

Managing Open Access Publication: a system specification

Contents

[1. Executive Summary](#)

[2. Purpose of this Document](#)

[3. Requirements Gathering Methodology](#)

[4. Scope](#)

[4.1 Summary of Core Requirements](#)

[4.1.1 Centred Around Academic Outputs](#)

[4.1.2 Task and Data Driven](#)

[5. Workflow](#)

[6. Logical Data Model](#)

[6.1 Notes on the Data Model](#)

[7. Integration with Other Systems](#)

[8. Functional Design and Wireframes](#)

[8.1 Overall Navigation and List views](#)

[8.1.1 Academic Output List View](#)

[8.1.2 Tasks View](#)

[8.2 Data and Task Display](#)

[8.3 Role Management](#)

[8.4 Reporting](#)

[Appendix A - Data Elements](#)

[Appendix B - Wireframes](#)

1. Executive Summary

This document provides a specification for UK HE institutions to manage data in relation to the publication of open access content. The specification was created through consultation with the UK HE community.

The system specified in this document is intended to enable institutions to record and report on data relating to the publication of open access outputs generated by their staff. This includes outputs that are published through either 'gold' or 'green' publication routes.

The core requirements of such a system are that it should be:

- centred around the output (proposed journal article, book chapter, etc.)
- driven by the available data
- based around small tasks rather than complex workflows
- able to capture data from other systems, whether internal or external to a particular institution
- designed to enable manual data entry quickly and accurately
- able to produce a wide variety of reports on demand
- configurable at an institutional level to support institution specific requirements

A data model is proposed for such a system, designed to enable the recording of sparsely populated data for any particular academic output.

It is proposed that integration with other systems be achieved through an adapter based approach (as is also found in SHARE¹, prompting opportunities for reuse).

Key aspects of functional design are outlined and illustrative wireframes provided for key screens in the user interface.

2. Purpose of this Document

The purpose of this document is to provide a specification for a system to help UK HE institutions manage administrative data in relation to the publication of open access Academic Outputs. The document is intended to:

- Describe the scope of such a system and the workflows it should support
- Describe an appropriate data model given the scope and workflows
- Provide illustrative wireframes for a user interface (UI) to such a system

¹ <https://sharewg.atlassian.net/wiki/pages/viewpage.action?pageId=8683527>

3. Requirements Gathering Methodology

This specification was written based on requirements gathered from the UK HE community in Q3 and Q4 2014, using the following methods:

- Reviewing in-house developed systems operating in this space at major UK research institutions
- A one day workshop with representatives from Imperial College London, University College London, the University of Edinburgh, and the University of Sheffield
- Consultation with the wider UK HE community through the Jisc Monitor face-to-face and online workshops
- Reviewing the work of Jisc Pathfinder projects, in particular the “End-to-End Open Access”² and the “OA Good Practice”³ pathfinders

4. Scope

The system specified in this document is intended to enable institutions to record and report on data relating to the publication of open access Academic Outputs generated by their staff. This includes Academic Outputs that are published through either ‘gold’ or ‘green’ publication routes.

The system is intended to sit in the space between a variety of existing systems that are already in place in many institutions including:

- Finance
- CRIS
- Institutional Repository

As far as possible the system is intended to avoid duplication of functions already present in such systems. The focus of the system is to record and report on information, some of which will be sourced from other systems such as the ones listed above.

The system is intended to help an institution:

- Support the collection of all necessary information related to the publication of open access Academic Outputs within the institution
- Ensure funding for article processing charges (APCs) is used appropriately
- Ensure appropriate decisions are made on Gold or Green open access publication for any particular Academic Output
- Enable reporting internally and externally on open access publication including reporting on the use of funds to support such publication

² <http://e2eoa.org>

³ <https://gw4openaccess.wordpress.com>

4.1 Summary of Core Requirements

4.1.1 Centred Around Academic Outputs

An Academic Output can be a research paper (intended to be) published in a journal but could equally be a book chapter, conference paper, a digital media artefact, or any other kind of output.

While the purpose of the system specified here is to record information relating to the publication process, it should be noted that the result of an attempt to be published is not always successful publication.

The system therefore puts the Academic Output (whether published or not) at the centre of the system.

