A Transformational Search and Discovery Interface for the ePrints Server Institutional Repository of the University of Southampton:

A Working Paper for the University of Southampton ePrints server

Revised and re-Submitted 25th January 2025

Local EPrints ID: 478240

PURE UUID: 266d3f06-828a-4830-8156-fcf6793391e9

David A Rew MChir FRCS

Honorary Consultant Surgeon,

Faculty of Medicine, University of Southampton

(https://orcid.org/0000-0002-4518-2667)

Professor Leslie Carr

Professor of Web Science, Faculty of Electronics and Computer Sciences (ECS)

University of Southampton (UoS)

https://orcid.org/0000-0002-2113-9680

Harry Nelson

Alexander Banwell

Philip Stambolov

Liulan Qian

Undergraduate Students

Faculty of Electronics and Computer Sciences

Correspondence: Mr David Rew: dr1@soton.ac.uk

Key Words: ePrints archive, Lifelines, data visualisation, dynamic interface, subject

taxonomy, dynamic iconography; synchronised timelines

Introduction

This preprint paper for the UoS ePrints server reports the development of a novel Interface API for the ePrints family of Servers for academic institutions. The work was undertaken as a student team project within the Faculty of Electronics and Computer Sciences of the University of Southampton in the Autumn of 2022. It builds upon the experience in the development and implementation of the Lifelines interface to the Electronic Patient Record (EPR) at University Hospital Southampton between 2009 and 2024.

UHS Lifelines represents a first-in-class practical exemplar of a novel form of interface to complex data sets, as exemplified by the clinical EPR and by institutional records repositories. The display of document links as dynamic icons on stacked, synchronised timelines which are defined by a relevant subject taxonomy allows the user to secure an immediate overview of the data set and the patterns within it, whether relating to the documentation in the EPR for clinical practice or the corpus of academic work relating to an individual researcher or a University Faculty.

We tested the hypothesis that it was possible to build an API around this interface concept as an undergraduate student project. In three months and at minimal cost, the resulting draft product fulfilled all of the key specifications of the "E-Print Interface" to a high standard of test functionality.

The project evidenced a simple and effective format for the real time overview of the entirety of academic document libraries at personal academic and faculty level. Subject to professional refinement of the draft API, we consider that the model can readily be adapted to the radical simplification of document search and navigation for ePrints servers, digital libraries and many other custodian information systems.

We would be pleased to explore the wider development of this concept in a range of practical applications.

Structured Abstract

Background: The Stacked and Synchronised Timeline and Iconographic Interface ("Lifelines") is a significant practical advance in the efficient overview, visualisation, navigation and interrogation of complex, heterogenous and time-structured data sets.

The Southampton ePrints system has been a repository for the outputs and open publication of all members of staff, Departments and Faculties of the University of Southampton (UoS) since 2004. Versions are used by some 400 institutions worldwide. ePrints has been optimised for internet searches on individual documents rather than for the visualisation of the global outputs of individuals or departments.

Methods: We used as a template for this data visualisation project the unique University Hospital Southampton (UHS) Lifelines interface, which has been created and adapted as a central element of the UHS Electronic Patient record (EPR). We created an external application programming interface (API) to the UoS ePrints server to extract the content on demand to populate an "ePrints Lifelines" API, using the ePrints taxonomy of output types, (eg papers, theses) to designate our content taxonomy and related timelines.

Results: In 3 months, using an agile methodology, a team for four undergraduate students created a working API which reliably accesses all content in UoS ePrints in real time. Specifically, it populates a timeline framework which permits direct visualisation and navigation to the source outputs of any individual, department or Faculty whose material is posted on ePrints. Citation metrics are also displayable for each output.

Conclusions: Our "ePrints Lifelines" interface demonstrates a transformative interface for the navigation and interrogation of heterogenous archival content in a fast and intuitive way, and which can be engineered with modest coding resources. It introduces a new approach for libraries, institutional, corporate, personal and national archives, ancestry record systems, bibliometric databases, and engineering systems substantially to enhance the value of their information holdings.

INTRODUCTION

The world abounds with large, complex, heterogenous and time-structured data sets. Libraries, personal, institutional, organisational and national information archives, ancestry record systems, bibliometric databases, and the maintenance records of complex engineered entities, are but examples. Globally, digitisation of catalogues and content is well advanced, as are private, public and commercial initiatives to make the underlying content searchable and readable.

