Research Data Spring

Idea to Project to Software & Tools

Contents

Background

## What is Research Data Spring?

Research Data Spring is our collaborative initiative for UK research. Its purpose is to gather fresh ideas and develop them into new solutions that would ease the pain points in research data management and improve the much needed software and standards, as well as prototype new shared services.

In the first phase we saw a high rate of participation. Approximately 600 researchers, librarians, publishers, developers and other third parties involved in the research data lifecycle posted 70 ideas and more the 150 comments on our **IdeaScale** community. We have selected 44 of these to join in at the sandpit workshop.

We hope that all participants will bring energy, knowledge and skills to make a full attempt at solving some of the key pain points in research data management. The first day will encourage exchange of knowledge, experience in the sector and feedback and we anticipate that 20 to 30 ideas will be pitching to the judging panel on the second day.

This is your day and your workshop and we more than welcome suggestions to the agenda and the structure. I will mention this a few times, but do contact **Daniela Duca** directly if you have something great in mind.

**Research data spring
Timeline**
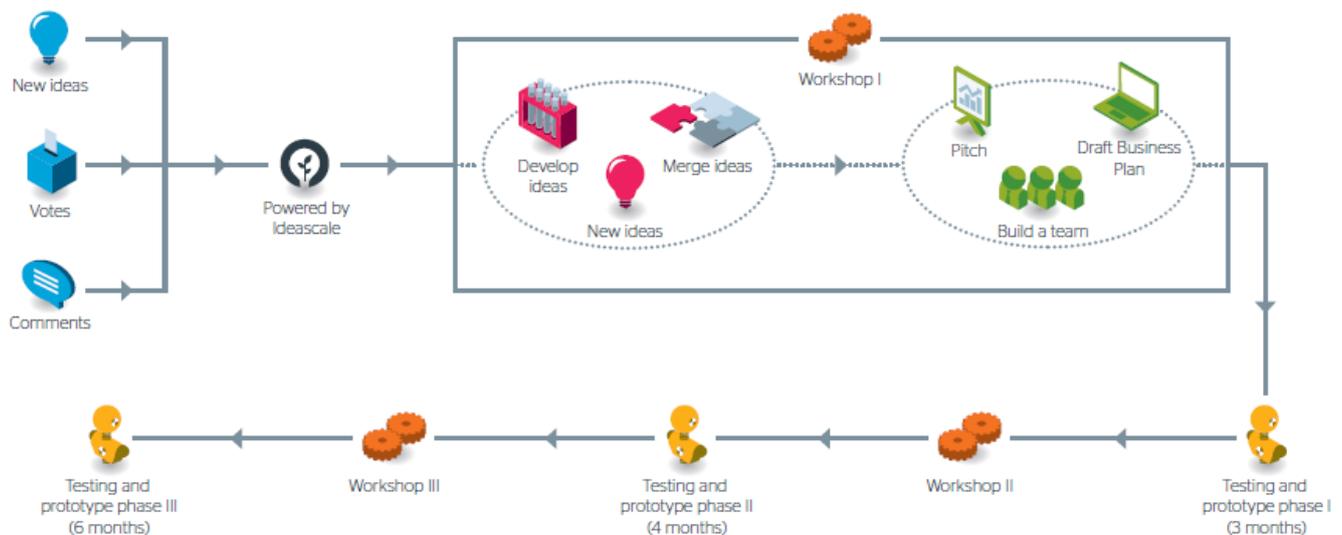
The sandpit workshop 26-27 February

What happens on Thursday?

The aim of the sandpit workshop is to open up the floor to collaboration, discussion, prioritisation and feedback. We want to ensure that, with your help, we can identify and support those ideas that will have the greatest positive impact on the research data management lifecycle.

Consequently, we have set out the sessions to allow as many people to have the opportunity to meet and reach out to their immediate groups as well as people and experts that they might not otherwise meet.

Most of Thursday morning and early afternoon sessions will be focused on group work around key themes. We have 5 birds of a feather groups, but these are by no means definite. Please do get in touch if you feel your idea fits better in a different group. This is flexible enough for you to choose a different group even on the day. During these group sessions we will encourage discussions around priorities and impact. Each author will have the opportunity to explain briefly their idea and what they are planning to bring forward.

We expect all ideas will develop. We anticipate that many of the discussions may lead to identifying commonalities or overlaps between ideas, points of collaboration or even new and improved solutions for key issues. Many ideas may merge into bigger projects or you may want to join another team. It is during these sessions that you would decide which ideas would go forward and the teams that you would like to join.