4.1.2 Task and Data Driven

The process of publication, and the process of recording information about publication can be seen as two distinct, but related, workflows. In most institutions responsibility for publication lies with the researcher, while responsibility for the recording of information lies elsewhere (typically, but not exclusively within library/information services).

The two workflows clearly overlap, in particular where decisions are required about the available funds for the APCs associated with Gold open access publication.

While in an ideal world the relevant information would be recorded as the process of publication occurs, this does not always happen. In extreme cases the publication process may have completed before any information about the publication is recorded.

The implications for a system to record and report on data relating to the publication of open access Academic Outputs are:

- It must support the recording of data around the publication of an Academic Output in any order
- It must clearly report on required data in relation to an Academic Output that has not yet been recorded

In order to support this the system specified here works on the basis of 'tasks' to be completed which are generated from the data currently recorded in the system. As the data recorded is updated, tasks are updated or created.

For example, if recording the ORCID for the corresponding author for an Academic Output is a requirement, but has not yet been recorded on the system, a task would be generated "Add ORCID" in relation to that Academic Output. Once the ORCID has been added, this task would be marked as complete.

4.1.3 Ease of Data Capture and Data Entry

Data capture (obtaining data from external systems) and data entry (manually adding data directly to the system) are key tasks in relation to the system described in this document. For the system to work efficiently it must support data capture from a variety of sources and make it easy to enter data quickly and accurately.

The system will adopt an adapter based approach to enable data capture from other systems as described below in the section “Integration with Other Systems”.

To support quick and accurate data entry the system should in particular:

- Support auto-completion on all relevant fields within the UI
- Support keyboard shortcuts for navigating the application, including (but not limited to) the support of key strokes to move between fields on any data entry form in the UI (typically the tab key is used in applications for this purpose)

4.1.4 Flexible and Powerful Reporting

The system specified in this document must support reporting on the data recorded in a flexible way. This should include:

- ability to run reports based on parameters supplied at run time
- ability to add new reports easily
- export of report results in at least csv format

4.1.5 Locally Configurable

While UK HE institutions share a large number of requirements in terms of collecting and reporting on data related to open access publication (driven by RCUK and other funder mandates), there are small variations in institutional practice which means any system should offer the ability for institutions to configure which data needs to be collected and how tasks are created to prompt such data collection.

5. Workflow

The workflow diagram in Figure 1 captures the core processes, triggers and decision points in the processing of an Academic Output by administrators at a single institution.

A set of inputs shown on the far left of the diagram encapsulate the main activities which trigger the processing of an Academic Output. These are:

1. Enquiry from author or
2. Pre-payment request from publisher or
3. Invoice arrived from publisher

The points in the process at which the administrator finds out about an Academic Output may differ enormously, varying from an enquiry from an author who is considering writing an article, to receiving an invoice from a publisher for a published article.

As a result a system designed to record and manage Academic Outputs must support flexible workflows and not be overly prescriptive in the order in which tasks are carried out.

Communication with third parties is an integral part of the process. Rather than explicitly model communication between actors we have instead highlighted which processes have touchpoints, or interaction with other parties.