Nevertheless, there are significant limitations in legacy software solutions to the visualisation and dynamic navigation of an entire archival data set. A data search may involve multiple screens, frames, windows, menus and lists to generate an overview of the content, which may be output of an author, a department, or a subject field from a large archive.

Legacy software systems impose therefore significant time and attention challenges upon users. These constraints lead to inefficient use of time and human resources, and to the risk of missing key elements and patterns in the data set. This in turn leads to errors in interpretation of the available information and in decision making.

The University of Southampton (UoS) ePrints system

The UoS ePrints server (https://eprints.soton.ac.uk/) was developed in the early 2000s as a tool for building high quality and highly configurable Open Archives Initiative (OAI)compliant repositories. It found widespread use in the collation, archiving and open access publishing of a range of academic outputs of members of the University. (1. Harnad 2001, 2. Hey et al 2004, 3. White 2008, 4. Gumpenberger 2009).

ePrints has since been adopted by some 400 other institutions. It also provides: for the import of data from other repositories and services; a quality-controlled code framework; and a flexible plug-in architecture. It is optimised for Google Scholar searches, and for a range of bibliography managers. ePrints 3 is the latest version of the software.

As such, an institutional ePrints archive may contain many hundreds of career outputs from an individual member of staff, and many thousands of outputs from a Faculty, School or Department. However, there is no legacy function in the server which permits the collation and overview of those outputs in a simple, coherent and temporally structured format.

The dynamic visualisation of digital data sets

The optimisation of human interaction with large, heterogenous and temporally structured computer data sets, has been an enduring and resource intensive challenge (5. Shneiderman et al 2000, 6. Morris and Yen 2004, 7. Morris and Boyack 2005).

In the mid 1990s, Professor Ben Shneiderman at the Human Computer Interaction Laboratory of the University of Maryland proposed a key mantra for the computer aided visualisation of complex data sets (8. Shneiderman 1996, 9. Shneiderman 1997). He opined that any data visualisation solution must provide an Overview of the entire primary data set to orientate the user. From this reference and orientation framework, the user can navigate to examine details on demand, while filtering out information which is extraneous to the immediate purpose.

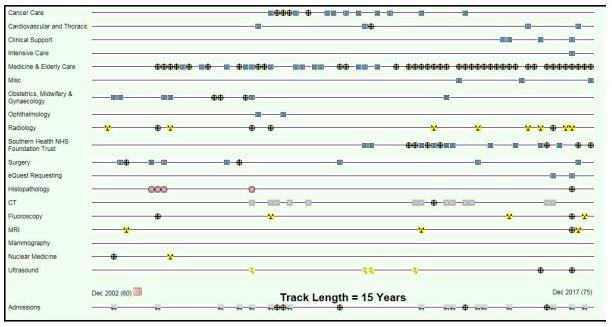
The Lifelines exemplar of a dynamic data visualisation system

Shneiderman's team subsequently introduced a concept for the visualisation of the individual police or medical record, which they designated "Lifelines" (10. Plaisant et al 1998, 11. Plaisant, Mushlin et al 1998) Lifelines is a two dimensional visualisation matrix which is characterised by stacked and synchronised timelines to present a subject taxonomy on the Y axis, while time is represented on the X axis.

Each time stamped data item, event, document or report is represented by a dynamic icon, which is accurately located in two dimensional Lifelines space by metadata on subject and time. The icon opens to a virtual third dimension, which is the subject content.

Working from this conceptual template, a small team at University Hospital Southampton (UK) (UHS) developed the Lifelines concept into a practical dynamic interface for the Electronic Patient Record (EPR). In the course of this project, we resolved a number of

practical challenges in the complexities of timeline structured data presentation, navigation and compression. UHS Lifelines now sits at the centre of the UHS clinical data estate for some 5,000 authorised users, and it addresses some 3 million patient record sets and Figure 1. (12. Hales et al 2019, 13. Rew et al 2021)



Copyright @2010-2017 University Hospitals Southampton

Figure 1. A live screenshot of the clinical record set of a 75 year old lady with multiple comorbidities in 2017. The entire dataset can be visualised and navigated on a single screen. Clicking on any icon opens the document(s) at that point on the relevant subject timeline.

