

UK PubMed Central: Re-imagining the venerable PubMed Central service for the life science community in the UK

Ernest Ong
Science, Technology and Medicine (STM)
British Library, UK

Abstract

Launched in 2007, UK PubMed Central (UKPMC) was originally released as a ‘mirror’ of the US PubMed Central repository. The feedback from our users and our experience of running the service prompted the organisations comprising UK’s principal funders of biomedical and health research to commission a new phase of development with aspirations to build the premier resource for biomedical and health research. This paper summarises the evolution of UKPMC, discusses the exploratory search facilities in the new web interface, and describes the architecture that underpins the new UKPMC service.

1. Introduction

UK PubMed Central (UKPMC) is an expanding information resource developed to serve the needs of the UK’s biomedical and health research community [1-2]. Launched in January 2007, the service is funded by the UK’s principal funders of biomedical and health research and was originally released as a ‘mirror’ of the US PubMed Central (PMC) repository developed by the National Center for Biotechnology Information (NCBI) as a free-to-access archive of full text journal articles. In 2008, the consortium delivering the service led by the British Library, in collaboration with the European Bioinformatics Institute (EBI) and the University of Manchester, was tasked by its funders, with the full support of the NCBI, to transform the full text repository into a premier resource and research tool for the biomedical and health research community.

The initial phase of developments included expanding the content base to incorporate all PubMed abstracts and a diverse range of grey literature, integrating advanced information extraction solutions based on text mining techniques, and developing a new web-based user interface that integrated these developments with a coherent user experience. The consortium took this opportunity to investigate enhancements to the “classic” PMC web interface, in close consultation with users from the research community. This paper describes the information retrieval and management facilities in the new UKPMC web interface that enable a fully

integrated exploratory search experience. An overview of exploratory search and review of the “classic” UKPMC web interface is presented before we introduce the new web interface, discuss its exploratory search facilities and describe the architecture underpinning the new developments.

2. Research Rationale

Exploratory search is open-ended, multi-faceted and intrinsically complex. It is characterised by an ill-structured or poorly understood problem context that motivates the search, and uncertainty in the search process to achieve some desired goal. The uncertainty and complexity associated with this class of search problems require the searcher to continually make qualitative judgements and exercise higher order thinking behaviours, including comprehension, analysis, synthesis and evaluation to facilitate learning and investigation as part of this process. Marchionini [3] distinguishes exploratory search from known-item searches, also known as lookups, that have a high degree of certainty involving problem domains the searcher is knowledgeable in. White and Roth [4], on the other hand, consider all search tasks to have an exploratory element to some degree, differing only in the extent of uncertainty. Moreover, they characterise the search processes for an iterative query refinement and exploratory search by the difference in traversals through the information space (Fig. 1).

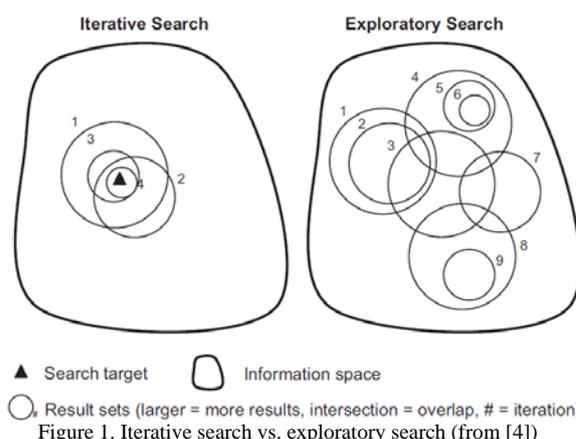


Figure 1. Iterative search vs. exploratory search (from [4])

The defining feature of the exploratory search process in Fig. 1 is its non-linear and multi-objective nature, not unlike probabilistic optimisation techniques, notably simulated annealing, that apply stochastic and opportunistic heuristics to minimise an objective energy function (cf. the perceived gap between the searcher's higher-order knowledge capabilities and the search goals). The analogy loosely based on simulated annealing provides a better characterisation of exploratory search where traditional deterministic linear models have been inadequate, as Wilson [5] noted in his paper describing an alternative model based on Tetris. In exploratory search, this energy function evolves over time as the search progresses reflecting the dynamic nature of the process, encapsulates the numerous search objectives (cf. to understand the problem domain or to define the search goal) and defines the size of the acceptable solution space (cf. the quantity and granularity of the information required). This macro-model of information seeking behaviour defines the overarching search process and complements the micro-models, such as those proposed by Ellis [6], that determine the form individual transitions might take. While Ellis' behavioural model does not prescribe the design for information systems, it identifies the range of activities a searcher may want to accomplish using these systems. The six primary behavioural patterns are:

1. **Starting:** Identifying relevant sources of interest as potential starting points of the search.
2. **Chaining:** Following up and connecting new leads from an initial source. Backward chaining involves following references from an initial source whereas forward chaining involves following sources referencing the initial source.
3. **Browsing:** Semi-directed searching in areas of interest by scanning content of identified sources.
4. **Monitoring:** Keeping abreast of developments in a given area by following selected sources.
5. **Differentiating:** Filtering, assessing and prioritising sources for usefulness based on prior experience.
6. **Extracting:** Systematically working through a given source to select material of interest.

Diriye et al. [7] revisited the definition of exploratory search from a human computer interaction perspective and proposed a model with four aspects:

- **Objective:** to retrieve a known fact or to produce some new knowledge. The objective of exploratory search is to produce new knowledge and inform some decision, and is generally abstract and open-ended.

- **Search Activities:** often described in terms of Bloom's taxonomy of learning objectives. Exploratory search involves higher-order activities, such as analysis, synthesis and evaluation that support learning and investigation.
- **Conceptual Complexity:** the degree of familiarity with and complexity of the problem context, search goal and search process. The search steps and target information are often vague in exploratory search.
- **Procedural Complexity:** the number and diversity of tasks required to achieve the search goal. Exploratory search tasks involve the execution and coordination of a number of complex search actions.

The conceptual complexity and procedural complexity aspects, in particular, have implications from a human computer interaction perspective and help inform the design of the user experience of exploratory search interfaces. White et al. [9] summarised the recurring themes that emerge in developing exploratory search tools, and reiterated the need to adapt these as user behaviours and our understanding of these patterns evolve over time. Some themes include:

- **Interactivity:** This covers a range of information and interaction modes including the use of facets, filters and actionable links to support navigation through the problem context and the search process.
- **Comprehensibility:** This aims to help the searcher formulate new insights, establish new connections and make informed decisions on navigation choices using techniques such as clustering, facets and visualisation. This addresses the lower order functions in Bloom's taxonomy thus enabling the searcher to focus on the higher order functions such as analysis and synthesis.
- **Usability:** As search tools become increasingly sophisticated, supporting a diverse range of activities, understanding the impact on user experience is of paramount importance. This orthogonal theme applies to the other themes and aims to reduce the searcher's cognitive load by minimising the conceptual and procedural complexity associated with using the search tool.

We will use elements of these models to discuss the exploratory search facilities in the new UKPMC web interface. First, we introduce the original UKPMC Classic web interface and discuss the motivation for the design of the new interface.

3. The UKPMC Classic Web Interface

The UKPMC Classic web interface is based on the portable PMC (pPMC) software developed by the NCBI that evolved from the US PMC software introduced in 2001. This ensured consistency in presentation style and, more importantly, compliance with the NCBI Business Rules that define the presentation requirements agreed in consultation with all publishers submitting to PMC. UKPMC was the first pPMC node in the network of PMC International sites, joined by PMC Canada in 2009. The availability and robustness of the pPMC software was instrumental in facilitating the launch of the UKPMC service in January 2007, under ambitious timescales.

In essence, the pPMC software is a full text article repository that provides access to articles and journal issues via a web interface. However, unlike most online repositories, the UK variant of pPMC does not provide any local search facility but instead relies on the search interface on the NCBI site in the US (Fig. 2). Consequently, searching UKPMC involved a roundtrip to the NCBI site (step 2, Fig. 2) with the article displayed back in UKPMC (step 3). This was sub-optimal for two reasons. First, to search UKPMC effectively required fluent knowledge of the NCBI search syntax and its search interface, with little support available in the local pPMC. Second, the one-service-two-systems user interaction model not only imposed a high cognitive load on the user, it limited the extent to which effective exploratory search could be integrated.