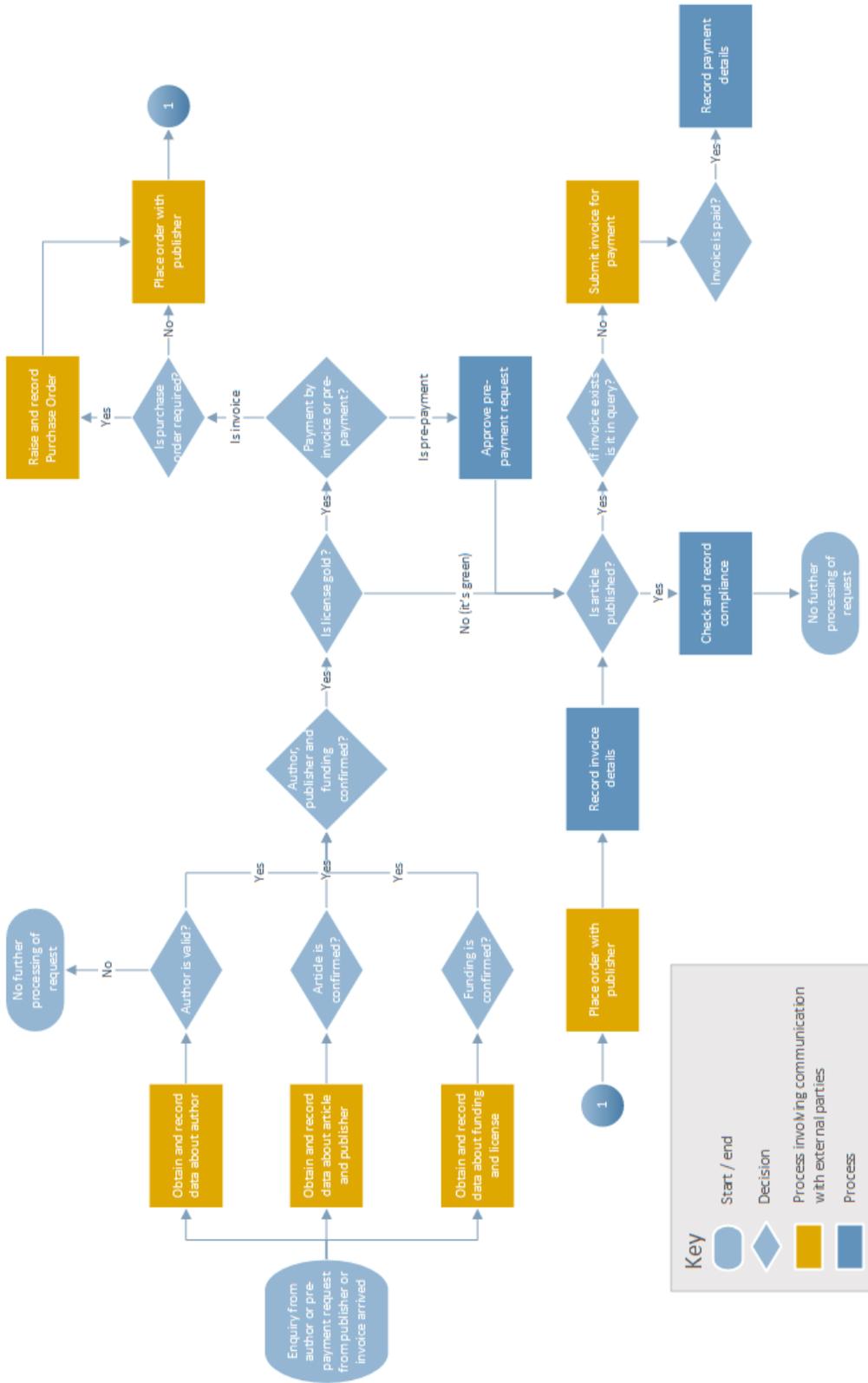


Figure 1, flowchart representing the processing of an Academic Output

6. Logical Data Model

The logical data model described in Figure 2 is designed to enable the recording of sparsely populated data for any particular academic output. This allows flexibility in the order in which data is recorded, and allows institutions to define their own 'minimum requirements' for data recording based on 'tasks' which can be configured at a local level.

6.1 Notes on the Data Model

It should be noted that systems users are not referenced in this model, but will be required with appropriate data structures for managing user roles and permissions. It should also be recognised that 'created by' and 'last edited by' stamps will be important.

An Academic Output can be a research paper (intended to be) published in a journal but could equally be a book chapter, conference paper, a digital media artefact, or any other kind of output. In addition to the key entities shown in Figure 2, there may be value on introducing some form of 'typing' to the Task entity, which could (for example) mark Tasks as valid only for specific types of Academic Output.

The Institution entity is likely to be related to a larger number of entities than shown in this logical model. Relationships to the Institution entity have only been included in the model where they specifically disambiguate the child entities. Other relationships to the Institution entity have been omitted for reasons of clarity.