"Lifelines" as a universal interface format for large data archives

We recognised that the UHS Lifelines EPR was an exemplar for a large number of potential applications in e-archive management, where data overload could overwhelm the user when using legacy computer visualisation methods.

Critically, the "Lifelines" interface adds or subtracts nothing from the source data. It simply reorganises information into a format which is optimised for the human eye and brain. We therefore wished to test the data model in the University of Southampton e-print server, as a well structured, locally accessible digital exemplar archive.

In this paper, we describe the successful proof of concept in an undergraduate project to extend the dynamic Lifelines interface concept into archive management using the UoS ePrints server as our development model. We have names this project "ePrints Lifelines".

METHODS

The project was allocated to four undergraduate students of the Faculty of Electronics and Computer Sciences (ECS) at UoS as a project over one semester, to write fresh code to build a Lifelines-like interface to the UoS ePrints server.

The General Brief

The brief was to address a range of user cases, which included the following:

- "As a senior executive of the University, I wish to be able to showcase the outputs of the University in a trustworthy, efficient and effective manner";
- "As a Faculty/Departmental lead, I wish to be able to visualise the outputs of individuals and units in a trustworthy, efficient and effective manner",
- "As an academic member of the University, I wish to be able to visualise my personal outputs for career development purposes".
- "As an external user of the UoS digital resources, I wish to understand the global outputs over time of an individual, department or faculty for my own purposes".

Specific objectives

The high-level requirements for the new interface to the ePrints server were for:

- -. A reliable and trustworthy visualisation of the ePrints server content for any researcher, department or Faculty of the University.
- Interactive functionality through the interface to each item in the data set
- The use of a logical subject taxonomy for the content timelines
- Speed of generation of the interface and onward navigation in any search.
- Intuitive usability to minimise user instruction
- The ability to view the output of multiple researchers or a Faculty on one page
- customisability of the X-axis timeline to view any defined period of outputs

General methodology

Overall, the project required:

- 1. Development of the code framework for the new timeline interface
- 2. An understanding of the data and records management within the ePrints system
- 3. The development of a system to identify and retrieve the specific information that we required from the ePrints server for any particular user query.
- 4. The development of a method accurately to allocate the retrieved data to our interface framework.
- 5. The ability to perform these functions economically and with minimal delay for the user. Given the time and resource limits of the project, we did not seek to apply any data analytics or other advanced functionality to our product.

We also determined at the outset to build a module that sat externally to the ePrints server, rather than remodel the intrinsic code of the live system.

Specifically in respect of the timeline interface design:

- We sought to replicate the general functionality of the UHS Lifelines model.
- We adapted the native subject taxonomy within the ePrints server.
- Timelines would only be displayed where content existed for that category.
- We had freedom to design a suitable iconography.

The accessibility of data from the University of Southampton e-print server

The ePrints server is based upon a large relational database. In response to submitted search parameters, it permits a range of export formats for the search results by generating Uniform Resource Locators (*URLs*) that return results in JavaScript Object Notation (JSON). Our "search" query for ePrints returned a URL that included the following parameters:

- name (for a single researcher)
- names (for multiple researchers)
- sort (whether to return the categories based on keywords or by type of record)
- dateStart (beginning of the period to search)
- dateEnd (end of the period to search)
- faculty (an id relating to a faculty category listed on ePrints)
- keywords (allowing for a similar functionality to the simple search, to search by topic)

