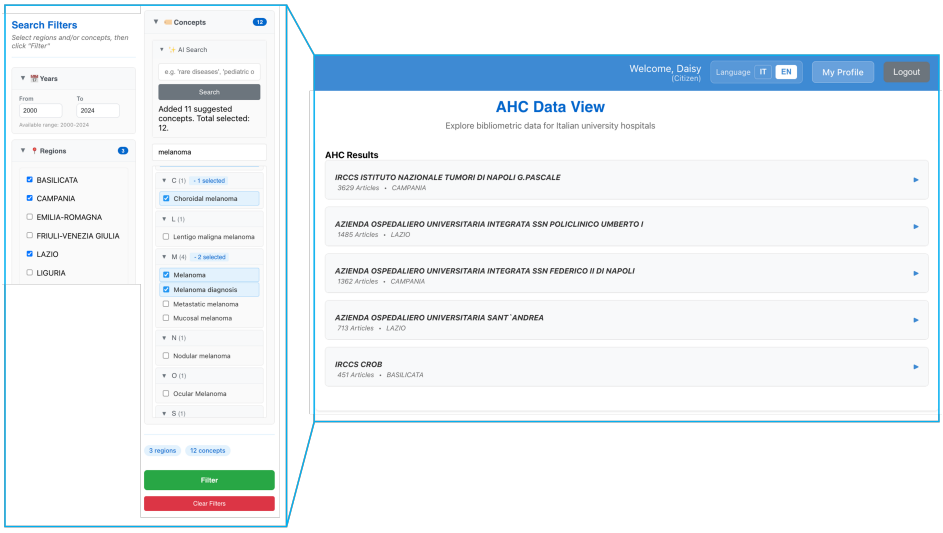

Fig. 2 Citizen Interface: AI-Assisted Concept Search and Ranked Institutional Results.

Ranked institutional results for the selected concepts and regions show IRCCS Istituto Nazionale Tumori di Napoli G.Pascale (CAMPANIA) as the top result, followed by Azienda Ospedaliero Universitaria Integrata SSN Policlinico Umberto I (LAZIO), AOU Integrata SSN Federico II di Napoli, AOU Sant'Andrea, and IRCCS CROB (BASILICATA). Results are ordered by number of publications addressing the queried concepts.

Institutional Overview Panel

Selecting IRCCS Istituto Nazionale Tumori di Napoli G.Pascale opens a five-tab institutional profile (Overview, Articles, MedStar, Researchers, Indicators). The platform labels its top-ranked researchers as **MedStars**. A MedStar is an internal author identified as among the institution's most productive and impactful contributors, ranked by publication count and

total citations on the queried topic (citizen interface) or across the full institutional portfolio (manager interface). MedStar rankings are recomputed dynamically for each query and institutional scope. Figure 3 shows the **Overview** tab, which distinguishes between overall institutional statistics and topic-filtered statistics for the current query. The upper section reports *Overall Statistics* across all research topics: 5,963 total articles, 21 grants, 582 clinical trials. The lower *Filtered Statistics* section reports metrics specific to the melanoma query (period 2000–2024): 3,629 articles on topic, 1,580 researchers active on the topic, 155,869 total citations, average 43.0 citations per article, 32 single-author articles, average article age 8.8 years, and average 16.6 authors per article. The left sidebar displays the institution's location on an interactive OpenStreetMap map with links to OpenStreetMap, Google Maps, and the OpenAlex institutional profile.