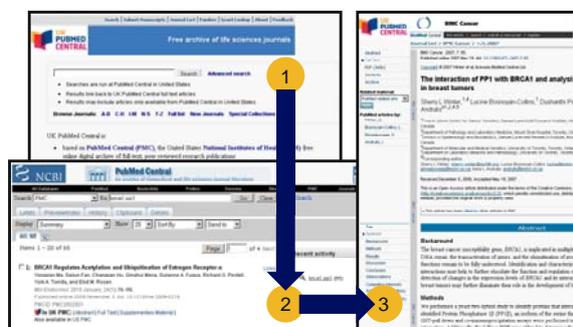


Figure 2. There and back again - the UKPMC Classic search workflow

Moreover, the content scope of pPMC was limited to full text articles in the PMC database (currently 2 million). A survey of UK biomedical and health researchers undertaken by the British Library in February 2008 highlighted the need to provide access to a larger and more diverse range of content in addition to enabling the discovery of and linking to pertinent content hosted *outside* UKPMC, ranging from biological databases to grey literature. In short, the users wanted a hub for biomedical and health research for the UK.

Finally, and most crucially, there was no means to modify or adapt the pPMC software developed and supplied by the NCBI, in order to respond to user needs or changes in scientific publishing practices. The feedback from our users informed a new 3-year phase of development to address the issues identified above. The result was a new UKPMC web interface, along with new backend systems underpinning it, including a modularised PMC, also referred to as the NCBI Backend. The new architecture provides a single point of access to a larger and more diverse range of content coupled with advanced information retrieval solutions and a richer and more coherent user experience.

4. The New UKPMC Web Interface

At the core of the technical development strategy for the new web interface is a user-centric co-innovation process based on the Living Labs model.¹ This entailed continuously engaging with our user community to understand their needs through scenario analysis, user stories and high fidelity interactive mockups, and evaluating how effectively the emerging developments addressed those needs. The engagement took the form of focus groups, one-on-one interviews and, to a lesser extent, social media and online feedback. To ensure that the technical developments were aligned with the open innovation process, we adopted agile development approaches with regular iterative releases, adhering closely to the tenet “release early, release often”. This strategy enabled users to keep abreast of emerging developments which, in turn, kept the developments in check.

To mitigate the risks and complexity inherent in the architecture for such an ambitious undertaking, we implemented a programme of inter-linked projects and developed strategic relationships to enable sustainable integration of external services. The result is a distributed architecture with three ‘families’ of services underpinning the web interface developed by the British Library:

- **EBI CiteXplore:** Deployed at EBI, CiteXplore provides the underlying search and bibliographic services in UKPMC, including full text search, query expansion, and contextually related citations (cited-by’s and references) and biological terms (curated and text mined).
- **NCBI Backend:** Deployed at the University of Manchester, the Backend produces modular output representing different fragments of a traditional pPMC page that the new web interface composes with other content fragments to render a complete page.

¹ http://en.wikipedia.org/wiki/Living_lab

- **NCBI Entrez eUtilities:** Deployed at NCBI and available publicly, the eUtilities suite of tools provide access across all NCBI databases including literature metadata, medical subject heading (MeSH) data, and contextually related articles and biological entities.

Rendering a single page in the new web interface can be highly data-intensive, involving numerous interactions with heterogeneous backend systems to produce a seamless and coherent user experience. Consequently, judicious data caching became a priority to maximise the responsiveness of the user interface. From a user perspective, the functional architecture comprises four elements, namely, information retrieval, content, content enrichment and information support (Fig. 3).

4.1 Information retrieval

This went beyond addressing the need for native resource discovery facilities and introduced user interface affordances to manipulate result sets and reformulate queries (Fig. 4).

The affordances enable the user to:

- initiate new searches or refine the existing search on a selected facet,
- revise an existing refinement search by removing an unwanted facet,
- restrict the results to specific content sets such as systematic reviews or guidelines, and
- switch between searching over citations and full text articles in the PMC database.