ePrints Timeline Sequence Diagram

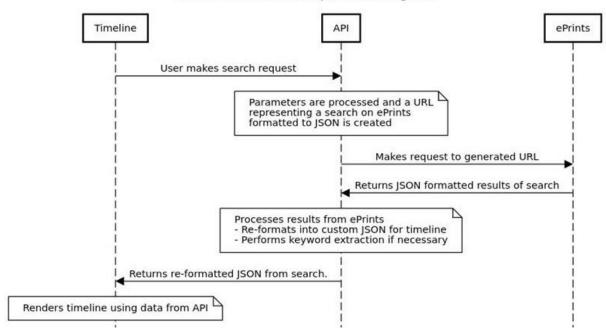


Figure 2a

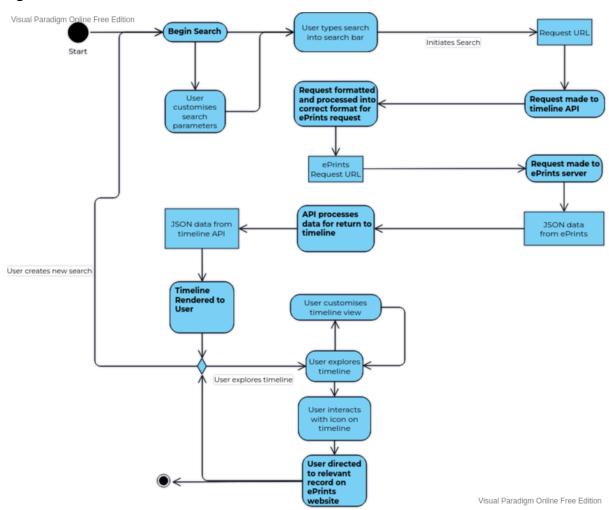


Figure 2b.

The frontend to our ePrints Lifelines interface, was designed to visit the Search URL, and then to process the JSON data which was returned into timeline interface format. A dictionary mapping tool allocated records to a specific category in the subject taxonomy of the interface (Figures 2a and 2b).

ePrints provides much information in each record which is not needed in our interface design. The required Attributes that were always present in the output from ePrints were:

- The URI of the record (string)
- Title (string)
- Contributors (list of strings)
- Date (string ePrints returns this as either an integer if it only has a year provided or a string in the format YYYY-MM-DD. We convert it to a string if it is an integer for consistency.
- Publication Title (string)
- Type (string conference, article, etc.)
- DOI (string)
- ePrints ID (string)

The only fields which were needed to create and populate each subject timeline were:

- a category for the Y-axis,
- a date for the X-axis,
- the title the record
- the URL, with which the user can track the record on the ePrints website.

There is also an option of a "keywords" search from the title of the publication. If the attribute was not present but was required by the timeline it would be auto-generated by the keyword generator.

The management of returns for large numbers of records

Searches on the outputs of multiple researchers or of a Department or Faculty could yield very large numbers of returns. This slowed the processing and overloaded the interface. We therefore put a cap of six months on the default date range on such searches to speed up the search to within three seconds and to de-clutter the interface. The date range could be

nevertheless be customised in the same style as in the ePrints advanced search, by specification of a start date and an end date.

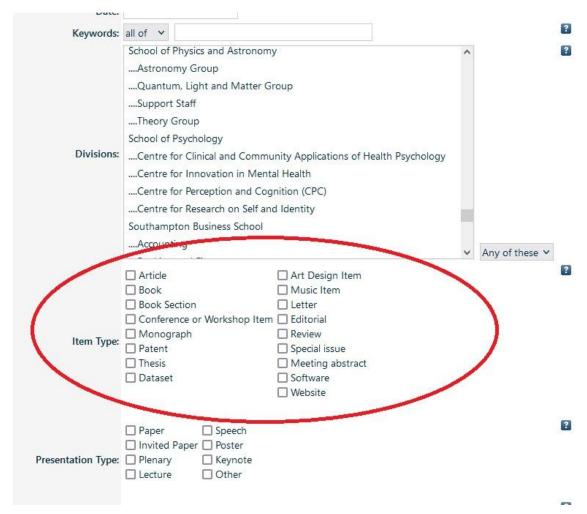


Figure 3. This is a screenshot from the search screen of the UoS ePrints system. In this figure, the "Divisions" section (upper screen segment) identifies all of the Departments and Faculties of the University, current and historic, for which academic outputs exist. The key outputs are themselves defined as Item Type (red circle) and Presentation types.

The selection of the Y axis subject taxonomy

The subject taxonomy of ePrints Lifelines was dictated by that in the ePrints server, as illustrated by the image of the search definition screen (Figure 3). In this figure, the "Divisions" section (upper screen segment) identifies all of the Departments and Faculties of the University, current and historic, for which academic outputs exist. The key outputs are themselves defined as Item Type (red circle) and Presentation types.