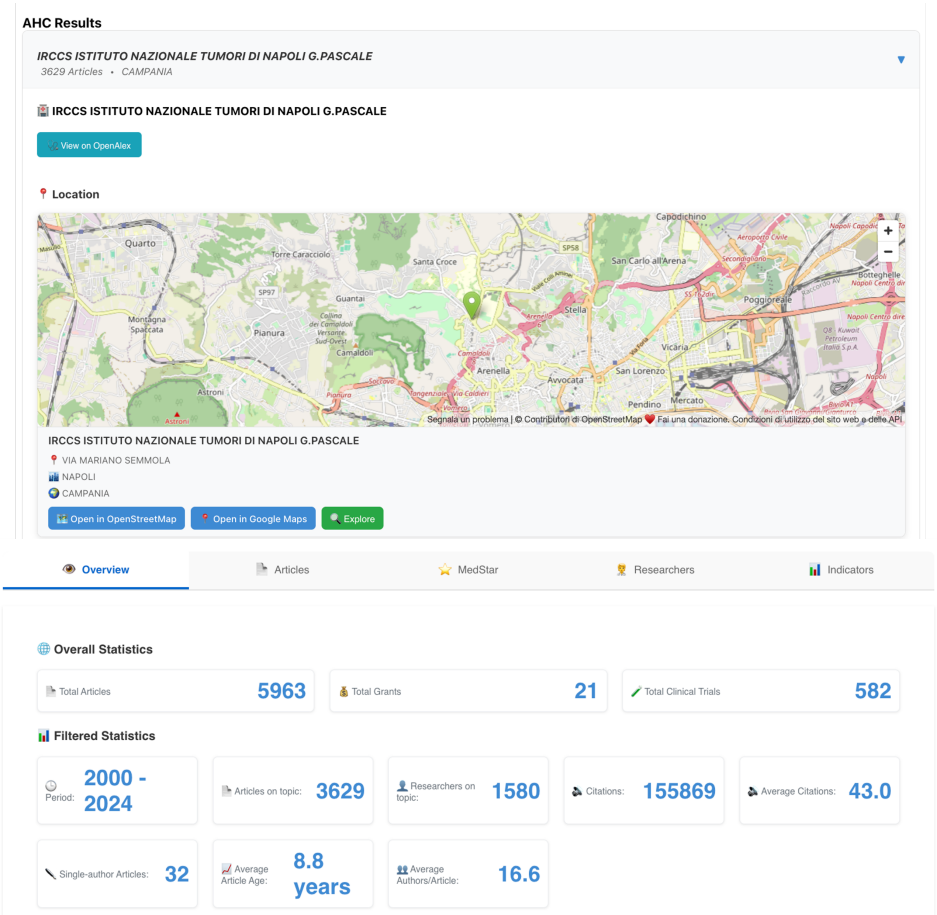

Fig. 3 Citizen Interface: Institutional Overview Panel — IRCCS Istituto Nazionale Tumori di Napoli G.Pascale.

Articles Panel

The **Articles** tab (Figure 4) displays melanoma-related publications and contextual analytics. The top section lists the top-10 most-cited articles with columns for PubMed link, title, citation count, and author list, enabling citizens to identify landmark studies and access full texts through Open Access repositories. The lower-left panel presents the annual publication production chart (2000–2023), documenting the institution's long-term and growing engagement with melanoma research. The lower-right panel displays a world map of international collaborations, with a ranked table of collaborating countries (United States leading, followed by France, Germany, United Kingdom, Spain, Switzerland, Belgium).

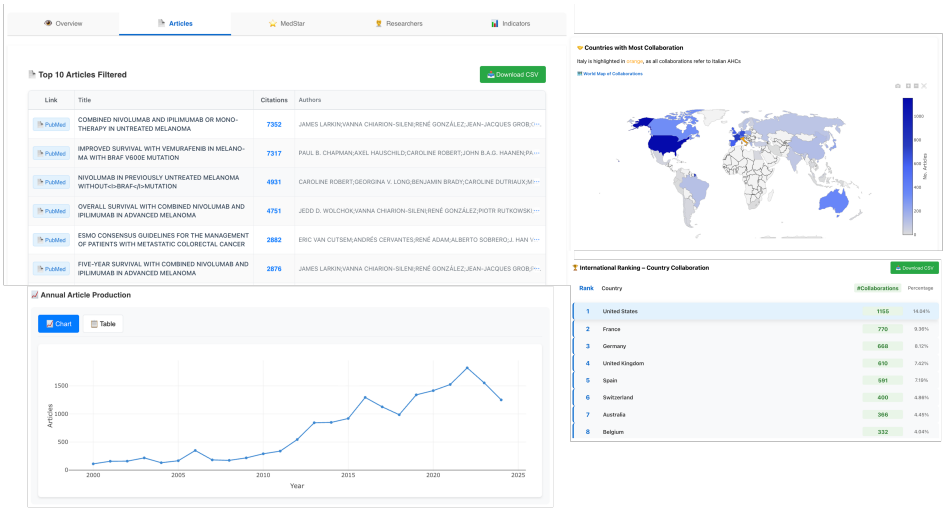

Fig. 4 Citizen Interface: Articles Panel for Melanoma Query at IRCCS Pascale.