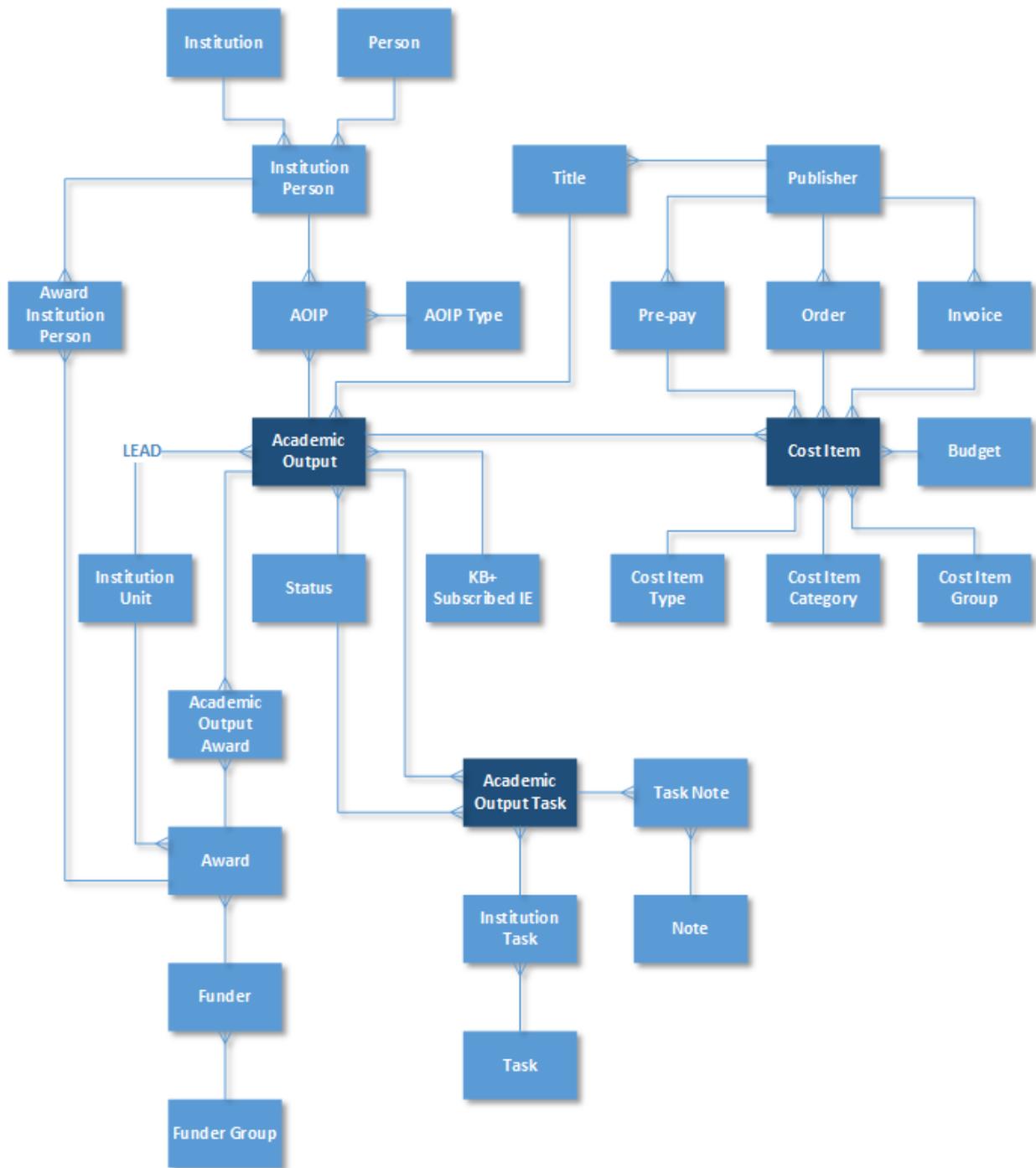


Figure 2, logical data model

7. Integration with Other Systems

As noted above, the ability to use data already recorded in other systems (whether internal or external to a particular institution) is a core requirement for the system specified in this document.

The Monitor suite of applications is typified by collecting, aggregating and enriching data from multiple sources, often aggregating over a different dimension to the norm, and then re-presenting that information to gain insight in a different context or dimension. For example, whilst most research output is classified primarily by subject area (at the most coarse level because journals group “like” materials together) Monitor is often much more interested in institutional and license based groupings and aggregations.

In order to fulfil the task of integration with other systems, Monitor must orchestrate a number of heterogeneous subsystems each with their own APIs and models of working. These APIs are supported by a number of substrates from open direct-call systems to publish/subscribe and polling interfaces.