We did not include the Presentation Types taxonomy in the final product, as there was limited user-populated data in these subsidiary categories.

The Use of Serverless Architecture and Amazon Web Services:

We used Amazon Web Services (AWS) to host the ePrints Lifelines project and to manage the deployment of the front and the back end of the product. We used the open source Vite Serverless Framework to construct and deploy our code to AWS.

The project was deployed using SST.Dev (Serverless Stack) software. SST is an open source framework for building full-stack applications on AWS. Full stack systems permit the building of both the front end and the back end of a software application. SST supports JavaScript, TypeScript, Python, Golang, and C#.

We wrote the application in Typescript using the VueJS framework. The SST Framework builds, packages, and deploys code with the one command: serverless deploy. This approach provided modularity for future development and integration, and simplicity of simultaneous development and testing by members of the team. It also allowed for a local debugging mode, although the logging output format had to be configured for each instance of deployment of the tool.

Disadvantages of this approach included limitations to the pre-processing of data. Data models and libraries had to be loaded into AWS whenever a request was made, thus limiting import of data from the ePrints Server at scale.

Testing of our development system

We worked in an agile manner to optimise the product to the conceptual objectives and to make minor design adjustments using iframes, as a simple technical solution. An inline frame (iframe) is a HTML element that loads another HTML page within the document. The demo pages were hosted using Git Pages, which are a free hosting service from Github. We also used the UoS Gitlab (https://about.gitlab.com/)

We used Figma (https://www.figma.com/) for the initial designs of the visual elements of the product. Figma is a cloud-based wireframing tool which allows for testing of basic functionality; and for planning and design purposes. It is also allows multiple team members to work on it simultaneously.

The ePrints Lifelines test models deployed with the relevant JSON files, which were read using the file system ("fs") module. The file logger was poorly signposted on the AWS website, leading to challenges and time penalties in debugging the test models.

Unit Testing: We used both manual and automatic testing during development. We lacked experience with testing libraries for NodeJS, and so we wrote our own Unit Testing functions to generate a series of sample searches which we could then validate manually.

Scenario Testing: We wrote step-by-step instructions which were based on the original user cases. We then tested each function, for example to generate the timelines, log the searches and check returns to verify the performance of the system.

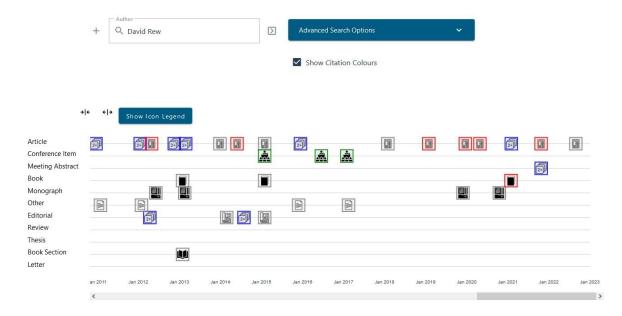


Figure 4. A live screenshot of the publications of the author, as listed on the University of Southampton ePrints server between 2011 and 2023. The X axis slider allows navigation along the timeline (back to 1978 in this case).

RESULTS

Our project resulted in a functional software product, ePrints Lifelines, which is written in Typescript Code using the VueJS framework. It is hosted in AWS on a free license, subject to data processing limits. It meets and exceeds the original user specification, and it has now been released to University colleagues across all Faculties for further testing.

ePrints Lifelines is an interactive data visualisation system for the ePrints archive. It is a free-standing Web Application which is connected to the ePrints Soton server, and which expresses features which are very similar to the original UHS Lifelines template.

It generates a 2-dimensional multiple timeline rendered computer screen interface (Figure 4). It is presently optimised for desktop and laptop screen use, as our resources did not permit rendering for hand held mobile devices in the available time.

Key features of the Display

The X-axis timeline is controlled by a slider, and the length of the timeline can also be modified by "expand" and "compress" arrow functions.

The Y-axis displays the content taxonomy as defined within the e-print server. It displays those timelines for any given individual, Department or Faculty for which content exists.

Each academic output is identified by a large icon measuring 24 by 24 pixels, which is centred on its publication date from the e-print server data. Each icon opens to the relevant content in the e-print server. Each timeline displays distinctive icons whose design reflects the particular subject taxonomy for that timeline.