MedStar Panel

The **MedStar** tab (Figure 5) identifies the institution's top-performing researchers on the queried topic, ranked by a composite of publication count and total citations, with OpenAlex profile links and a CSV download option. Paolo A. Ascierto ranks first with 736 topic articles and 72,692 citations. MedStar production-over-time bubble chart tracks researcher's annual publication volume and impact, and a collaboration network graph displaying co-authorship connections with other researchers.

Researchers Panel

The **Researchers** tab (Figure 6) provides a searchable directory of all internal authors affiliated with the institution, enabling citizens to identify specialists not appearing in the top-10 MedStar ranking. A free-text search for "Paolo A. Ascierto" returns the researcher's entry with 878 indexed articles and 101,030 total citations, an OpenAlex profile link, and a Details button opening the full researcher profile.

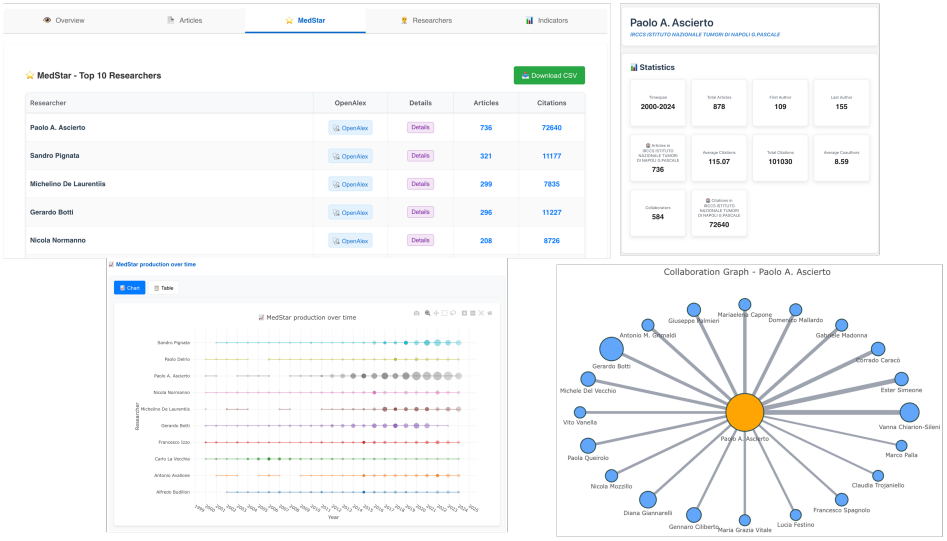

Fig. 5 Citizen Interface: MedStar Panel for Melanoma Query at IRCCS Pascale.

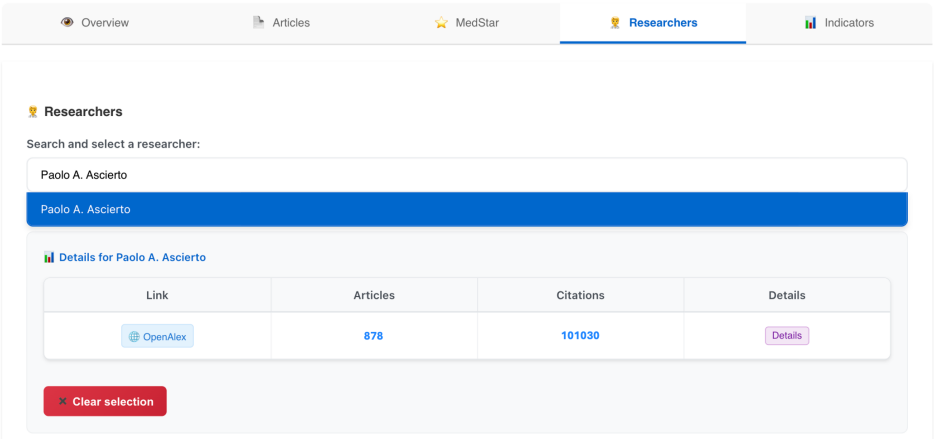

Fig. 6 Citizen Interface: Researchers Panel.