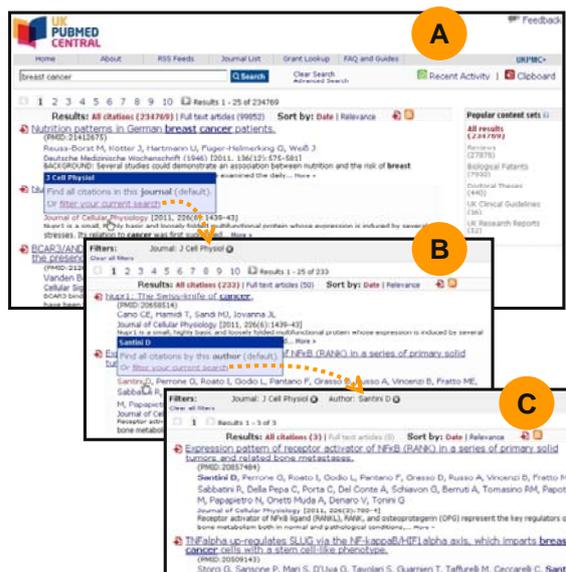


Figure 4. Query reformulation in the new UKPMC interface

4.2 Content

In addition to the 2M full text articles in the PMC database, UKPMC now provides access to 26M citations from authoritative sources of relevance to biomedical and health research including:

- 21M PubMed abstracts most of which are indexed with MeSH terms,
- 3.8M biological patents from the European Patent Office that includes pertinent patents filed in the United States,
- 528K records from AGRICOLA² relevant to biomedical research,
- 53K UK and international doctoral theses sourced from EThOS³ and NDLTD⁴ respectively, and
- 3K clinical guidelines sourced from the NHS.⁵

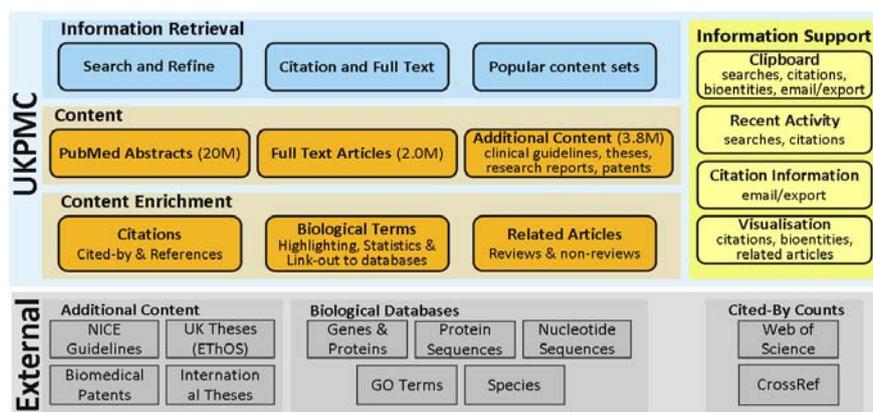


Figure 3. UKPMC Functional Architecture

² US National Agricultural Library citations

³ UK Electronic Theses Online Service

⁴ Networked Digital Library of Theses and Dissertations

⁵ UK National Health Service

4.3 Content enrichment

In keeping with advances in scientific publishing, UKPMC augments abstracts and full text articles with contextually relevant information drawn from both within UKPMC and without (Fig. 5). Due to the number and diversity of supplemental information available, enabling access to these sources while preserving the context of the primary content, the abstract for example, is crucial in reducing the cognitive and procedural demands on the user. The range of supplemental information includes:

- descriptions for MeSH terms (Fig. 5, A),
- biological terms highlighted in and linked from abstracts to authoritative databases such as UniProt and NCBI Gene (Fig. 5, A),
- related citations (cited-by's and references) including cited-by counts (Fig. 5, B),
- biological terms identified in the full text (curated and text mined) linked to authoritative databases (Fig. 5, C),
- related articles, comprising reviews and non-reviews, retrieved from NCBI Entrez eUtilities (Fig. 5, D), and
- links to 21K abstracts of systematic reviews and economic evaluations on CRD.⁶

4.3 Information support

Information support concerns the organisation and manipulation of known information objects to facilitate higher order thinking behaviours. These facilities are largely applicable to other functional elements in the architecture and include the following:

- *Recent activity* tracks all searches performed and citations viewed, and enables the user to revisit these objects and transfer them to the clipboard (Fig. 6, A). It is worth noting that refinement searches are displayed under its respective base search to preserve the context of the refinement. This helps the user understand the context of the search exploration and enables the identification of opportunities for new explorations.
- The *clipboard* is a pervasive facility that enables the user to collect information objects of varying granularity across multiple search episodes. These comprise search queries, citations including related citations and related articles, and curated and text mined biological terms (Fig. 6, B). In addition to providing convenient access to these objects, the clipboard can be exported to file or via email. The export function further enables lightweight collaboration between two or more users by supporting simultaneous use of a shared workspace for the duration the clipboard remains active. Unlike clipboards on most systems, user authentication is not required to access to the clipboard facility.
- The user can export *citation information* in the Research Information Systems (RIS) format that most reference managers support or send the same information via email. This facility is available for individual citations in addition to bulk export of citations from the clipboard. Furthermore, users can monitor search results for updates via RSS using the OpenSearch protocol.
- The *visualisation* of citation trends by

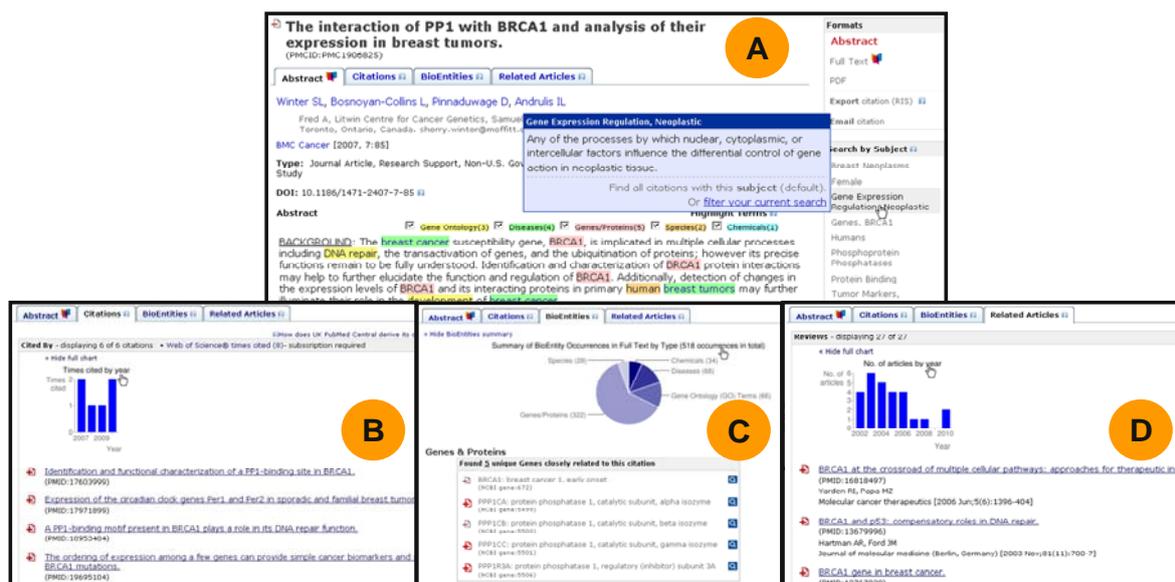


Figure 5. An abstract (already in Clipboard) with content enrichment

⁶ Centre for Reviews and Dissemination

publication year for cited-by's, references and related articles helps users make informed navigation decisions especially for large citation sets (Fig. 5, B). The cited-by trend in particular enables users to make qualitative judgements about the perceived scientific reputation of a paper over time. Likewise, the visualisation of the relative composition of a full text article based on the number of biological entity occurrences by type provides an insight to the main biological themes in the paper (Fig. 5, C).

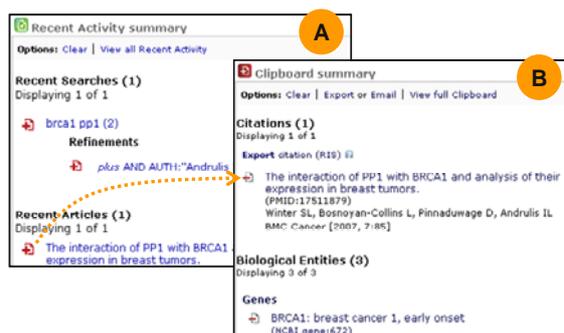


Figure 6. Recent Activity and Clipboard: transferring an article from Recent Activity to the Clipboard