On Thursday afternoon you would be settling into teams and start working on the project plans. The set up would be similar to a hackathon, except we would be planning the projects and pitches. Wi-Fi will be available so you can get in touch with your colleagues at any point.

Additionally, there will be a group of experts representing different initiatives relating to research data and research systems that will roam around and may join your team discussions. Their names and affiliations will be circulated, so feel free to seek them out during this time as well.

During or just before dinner, we will have a signup sheet for the pitches and have tried to fit as many time slots as possible. There are 28 slots with 4 minutes per pitch and 4 minutes for questions.

You may continue to plan the projects, exchange business cards and ideas during dinner as well. Dinner will be quite informal with a great selection of food, buffet style, to allow for discussions to continue.

And Friday?

Friday morning, you will have about 3 hours to prepare the pitch. The pitching sessions will start at 11:15am. There are 4 sessions with 7 pitches of 4 minutes each. There are 4 additional minutes for questions. We will keep to the time allocated strictly to be fair to all teams. You may choose to use PowerPoint or free style the pitch. We will be providing a PowerPoint template with the key sections that you would need to include that you may use. More details in the pitching section of this document.

The judges will be taking notes and will be meeting the following week to decide on the funding.

Programme

Day 1 (Thursday, 26 February 2015)

Time	Activity	Notes
10:00	**Workshop opens** – tea/coffee and registrations	
10:30	**Welcome**, introduction to aims of workshop and structure	Plenary
11:10	**The 'mingle'** – meet and greet the idea authors, experts; developers, librarians, publishers, learned society members…	Active break
11:30	**Birds of a feather** – within smaller groups based around key topics and 8 ideas participants will identify and discuss the issues and priorities, and brainstorm solutions. Topics and ideas have been posted on our blog and the 5 groups are structured below.	This first part will be focused on discussing the key issues in this area.
12:25	**Birds of a feather active planning** – within idea teams, we encourage you to work on identifying the ideas going forward, the types of skills needed for the project and major questions that need to be further fleshed out.	The second part of this session will start with each author/team introducing their idea and follow into discussions around needs for the teams, collaborators and partners.
13:00	**Lunch**	
13:45	**Ideas market** – based on the active planning session, members of the teams can interact with other participants to collect information that would provide initial answers to key questions within their projects and find collaborators with required skills.	Active break to find more partners and flow into idea teams.
15:00	**Pitch and project planning marathon** – teams will reconvene in their new formations and will start working on the pitch and the project plan. Use **figshare** to save and share your work and add Research Data Spring and sandpit1 as tags.	Work on the project plan, think about activities within the next year, break it down for the first three months; identify collaborators and skills for the team. More details on planning in the next section of this document.
19:00	**Informal dinner**	Work and discussions can continue during dinner and drinks.

Day 2 (Friday, 27 February 2015)

Time	Activity	Notes
7:00	**Breakfast**	
8:30	**Pitch and project planning marathon** (continued). Use **figshare** to save and share your work and add Research Data Spring and sandpit1 as tags.	Work on preparing the pitch and finalising any gaps on the project plan within idea/project teams.
11:15	**Pitches session 1**	We anticipate up to 7 ideas per session with 4 minutes per pitch and 4 minutes for questions. We will prepare a few template slides and questions to be considered. More details in the pitching section of this document.
12:30	**Pitches session 2**	
13:30	**Lunch**	
14:15	**Pitches session 3**	
15:30	**Pitches session 4**	
16:30	**End workshop**	

What to bring?

This is a list of suggested items that would help you during the workshop:

» Laptop/tablet or any device that you feel comfortable using for note-taking, tweeting, skyping, searching, planning and getting connected with new people
» Business cards
» Open mind and positive attitude!
» Feel free to upload all your work, slides, and scribbles there. Please use these tags: Research Data Spring, Sandpit1, in addition to any ones you feel relevant.

What else to consider?