Indicators Panel and Data Export

The **Indicators** tab (Figure 7) provides access to a curated subset of performance indicators organised across eight expandable dimension panels: *Publication Productivity* (PP1–PP9), *Publication Impact* (PI1–PI11), *Publication Collaboration* (PC1–PC2), *Patents Data*, *Clinical Trials Data*, *Datasets Data*, *Online Attention & Digital Engagement* (AL1–AL19), and *Grants Data*. Citizens select individual indicators or entire categories, specify the temporal range using the year slider (default 2005–2024), and click Filter to generate a tabular view. The resulting table (25 rows × up to 211 columns for the full indicator set) can be downloaded as a CSV file for offline analysis. Indicators are displayed with plain-language labels; tooltips provide brief interpretive guidance for non-specialist users.

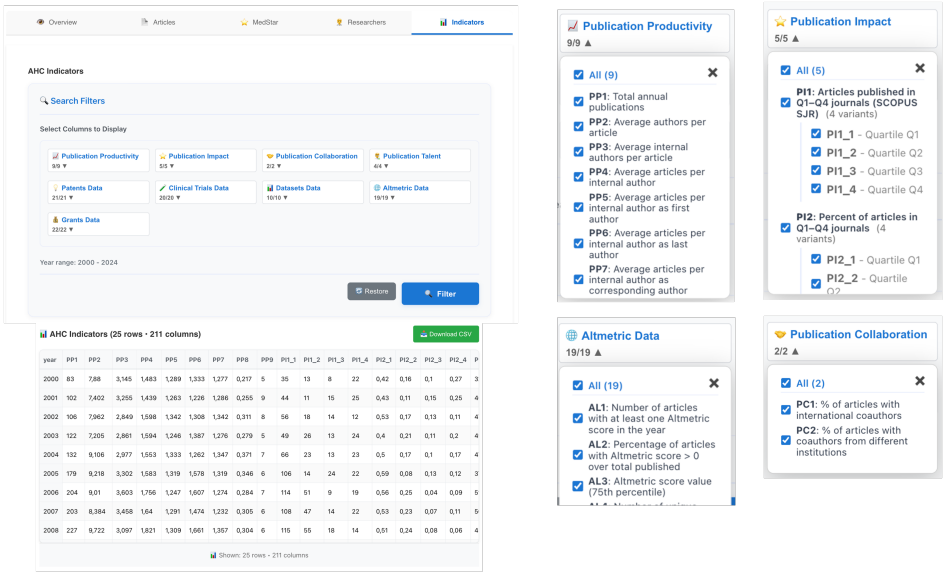

Fig. 7 Citizen Interface: Indicators Panel.

B.2 Manager Interface: Institutional Performance Monitoring

This scenario illustrates how a clinical manager or AHC director uses the authenticated manager interface to monitor comprehensive institutional performance, identify strategic priorities, and support evidence-based resource allocation decisions. The scenario follows a manager at Azienda Ospedaliera di Perugia (AOU_PERUGIA, UMBRIA region).

Authentication and Direct Institutional Access

Managers access the platform through secure authentication following a formal access request process: institutional affiliation, professional role, and intended use are documented and verified by platform administrators before credentials are issued. Unlike the citizen interface, which requires concept-based queries, the manager landing page immediately exposes the complete research profile of the affiliated institution without any query formulation step.

Figure 8 shows the authenticated landing page. The left panel provides a **Search Filters** sidebar (year range selector 2000–2024 and concept filter, activated via *Apply Filters*) that enables scoping the view to specific topics or time windows. The right panel displays the institutional header, **AZIENDA OSPEDALIERA DI PERUGIA**, AHC code AOU_PERUGIA, UMBRIA, with an OpenAlex profile link, an interactive OpenStreetMap map centred on the institution's address, and direct navigation buttons to OpenStreetMap and Google Maps.