A number of modern web based idioms have sprung up to provide RESTful interfaces to the traditional messaging components.

The question before Monitor is which standards to adopt for publishing and consuming events so as to achieve the widest possible impact. Some projects in the area have decided to “Bet the house” on PubSubHubbub. This seems like a risky strategy for Monitor, given the lack of non-repudiation and guaranteed delivery options. These protocols do offer message counters, so consumers (subscribers) can detect dropped messages, but ultimately, this means systems still need some pull capability.

Given any cursor based mechanism (for example the algorithm underlying OAI-PMH) a pull based system can be turned into a push based service by an agent which polls the pull service and pushes out events.

Given this, it seems appropriate to adopt an adapter based approach (as is also found in SHARE⁴, prompting opportunities for reuse) which will serve the project well for both collecting information from remote systems, for example financial systems, and for pushing messaging out using a variety of messaging substrates such as OAI-PMH, WebHooks, PubSubHubbub, etc.

⁴ <https://sharewg.atlassian.net/wiki/pages/viewpage.action?pageId=8683527>

8. Functional Design and Wireframes

Only key aspects of the functional design are described here. Illustrative wireframes are provided for key screens.

8.1 Overall Navigation and List views

The key entry points to the system for a user will be via Academic Outputs (the user wishes to create or edit a record), or via Tasks (the user wishes to complete a set of tasks). To support this users will be able to access lists of Academic Outputs and lists of Tasks which can be filtered by a set of on-screen controls.

8.1.1 Academic Output List View

The Academic Outputs page will provide a summary view of all outputs for an institution. From this screen the user will be able to locate existing outputs and also choose to create new outputs.

A set of filters and search can be used to retrieve individual outputs or groups of outputs which match specific criteria. A range of filters are suggested in Figure 3 which would allow the user to retrieve, for example, all published outputs with unpaid APCs, or all outputs for a specific publisher which cost over £2,000.

The screenshot displays the 'Academic Outputs' page. At the top left, the user 'Jo Smith' is logged in. The breadcrumb trail is 'Home > Academic Outputs'. A search bar contains 'Science' and a 'Go!' button. Below the search bar are filter controls for 'Filter by' with categories: Corresponding author (Author name), Publication (Publication name), Publisher (Publisher name), Cost (Min cost, Max cost), Status (Status), and Payment (All, Paid, Unpaid). A 'More search options' dropdown is also present. The main content area shows 'Showing 10 of 330 academic outputs. Searching for "Science"'. Below this is a table with the following data:

Output ID	Article title	Corres Author	Publication	Publisher	Cost (£)	Status	Payment
1256	Automatic yield-line analysis of slabs using analysis ...	Alison Raynor	Institute of Science	John Wiley & sons	1 440	Published	Unpaid
1278	Congue ipsum dolor sit amet eros viris ea qui pri ei ...	Dr D. Weymouth	Journal of Life Sciences	Reed Elsevier	1 880	Published	Paid
1386	Altera aeterno complectitur no est sea cu explicari ...	Dr A. G. Shan	International Journal ...	Springer Science	1 998	Unpublished	Unpaid
1389	Appellantur ad eos congue ignotas an exerci aliquip ...	J. Harvey-Smith	Fatigue and Fracture ...	Springer Science	2 314	Published	Unpaid
1412	Deterruisset ut incidereit adversarium deterruisset ...	M. K. Sarkar	Journal of Transactions ...	Springer Science	1 807	Unpublished	Unpaid
1434	Eum petentiums voluptatibus in idit nulla suscipit a ...	Dr L. Pye	Journal of Life Sciences ...	Reed Elsevier	1 034	Unpublished	Unpaid
1443	Cu quo vim eu brute posse nonumy ut incidereit ...	Jason Able	Geotechnical journal of ...	Springer Science	2 214	Published	Paid
1448	Cu qualisque incidereit per hist nostrud alienums ...	G. R. Susmal	Frattura ed Integrita ...	Springer Science	2 598	Published	Unpaid
1456	At utinam mediocrem mediocritatem has eos ques ...	Prof. Wes Archer	International Journal of ...	Springer Science	1 970	Published	Paid
1502	Quem feugiat detracto cu ne his solums nostrum ut ...	Lee Kang	Theoretical and applied ...	Reed Elsevier	1 300	Unpublished	Unpaid

At the bottom of the page, there is a pagination control showing 'Page' with a 'Go!' button and a set of numbered links: « previous, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, next ».