Hovering over the icon displays the date of accession, the title/description metadata and the citation count. The colour of the icon can be set using the "Show citation colours" icon to indicate whether or not an output has been cited, and if so, how many times. The citation data is obtained through an external API (14. Harnad S and Carr L 2000) which can slow the load of the interface.

The management of overlapping events and icons

In some circumstances, more than two documents or events coincide or are closely related in time. This situation leads to two or more records overlapping or displacing each other on the timeline. We therefore combine such events into a unitary "2+" icon to preserve both the temporal structure of the interface and easy access to each item of content. Hovering over the "2+" icon displays a list of the underlying content. Clicking on such an icon opens to a list of hyperlinks to the relevant ePrints server content pages.

Search functionality

A user can search for an author (or multiple) by name (forename, surname); by their ePrints ID; by a keyword, or by a Faculty, and every record posted in the given Lifelines will be shown on the timeline. Additional options for customising the search are contained within an "advanced settings" drop-down that can be opened and closed to improve viewing of the timeline.

DISCUSSION

General observations

The stacked and synchronised timeline and iconographic data visualisation interface is a powerful tool for the unitary display and interaction with heterogenous data from single or multiple data systems. It provides a highly efficient way of visualising "whole of life" data, whether in a medical record or an academic career.

The purpose of this project was to test the adaptability of the concept from an established medical electronic Patient Record (EPR) application to an exemplar data archive with time stamped records and a bespoke subject taxonomy, of which there are many other such examples worldwide.

To this end, a small team of undergraduate students were able to build a working interface to the Southampton ePrints server which achieved all of the key design objectives. For the first time, using the beta version of the software product, and subject to final acceptance

and enterprise hardening of the product, users are now able to see, navigate and access the entirety of their academic output on a single and highly intuitive interface.

By developing the software as an external API on AWS, work progressed with no disruption to the code or performance of the ePrints software itself.

Our solution, ePrints Lifelines, promises to revitalise ePrints and create new user cases for a system which has been widely adopted by academic institutions for open access publication of locally created content. This upgrade offers a number of advantages to users, in that users can:

- easily review and audit their listed content, and check for errors that have crept into the system during data entry
- easily navigate the ePrints server without having to search for each output individually via Google or Google Scholar;
- use an alternative, fast and efficient search system to review the outputs of colleagues, staff, Departments and Faculties for a range of use cases.

Constraints

We recognise that this project was conducted at pace under the constraints of an undergraduate project, and that the ePrints Lifelines software product has room for improvements and for further functionality. It has also highlighted room for improvement and modernisation of the ePrints system itself.

One evident problem is the lack of selectivity of the search function for common names, wherein the search will return all examples of outputs of all "John Smiths" on the interface. The obvious solution to this problem would be to adopt the Orcid identifier for all individual person searches.

Orcid post-dates the ePrints servers, so universal adoption of Orcid Identifiers (ids) would involve retro-filling of the ids of all current staff. This in turn would require that all academically active staff register with Orcid and confirm their id to suitably updated EPrint system. This would still leave "Orcid-naive" content from authors who had retired, had

moved to other institutions or died. It is not clear whether such an updating effort on retrospective content would be justified at this point.

Like many academic institutions, the University of Southampton has adopted modern Research Information Management Systems which in our case is the Elsevier PURE system. PURE in turn takes citation data and titles from SCOPUS, from the Web of Science, and (in the UK) from the 2014 Research Excellence Framework data set. PURE data is in turn exported to Orcid, which itself is now creating a repository of outputs for Orcid-listed researchers.

Therefore, the ePrints server sits in a much more complex public data environment than that which existed when it was originally conceived. Nevertheless, PURE is an internal UoS system. The opportunity for institutional academics to demonstrate the totality and the specifics of their academic output to the outside world, and to publish in an open access format content that would not normally find its way into the catalogues of commercial publishers, (for example Theses, internal papers and multimedia and performance outputs), suggests an important future for the institutional ePrints system.

Technical Matters

We are conscious that our solution is currently specific to desktop and laptop devices. Clearly, a solution that could displayed as an App in mobile and hand held devices would extend the UoS ePrints server content more conveniently and to a much wider research ecosystem.