Overview, Articles, MedStar, and Researchers Panels

The institutional profile is organised across the same five tabs as the citizen interface (Overview, Articles, MedStar, Researchers, Indicators), but with a critical functional difference: all panels display the *complete institutional portfolio* across all research domains rather than topic-filtered subsets. Figure 9 shows the four content panels side by side.

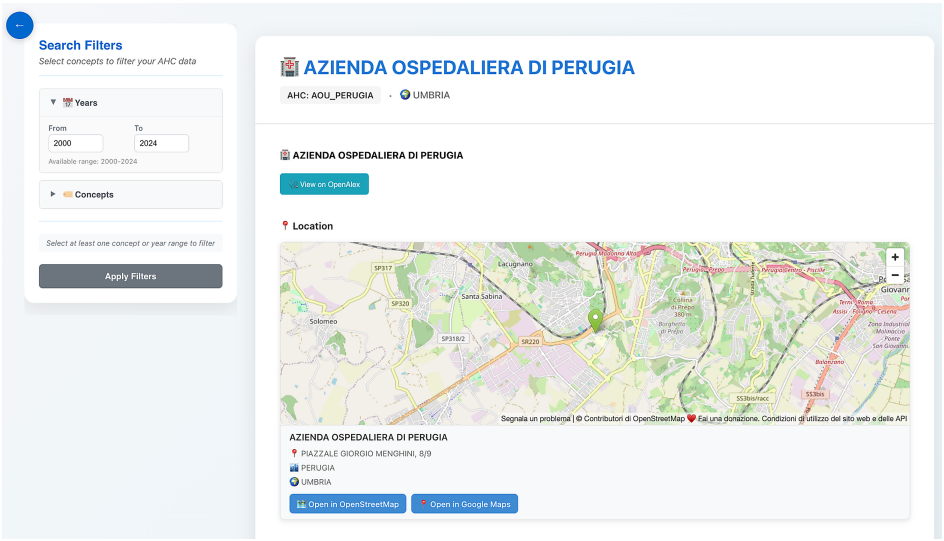

Fig. 8 Manager Interface: Authenticated Landing Page for Azienda Ospedaliera di Perugia (AOU_PERUGIA).

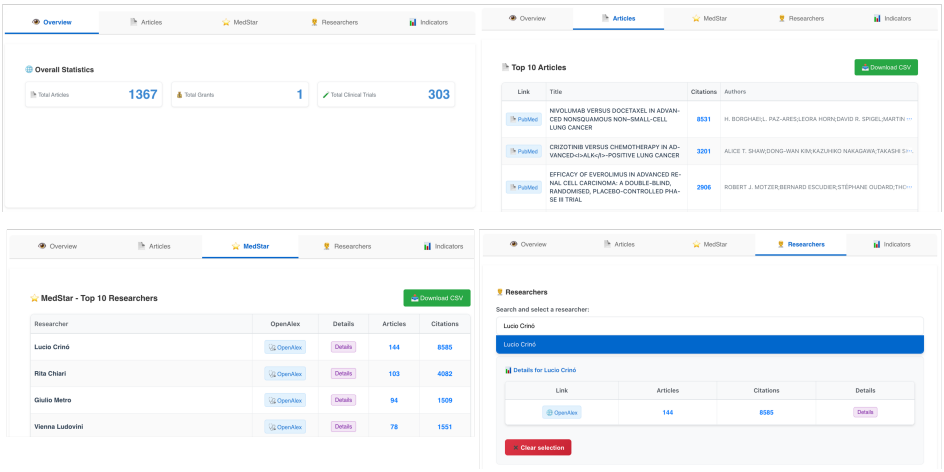

Fig. 9 Manager Interface: Overview, Articles, MedStar, and Researchers Panels for AOU Perugia.

Overall Statistics show 1,367 total articles, 1 grant, and 393 clinical trials, documenting substantial clinical research activity across all research domains (2000–2024). These summary metrics allow managers to position their institution relative to peers and detect areas requiring attention (e.g., low grant volume may signal limited competitive funding capacity).