Figure 3, wireframe showing the Academic Outputs list view

Selecting an output ID or article title will open the individual output where he/she can review or edit the output details (see “Data and Task Display” below).

8.1.2 Tasks View

The tasks view will play a key role in supporting an administrator’s day to day activities relating to Academic Output processing. Its main purpose will be to present an overview of the status of tasks associated with outputs and a clear view of outputs needing attention.

Administrators will then have the tools needed to better manage and prioritise the processing of Academic Outputs at their respective institutions. For example, to focus on tasks which have been in progress for a long time or to reveal high priority tasks which are in progress or not started.

The wireframe in Figure 4 shows an extract of the task list for an institution and examples of the filters which an administrator could apply to retrieve specific tasks.

Filtering tasks by date would allow a user to prioritise tasks based on an expected completion date (being the date by which the task should be completed).

It would also be possible to filter the task list by type - for example all ‘verify author tasks’ or by broader areas such as ‘all tasks associated with author data’. This would allow a users to tackle tasks either based on a particular workflow (it may be more efficient, for example, to check author details from several different Academic Outputs at the same time if it requires login to a specific system).

The screenshot shows a 'Tasks' dashboard for user Jo Smith. The left sidebar contains navigation options: Dashboard, My-tasks, Academic Outputs, Add New Output, Articles, Invoices, Tasks, ADMINISTRATION, Reports, Manage Lookups, Manage Users, Manage Tasks, and More... The main content area is titled 'Tasks' and includes a search bar and a 'Filter by' section. The 'Filter by' section has dropdowns for Task (Task name), Output ID (Output ID), Article (Article title), Date task due for completion (From, To), Overdue, Priority (Priority), and Status (Status). Below the filters, it says 'Showing all tasks' and 'Results per page 10'. The task list table has columns: Task, Output ID, Article title, Task due date, Priority, Status, and Inits. The tasks listed are:

Task	Output ID	Article title	Task due date	Priority	Status	Inits
Verify corresponding author	1260	A mass matrix formulation for cohesive surface ...	02/09/14	High	Completed	JS
Confirm article publication	1260	A mass matrix formulation for cohesive surface...	12/12/14	Medium	Not started	
Confirm funding	1260	A mass matrix formulation for cohesive surface ...	08/10/14	High	In progress	JS
Payment made in full	1260	A mass matrix formulation for cohesive surface ...	31/12/14	Low	Not started	
Verify compliance	1260	A mass matrix formulation for cohesive surface ...		Medium	Not started	
Verify corresponding author	1261	On the multiaxial fatigue assessment of complex three-dimensional ...	09/09/14	High	Completed	JS
Confirm article publication	1261	On the multiaxial fatigue assessment of complex three-dimensional ...	19/12/14	Medium	Not started	
Confirm funding	1261	On the multiaxial fatigue assessment of complex three-dimensional ...	15/10/14	High	In progress	JS
Payment made in full	1261	On the multiaxial fatigue assessment of complex three-dimensional ...	07/01/15	Low	Not started	
Verify compliance	1261	On the multiaxial fatigue assessment of complex three-dimensional ...	07/01/15	Medium	Not started	

At the bottom, there is a pagination control showing 'Page' and 'Go!' buttons, and a set of numbered links from 1 to 10, with '9' highlighted.

Figure 4, wireframe showing all tasks on the system

Figure 5 shows an example of how the filters could be tailored to suit the preferences of a user or institution. Here the user has decided to follow up on tasks related to funding, specifically those Academic Outputs where funding for related APCs have not yet been confirmed.