5. Exploratory search in the new UKPMC web interface

Marchionini [3] emphasises the need for highly interactive user interfaces that enable exploratory searchers to engage with the search process. This reflects the open-ended and multi-faceted nature of exploratory search where the goal is higher order knowledge “creation” as opposed to known-item retrieval. Ellis’ model of information behaviour [6] provides a framework for discussing the extent to which search tools support the identified behavioural patterns. UKPMC addresses these as follows:

1. **Starting:** The three main sources for initial exploration are search, abstracts and full text articles. The entry routes vary and include direct access, bookmarks, RSS feeds via OpenSearch and, increasingly important, Google search.
2. **Chaining:** Chaining involves following up on links emanating to and from an initial source as a means to broaden the search. This includes following links to external providers of primary content such as full text articles on publishers’ sites and doctoral theses on EThOS, in addition to exploring contextually relevant supplemental information including related citations, biological terms and related articles.
3. **Browsing:** Browsing is the process of semi-directed searching in areas of interest to identify potential search opportunities.

UKPMC facilitates this on two levels. Firstly, by enabling convenient access to diverse supplemental information in the context of the primary content and secondly, by facilitating faceted search and refinement on both search and citation views. Moreover, the low procedural complexity associated with exploring related sources encourages exploratory browsing. For example, the same affordances for faceted search and refinement are used across both search and citation views thus ensuring consistency and enhancing usability.

4. **Monitoring:** Monitoring helps the user maintain awareness of emerging developments as potential sources of new search opportunities and occurs at three levels. Firstly, at the personal level, in the form of reviewing the audit trail in the recent activity and the objects in the clipboard to identify potential gaps. Secondly, at the social level, by monitoring contributions by co-searchers in the shared clipboard workspace. And, thirdly, at the content level, by enabling users to monitor search results for updates via RSS using the OpenSearch protocol.
5. **Differentiating:** Differentiating involves exploiting the characteristics of sources to evaluate their nature, quality and usefulness as a means to enable filtering and prioritisation. UKPMC enables both qualitative and quantitative evaluation as follows:
 - *qualitative evaluation:* (a) The availability of high quality metadata enables the user to make informed decisions based on the generic and, more importantly, domain-specific metadata, namely, MeSH. (b) Supplemental information provides the user with an overview of the themes in an article. This includes the provision of descriptions for MeSH terms and a summary of the biological terms in the full text.
 - *quantitative evaluation:* (a) Cited-by counts from UKPMC, based on CrossRef and the Web of Science enable the user to make qualitative judgements about the scientific community’s perception of quality. (b) The number of occurrences of biological terms in the full text may highlight the relative importance of specific entities and processes to a given scientific enquiry or method, for example. (c) The visualisation of citation trends and biological themes

enable the interpretation of large datasets.

6. **Extracting:** Extracting entails systematically working through sources to identify and select relevant material. UKPMC supports users on two levels. Firstly, by enabling the collection, sharing and export of information objects of interest using the clipboard in addition to integration with external tools such as reference managers. This focuses the scope of the review and forms the basis for subsequent exploration and search refinement to identify materials. Secondly, by highlighting, summarising and thereby “extracting” biological entities produced by curation teams and using text mining.

Ellis’ emphasis on behavioural patterns omits one crucial factor from a human computer interaction perspective - usability. The common user interface design goals of simplicity, consistency and learnability all have one aim – “to make simple things easy and difficult things possible”.

These are compatible with the factors proposed by Diriyee et al. [7] wherein the aim is to reduce the conceptual and procedural complexity associated with using the search tool while simultaneously providing affordances that aid the user to engage in higher order search activities. These were the considerations foremost on our minds when designing the user interaction model for the new web interface. The extent to which we succeeded in fulfilling these goals will be the subject of a subsequent study.

6. Architecture

The British Library led the consortium in the definition, implementation and management of the architecture that underpins the new UKPMC service in a 3-year programme of work. As consortium lead, the British Library was also responsible for

programme and quality management in addition to product design, agile user-centric functional requirements elicitation, non-functional requirements modelling and verification, system integration, release management, change management, and the implementation of the new UKPMC web interface.

6.1 Rationale

The definition of the target architecture was influenced by a number of factors:

- An aggressive development lifecycle including continuous user engagement throughout the lifecycle.
- Pre-existing knowledge and practices within the consortium prior to the development of the new UKPMC architecture.
- Distributed nature of pre-existing UKPMC sub-systems and other relevant services most notably the NCBI Entrez eUtilities and Thomson Reuters Web of Science hosted in the US.