The **Articles** tab displays the top-10 most-cited publications across the full institutional portfolio enabling managers to identify the institution's landmark research contributions regardless of disease area. The Top-10 researchers ranked by productivity and citation impact across all topics: Lucia Crimi (148 articles, 5,362 citations), Rita Chiari (123 articles, 4,685 citations), Giulia Mena (84 articles, 1,108 citations), Maura Lupecini (79 articles,

1,321 citations). The manager's MedStar ranking reflects overall institutional talent, supporting succession planning and resource allocation across all research programmes. The **Researchers** tab provides a searchable directory of all affiliated internal authors. A free-text search returns the matching researcher's total articles and citations with an OpenAlex link and a Details button for the full profile.

Manager-Specific Indicator Access

Managers access an expanded indicator set covering all eight dimensions, with fine-grained sub-indicators unavailable to citizen users (Figure 10). The Indicators tab combines a left-side selector panel with a temporal filter (year range 2000–2024) and a central data table that can be exported as CSV.

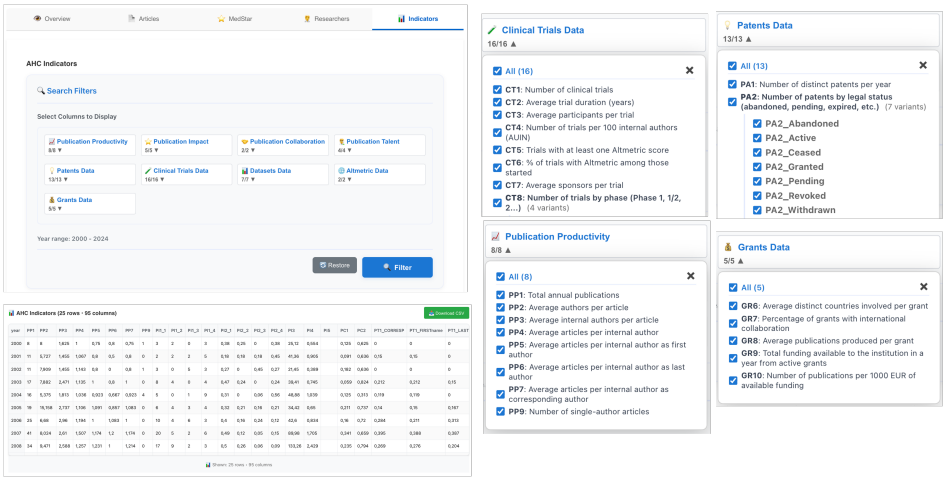

Fig. 10 Manager Interface: Indicators Panel with Expanded Dimension Categories.

Managers can select individual indicators or entire categories, apply temporal filters, and export the resulting dataset as CSV for integration with institutional reporting systems or accreditation documentation.

B.3 Funder Interface: Portfolio Analysis and Strategic Investment

This scenario illustrates how policy makers and research funders use the platform's cross-institutional access to identify research excellence, assess strategic alignment with national priorities, and support evidence-based competitive funding allocation.

Access Model and Interface Structure

The funder interface shares the same five-tab layout as the citizen and manager views (Overview, Articles, MedStar, Researchers, Indicators) but differs in two critical respects. First, funders can navigate freely across *all 49 Italian public AHSCs*: selecting any institution in the interface exposes its complete profile without restriction. Second, funders retain the LLM-assisted concept search available to citizens, enabling topic-targeted institutional discovery (e.g., querying "antibiotic resistance" or "cardiovascular biomarkers" returns a ranked

list of AHCs by publication productivity on those topics, which the funder can then explore individually). This dual access model distinguishes the funder role from both the citizen (topic-filtered only) and the manager (single institution only).

Indicators Panel: Cross-Institutional Evidence for Funding Decisions

Figure 11 shows the funder's **Indicators** tab for a selected institution. The left selector panel exposes the full eight-dimension indicator set across all nine dimension categories. The resulting AHC Indicators table (25 rows × 114 columns for the configuration shown; up to 311 columns with all indicators selected) reports annual values for 2000–2023. Because the funder can switch freely between any of the 49 AHCs and export each institution's table as a CSV, the interface supports longitudinal, institution-by-institution analysis as well as batch data extraction for portfolio-level statistical work.