The screenshot shows the same 'Tasks' dashboard but with filters applied. The 'Filter by' section has 'Confirm funding' selected in the Task dropdown, 'Output ID' in the Output ID dropdown, and 'In progress' and 'Not started' selected in the Status dropdown. Below the filters, it says 'Showing tasks: Confirm funding, Not started, In progress' and 'Results per page 10'. The task list table has columns: Task, Output ID, Article title, Task due date, Priority, Status, and Inits. The tasks listed are:

Task	Output ID	Article title	Task due date	Priority	Status	Inits
Confirm funding	1251	Arachnids submitted as suspected brown recluse spiders	02/09/14	High	In progress	JS
Confirm funding	1260	A mass matrix formulation for cohesive surface elements	08/10/14	High	In progress	JS
Confirm funding	1264	Functional analysis of tomato calmodulin gene family ...	08/10/14	High	Not started	
Confirm funding	1263	False traitors or worthy knights? Treason and rebellion ...	31/12/14	High	Not started	

Figure 5, wireframe showing tasks filtered by “confirm funding”, “not started” and “in progress”

8.2 Data and Task Display

The complete data associated with an Academic Output is too profuse to be presented as a whole in the UI and so will need to be separated into task related groups for entry and presentation. Based on our understanding of the workflow and data involved in the management of an Academic Output we suggest the following division:

- **Home** - data about the output such as its ID, creation date and approval status, the person submitting the request for an Academic Output (this is not necessarily an author), the administrator managing the output, and notes about preliminary communications.
- **Authors** - the Corresponding Author’s details and verification of the author’s relationship with the institution, and basic information about any additional authors.
- **Article** - information about the publisher, journal and article (in the case of other types of academic output, alternative labelling and data may be appropriate to describe the details of the published output).
- **Funding** - details of funding streams in relation to the publication of the academic output (e.g. COAF, RCUK).
- **Finance** - publication costs and payment details to handle both pre-payments and invoices.
- **Compliance** - a record of the article’s compliance to the requirements of publications paid for by the relevant funding streams, or other requirements such as local requirements from the institution.

Figure 6 shows a wireframe of the input interface for the “Authors” grouping.

Home > Academic Outputs > 2647

Academic Output: 2647

Home Authors Article Funding Finance Compliance

Settings Show tasks Comms trail

Dashboard My-tasks Academic Outputs Add New Output Articles Invoices Tasks ADMINISTRATION Reports Manage Lookups Manage Users Manage Tasks More...

Corresponding Author

Author ORCID	Author ID (internal)	Author name	Email address
0000-0011-5000-0007		Abdul Badani	abdul.badani@gla.ac.uk

Institution ID	Institution	Department / Faculty
GLA495	Glasgow Caledonian University	School of Psychology

Author verified? No Yes

Method of verification

Additional Authors

Author ORCID	Author ID (internal)	Author name	
0000-0011-4757-9292		Dr Lara Howard	Add row +

Institution ID	Institution
GLA495	Glasgow Caledonian University

Notes

Note	Staff name	Note date	
	Staff name		Add row +

Issue type

Academic output ID Go! « previous next »

Figure 6, wireframe showing the input screen for an Academic Output's author details

Tasks within the system are driven from individual data fields. Tasks can be created, or marked as completed, depending on the value of a particular field.

For example, if recording the ORCID for the corresponding author for an Academic Output is a requirement, but has not yet been recorded on the system, a task would be generated "Add ORCID" in relation to that Academic Output. Once the ORCID has been added, this task would be marked as complete.

Because of the relationship between a task and a data field, relevant tasks can be grouped based the part of the Academic Output they relate to, and displayed in that context. Figure 7 shows a task list displayed in the context of an Academic Task, grouped by the divisions listed above.