These factors led to a distributed architecture (Fig. 7) that aimed to:

- Minimise the changes to the pre-existing architecture.
- Exploit pre-existing knowledge and practices within individual consortium members to minimise the disruption required to realise the new architecture.
- Reduce the need for re-skilling staff to adopt and manage new systems, practices and communications plan.
- Enable the flexible integration of external services to enhance core UKPMC offerings.

The clear separation of concerns and division of responsibilities within the consortium resulted in lower overall programme risks and enabled the delivery of the ambitious implementation plan within a highly aggressive schedule. However, the quality of the architecture was not compromised because of

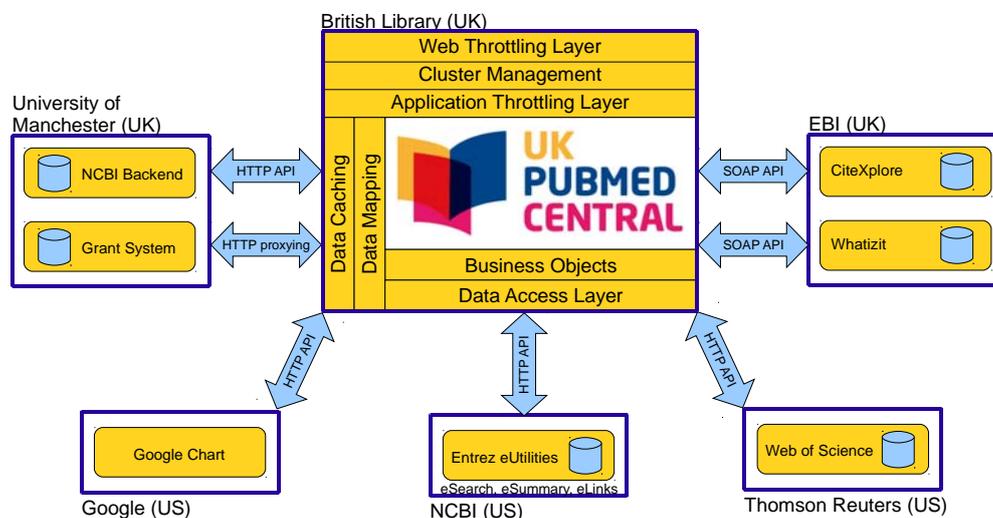


Figure 7. Logical System Architecture

these challenges. The new UKPMC web interface, for example, adopted a set of architectural and design patterns that provide an *extensible dependency layer* using service adapters, data access abstraction, data caching, and data mapping and transformations. This reduced the cost and complexity of implementing robust and performant integration solutions. Examples include the integration of a range of tools developed and enhanced as part of the UKPMC programme, including NCBI Backend and EBI CiteXplore, in addition to an array of reputable and pertinent services available outside UKPMC including NCBI Entrez eUtilities and Thomson Reuters Web of Science (Fig. 7). Furthermore, the reuse of *modular user interface components* ensure a consistent user experience, the pervasive clipboard facility being a prime example, while minimising code duplication and reducing technical complexity. The resulting framework made it possible to integrate the EvidenceFinder service, an advanced semantic search developed by the National Centre for Text Mining (NaCTeM) based on deep linguistic analysis, incorporating extensive user interface enhancements into a “UKPMC Labs” site in under a week with minimal resources. From a technical perspective, UKPMC is a service broker and orchestration engine that provides an optimised user experience for accessing high-value biomedical and health research content and tools based on service-oriented architecture principles.

6.1 Quality Management

A comprehensive approach to quality management is vital to the delivery of a service of this size and complexity particularly given the aggressive delivery schedule. In broad terms, these aspects included:

- **Architectural requirements:** This entailed an intensive consultation process early in the programme involving key stakeholders, development partners in the UK and the US, and the end-user community to evaluate trade-offs and risks, and to ensure that the target architecture was fit for purpose.
- **Functional requirements:** The programme adopted a user-centric co-innovation process based on the Living Labs model to ensure continuous user involvement throughout the agile implementation lifecycle.
- **Non-functional requirements:** These complement the user experience design undertaken as part of the user consultation process and included non-functional requirements modeling, accessibility audit, capacity and load testing, and extensive web analytics to facilitate deep log analysis.
- **Release management:** In addition to rigorous configuration management and

metrics-based change management practices, the programme also adopted short regular release cycles and user acceptance testing in the staging environment in conjunction with pre- and post-deployment testing in the production environment.