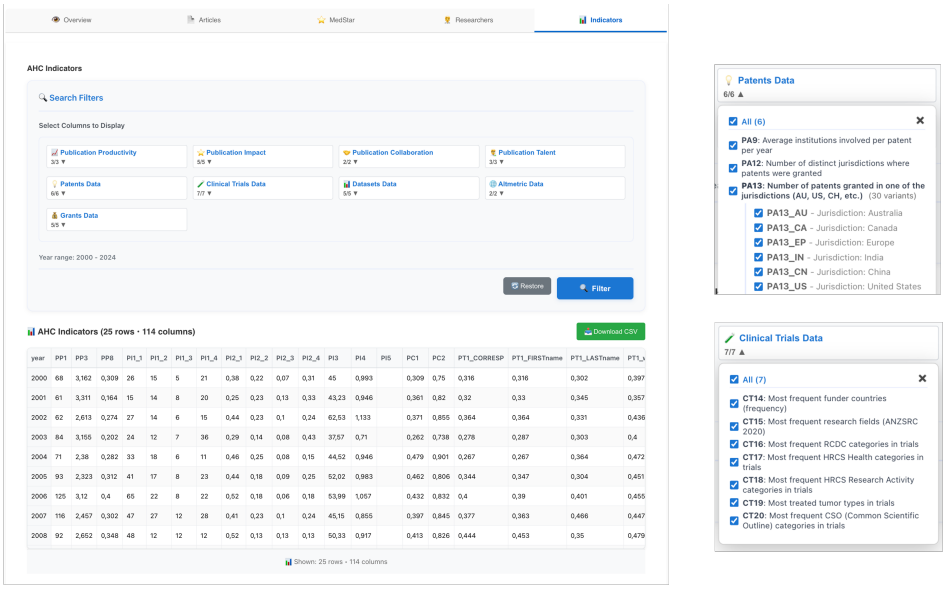

Fig. 11 Funder Interface: Indicators Panel with Funder-Relevant Expanded Categories.

The two expanded panels visible in Figure 11 illustrate indicator categories of particular strategic relevance for funding bodies:

Patents Data — jurisdictional sub-indicators (PA13). The PA13 family records the number of patents granted in specific jurisdictions (Australia, Canada, Europe, India, China, United States). These sub-indicators allow funders to assess each institution's international intellectual property strategy: an institution with patents concentrated in the US and EP jurisdictions signals a strong translational orientation towards high-value commercial markets, while absence of international filings may indicate limited technology-transfer capacity requiring targeted support.

Clinical Trials Data — thematic category indicators (CT14–CT20). This group documents the thematic profile of an institution's clinical research activity through multiple classification schemes: funder country distributions (CT14), ANZSRC research fields (CT15),

RCDC categories (CT16), HRCS health and research activity categories (CT17–CT18), cancer types (CT19), and CSO categories (CT20). Cross-referencing these indicators with a funding programme's stated priority areas enables rapid identification of institutions whose clinical research portfolio is already aligned with targeted calls, reducing evaluation overhead and improving grant allocation precision.

Strategic Use Cases

Combining indicators across all eight dimensions, funders can pursue four analytical goals:

- **Scientific excellence identification**: rank AHCSs on field-normalised citation impact (PI10), international collaboration rates (PC1), and journal quality positioning (PI5–PI8, with the caveat that these are secondary metrics per DORA/CoARA) to identify sustained research leaders
- **Innovation and translation capacity assessment**: integrate PA13 jurisdiction distributions, PA2 legal-status breakdowns, and CT8 phase distributions to profile each institution's translational pipeline from basic research through clinical application to commercial protection
- **Funding efficiency analysis**: use GR9 (total funding available), GR8 (publications per grant), and GR10 (publications per 1,000 EUR) to compare return on investment; institutions with high GR10 and high PI10 represent efficient investment targets
- **Research priority alignment**: CT14–CT20 and GR15–GR22 thematic classifications enable systematic matching of institutional research profiles to national health research strategies and European funding programme priorities

All indicator tables are exportable as CSV, supporting integration with funding agency information systems, preparation of portfolio reports for ministerial oversight, or downstream statistical analyses (clustering, ranking, regression) on the full 49-institution panel.