The decision to develop an external API to access the ePrints system gave us considerable flexibility in the technologies we tested when compared with attempts to develop a solution which was internal to the ePrints software. We were also able to work independently of the ePrints Development Team. However, we had to develop workarounds to access the source software and to reverse engineer the request format, as there was no documented external API for ePrints. The search to the public ePrints page therefore had to be simulated.

The lack of internal access to the ePrints database also created an added overhead for the network request to ePrints for each search made in our product. This reduced the time available to process the data for rendering on ePrints Lifelines.

Nevertheless we have delivered a functional product with features over and beyond the original general specification at the outset of the project, as for example, by building a "Faculty Search Function" into the product. This outcome was facilitated by an agile and iterative process with regular discussions and supervisions as the project progressed, rather than working to a rigid waterfall design specification.

Critically, were able to build a fast, responsive and intuitive data visualisation Application which we believe improves upon the legacy ePrints search tool at negligible cost but for the time and commitment of the development team.

Acknowledgements

We are grateful to Matthew Warren and to Alan Hales for advice on the Lifelines architecture.

References

- 1. Harnad S.(2001). For Whom the Gate Tolls?: How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now file:///C:/Users/Acer/Documents/Army/harnad%20ePrint%20server%202001.pdf
- 2. Hey, J. M.N. and Simpson, P. (2004) Building an Institutional Research Repository Based on User Requirements a University of Southampton Case Study. At, *Institutional Repositories: the Next Stage, Washington, DC, USA, 18-19 Nov 2004*.
- 3. White W. (2008): Case history: University of Southampton, e-Prints Soton OR2008 Open Repositories Conference 1-4 April 2008
- 4. Gumpenberger C. (2009): The EPrints story: Southampton as the cradle of institutional archiving . GMS Medizin Bibliothek Information 2009, Vol. 9(1), ISSN 1865-066X http://www.egms.de/en/journals/mbi/2009-9/mbi000138.shtml
- 5. Shneiderman B, Feldman D, Rose A, and Ferré Grau X. (2000). Visualizing digital library search results with categorical and hierarchical axes. In Proceedings of the fifth ACM conference on Digital libraries (DL '00). Association for Computing Machinery, New York, NY, USA, 57–66. https://doi.org/10.1145/336597.336637
- 6. Morris, S. A., & Yen, G. (2004). Crossmaps: visualization of overlapping relationships in collections of journal papers. Proceedings of the National Academy of Science of the United States, 101(suppl. 1), 5291-5296.
- 7. Morris S.A. and Boyack K. (2005). Visualizing 60 years of anthrax research. Proceedings of ISSI pages 45–55. Edited by P. Ingwersen and B. Larsen. Pages 45–55, 07 2005.
- 8. Shneiderman B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proceedings IEEE Symposium on Visual Languages 1996; 336-343

- 9. Shneiderman B. A Grander Goal: A Thousand-Fold Increase In Human Capabilities Educom Review, 1997; 32: 6, 4-10.
- 10. Plaisant, C., Shneiderman, B., Mushlin, R. (1997)

 An Information Architecture to Support the Visualization of Personal Histories. *Information*

Processing & Management, 1997; 34: 5, pp. 581-597.

- 11. Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., Shneiderman, B. (1998). LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records Proc AMIA Symp 1998; 76-80
- 12. Hales AA, Cable D, Crossley E, Findlay C, and Rew DA. (2019)

 Design and implementation of the stacked, synchronised and iconographic timelinestructured electronic patient record in a UK NHS global digital exemplar hospital. BMJ

 Health & Care Informatics, 26(1).
- 13. Rew DA, Hales AA, Cable D, Burrill K. and Bateman AC. (2021)

 New life for old cellular pathology: a transformational approach to the upcycling of historic e-pathology records for contemporary clinical uses

 Journal of Clinical Pathology Published Online First: 16 February 2021.

 doi: 10.1136/jclinpath-2021-207385
- 14. Harnad, S., & Carr, L. (2000). Integrating, navigating, and analysing open EPrint archives through open citation linking (the OpCit project). *Current Science*, *79*(5), 629–638. http://www.jstor.org/stable/24105079