- **Service operations management:** This encompasses a range of aspects including:
 - *End-user support* via the UKPMC Helpdesk that feeds into the change and incident management processes.
 - *Cluster management with redundancy and failover* across application sub-systems including web applications and database services (Fig. 7).
 - *Access throttling* via independent web and application throttling capabilities built on a set of iteratively refined heuristics (Fig. 7). This aimed to strike a balance between enabling access to legitimate end-users against preventing access to scripted and harvesting bots that violate the terms of use of the service.
 - *Runtime profiling* of service invocations and user scenarios to monitor compliance against specified service level targets.
 - *External monitoring* of the UKPMC site from cloud-based services.
 - *Internal monitoring and self-healing services* capable of detecting and reporting faults on individual UKPMC sub-systems in addition to healing and resolving faults identified in sub-systems managed by the British Library.
 - *Runtime self-diagnosis and dynamic adaptation* of reliability- and performance-critical components in the UKPMC web interface systems.
 - *Remote management* of servers and sub-systems for authorised personnel via approved mobile devices using secure tools.

7. Conclusion

UKPMC launched in 2007 as a remote ‘mirror’ of the US PMC repository with minimal search support. It has since evolved into a rich information hub for biomedical and health research in 2010 that supports a rich set of exploratory search facilities, underpinned by a robust architecture and an expanding content base currently comprising 26M citations and 2M full text articles. Since the launch of the new UKPMC web interface, its user base has

expanded significantly with non-UK traffic now accounting for more than 80% of all traffic. At the core of the technical development strategy is a user-centric co-innovation process that aims to understand the information seeking needs of the biomedical and health research community and align the technical developments with these.

We took this opportunity to implement significant enhancements to the underlying architecture that enabled the realisation of exploratory search not previously feasible under the pPMC-based architecture. The new capabilities provide an optimised user experience for accessing best of breed and reputable biomedical and health research content and tools.

It is still too early to evaluate the effectiveness of exploratory search support in UKPMC. To do so will require detailed quantitative and qualitative analysis of user behaviours and preferences over a longer period. The existing exploratory search facilities serve merely as a starting point to understand the challenges in enabling exploratory search in biomedical and health research. We anticipate that these developments will be supplemented with technical application programming interfaces (API) that will enable the integration of these exploratory search capabilities on external sites. The outcomes of this analysis in conjunction with insights gained from on-going engagement with the biomedical and health research community will inform subsequent refinements of the UKPMC service and its architecture.

References

- [1] P. Davey, "UK PubMed Central: becoming the information resource of choice for the UK's life sciences research community", *Interlending & Document Supply*, 2010, 38(2):102 - 107.
- [2] J. McEntyre et al., "UKPMC: a full text article resource for the life sciences", *Nucleic Acids Research*, Vol. 39, Database issue, 2011.
- [3] G. Marchionini, "Exploratory search: from finding to understanding", *CACM*, 2006, 49(4):41-46.
- [4] R.W. White and R.A. Roth, "Exploratory Search: Beyond the Query-Response Paradigm", *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 2009, 1(1):1-98.
- [5] M. Wilson, "The Tetris model of the information seeking process", Technical Report, Department of Computer Science, Swansea University, 2009. (unpublished)
- [6] D. Ellis, "A behavioral model for information retrieval system design", *Journal of Information Science*, 1989, 15(4-5):237-247.

[7] A. Diriye, M.L. Wilson, A. Blandford and A. Tombros, "Revisiting Exploratory Search from the HCI Perspective", in: *4th Int. Workshop on Human-Computer Interaction and Information Retrieval*, 2010. (in press)

[8] T.D. Wilson, "Models in information behaviour research", *Journal of Documentation*, 1999, 55(3):249-270.

[9] R.W. White, B. Kules, S.M. Drucker and M.C. Schraefel, "Supporting exploratory search", *CACM*, 2006, 49(4):36-39.