C Technical Implementation Details

C.1 Deployment and Security Architecture

SciK-Health runs as a containerised four-service stack managed via Docker Compose: MySQL 8.0 for user authentication and role management; Neo4j 2025.04 for the scientometric knowledge graph; a Node.js/Express backend for data retrieval and business logic; and a React/Nginx frontend. Services communicate over an isolated bridge network; the backend and graph database are not exposed to the public interface. All API endpoints require a JWT Bearer token (24-hour expiration); passwords are stored using bcrypt. Role assignment is stored in MySQL and embedded in the JWT payload at login.

C.2 Graph Database Schema and Query Design

The Neo4j knowledge graph organises data through typed nodes (`:AHC`, `:Researcher`, `:Article`, `:Grant`, `:Clinical_Trial`, `:Year`) and semantically labelled relationships. Precomputed performance indicators are stored as properties of `:HAS_STATS_IN_YEAR` edges, enabling efficient annual time-series retrieval without on-demand aggregation. Three query families exploit graph traversal capabilities: (i) concept-filtered institution ranking via a three-hop `AHC` → `Researcher` → `Article` traversal with runtime concept injection; (ii) researcher collaboration networks via `COLLABORATED_WITH` edge traversal with co-authorship counting; and (iii) aggregate statistics combining institutional, researcher, and article nodes in a single pass. Representative Cypher implementations for each query family are provided below.

Query family (i): Concept-filtered institution ranking

```
// Rank AHCs by number of publications matching a set of queried
    concepts
MATCH (a:AHC)-[:HAS_RESEARCHER]->(r:Researcher)-[:AUTHORED]->(art
    :Article)
WHERE any(c IN art.concepts WHERE c IN $queriedConcepts)
RETURN a.name AS institution, a.region AS region,
       count(DISTINCT art) AS articleCount
ORDER BY articleCount DESC
```

Query family (ii): Researcher collaboration network

```
// Retrieve co-authorship pairs for a given institution
MATCH (r1:Researcher)-[c:COLLABORATED_WITH]->(r2:Researcher)
WHERE r1.ahcId = $institutionId AND r2.ahcId = $institutionId
RETURN r1.name AS author1, r2.name AS author2,
       c.coauthoredArticles AS sharedArticles
ORDER BY sharedArticles DESC
LIMIT 50
```

Query family (iii): Annual aggregate statistics retrieval

```
// Retrieve pre-computed indicator values for an institution over
     a time range
```

```
MATCH (a:AHC {id: $institutionId})-[s:HAS_STATS_IN_YEAR]->(y:Year
    )
WHERE y.year >= $startYear AND y.year <= $endYear
RETURN y.year AS year,
       s[$indicatorCode] AS indicatorValue
ORDER BY year ASC
```

C.3 Frontend Implementation

The React presentation layer renders analytical visualisations using Plotly.js (interactive time-series, bar, and scatter charts) and supports CSV export of indicator tables. Role-specific page components (`CitizenPage`, `ManagerPage`, `FunderPage`, `AdminPage`) resolve to a shared `Dashboard` component with different API endpoint configurations, ensuring consistent UX across roles.

C.4 NLP Component

Natural-language user queries are handled by a **prompt-based LLM concept selector**. The complete list of biomedical concepts indexed in the Neo4j database (approximately 30,000 entries from OpenAlex concept hierarchies) is passed at inference time to Gemma3:27b (27B-parameter open-weight model served locally via Ollama; temperature = 0.0, context window 131,072 tokens) alongside the user's natural-language query. The model returns a JSON array of 3–10 concepts selected from the provided list that are semantically most relevant to the query.

This design differs from classical retrieval-augmented generation in that concept selection is performed through in-context reasoning over the full controlled vocabulary rather than through vector similarity search over embedded representations. Because the model can only return concepts explicitly present in the prompt, all selected entries are guaranteed to be valid database references, eliminating hallucinated or out-of-vocabulary terms.

This structural correctness guarantee applies to concept validity but does not address retrieval effectiveness with respect to user intent. A formal precision/recall evaluation against a held-out expert-annotated query set has not yet been completed; it is listed as a planned deliverable in Section 6 of the main text.