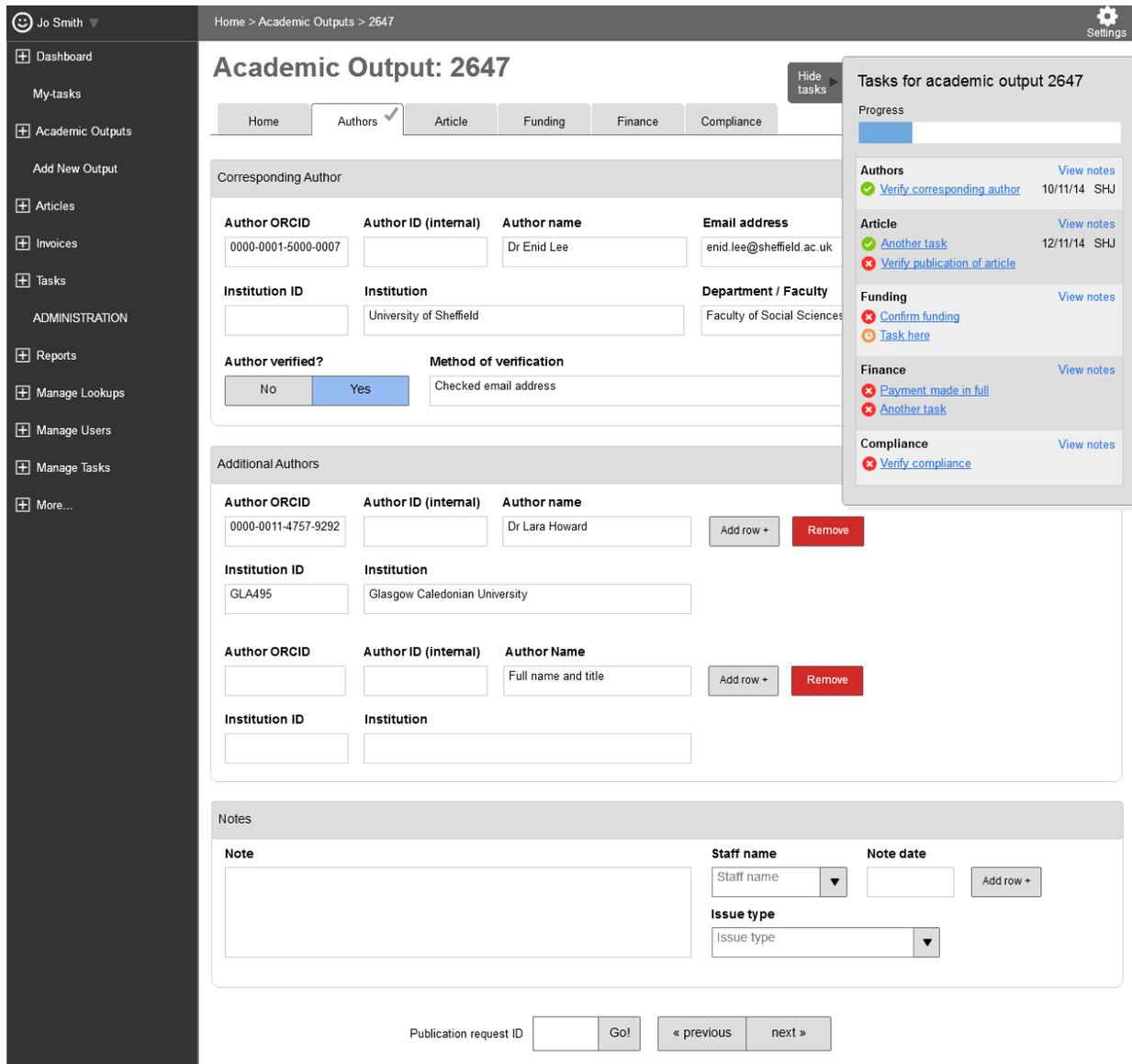


Figure 7, wireframe showing an Academic Output with the task list expanded

8.3 Role Management

User permissions on the system would be role based. It is likely that at least the following two roles would be required:

1. Standard user - able to work with Outputs and associated components such as invoices, payments and tasks. Will have create, read, update and delete (CRUD) access.

2. Administrator - is able to manage configuration of the system for the institution. For example defining rules for tasks, adding and managing system users, manage the system lookups.

A user's role settings will define the options available to him/her in the left primary menu (shown on all the above wireframes)..

8.4 Reporting

Some aspects of reporting may be served through the use of the 'List Views' described above. However, users will also need access to more flexible reporting as described in the Summary of Core Requirements above.

Appendix A - Data Elements

See <http://jiscmonitor.jiscinvolve.org/wp/files/2015/01/141205-APC-data-elements-v2.3.pdf>

Appendix B - Wireframes

High quality versions of the Wireframes illustrated in this document are available at http://jiscmonitor.jiscinvolve.org/wp/files/2015/01/Managing-Open-Access-Publication_-Wireframes-2.pdf