In our first email we encouraged you to:

» Review the comments on your idea and do get in touch with those that have contributed expert knowledge
» Think about your first 3-month goals and the longer-term goals for your idea
» Are there other similar and/or related initiatives in your area? How does your solution align with these?
» If your solution is discipline specific, have you thought about transferability to other groups or areas?
» Read over the Jisc **standard grant T&C**; note that all Jisc grant funded projects must make all outputs openly accessible

Birds of a Feather – the idea groups

Groups	Topics	Ideas
Group 1	Analytics & Visualisation	3D data workflow
		Data browsing tools for repositories
		Seeing is believing - visualising research data
		RDM Administration Analytics
		Enabling complex analysis of large scale digital collections
	Preservation	Develop a DataVault
		Preserving research data through deposit
	Researcher experience	Cloud Work Bench
		Collaboration tool for qualitative data analysis
Group 2	Provenance	GeosMeta: a metadata and provenance service to support research
		Provenance and Packaging
		Linked data notebook
		Standards and Schemas for Digital Research Notebooks
		Use semantic desktop to capture contextual research data
	Researcher experience	An Intelligent Data Cleaning Software Tool for Research Data
		Badges as a proxy for peer review of data
		Giving researchers credit for their data
Group 3	Researcher experience	Clipper: Enhancing time-based media for research
		Using Alfresco for Management of Research Data
	Reuse	Computational Experiments as Data Objects
		Recomputation: enabling archiving of computational experiments
		Research Software: assigning persistent identifiers
		Research Data as a Unique and Distinctive Collection (UDC)

Groups	Topics	Ideas
		Unlocking the UK's thesis data through persistent identifiers
	Sensitive data	Dissecting digital humanities data with biomedical tools
		Managing Sensitive data
Group 4	Services	Open Source Database-as-a-Service with Data Publishing
		Pushing Database Wiki out of the nest
	Standardising discourses	Architecture for end-to-end data archiving
		DAF Question Bank
		Research Data requirements vocabulary
		Extending the Organisational Profile Document (OPD) to cover RDM
		Sound Matters: a framework for use and re-use of sound
		Standard protocol for research equipment
Group 5	System integrations	Distributed Knowledge Sharing Platform for RDM
		Integrated RDM toolkit/service
		Link EPrints to remote research data
		Protocol for a shared approach to Research Data Management Systems
		Research Data Infrastructure for the Visual Arts (RDIVA)
		Seemless RDM, Seemless Integration
		Template Solution for Integrated Repository and RMS
		Streamlining Deposit: An OJS to Repository Plugin
		Wrapping up RDM functionality

What happens after the sandpit?

We are hoping to send out the funding letters to the shortlisted projects by in the first or second week in March. We will also post the results up on our blog.

The teams should already be able to start working on their deliverables. We will expect the teams to return to the second workshop in June 2015, where they would be presenting their progress and suggesting the next directions of development. If shortlisted then, the teams would get funding for another 4 months of development and come back to the last workshop in November. This will be followed by 6 months of development.

Just to flag here, that all plans and pitches presented at the first workshop should take into account the end goals and the bigger period of development. The final outcomes and deliverables should be presented during the pitch, while focusing on the milestones and tasks required for the work in the first 3 months.

During the development phases, the funded projects will track and share their progress on our research data management blog. These would be short snippets on latest team work, but more details on this will follow the workshop.

Planning

This section lays out details and suggestions for preparing the project plan. The afternoon sessions on Thursday will be focused on developing an initial rough draft of the project plan. You and your team will submit these drafts to Jisc at the end of the workshop and, if selected for the 3 months funding, you will work based on these plans. The project plan will constitute the base of the pitch and will be assessed via the pitch by the panel of judges.

The project plan is a dynamic and living document. And even though the funded projects would submit a more finalised version within March 2015, by no means would this be fixed. We recommend to use Google or a similar platform to work on the draft project plan during the workshop. Before the pitch, you can upload the rough draft on figshare using Research Data Spring and Sandpit1 as tags. You can link these to your ORCID as well and generate some DOIs. More details on how we plan to use figshare will be provided on the day. Also, Mark Hahnel will be there to answer any questions.

The funded teams will then present their progress during the second workshop based on the devised deliverables, tasks and milestones that are included in this plan. The team can propose any changes or alternate directions in achieving deliverables during the second workshop. If these are approved by the panel, the team would receive funding for 4 more months of development. Returning to the third workshop, the teams will present progress, adding more detailed milestones for the last 6 months of development. Again, these would go through a panel that will make suggestions and approve the projects for further funding.

We expect the shortlisted teams to deliver tangible and less tangible outcomes after each development phase.

Preparing the project plan

You have probably heard dozens of times if not more, that the key to a successful project is in the planning. That is why, the first thing we would encourage you to do once you settle in your teams at the sandpit is to create a project plan.

The plan does not need to be in its final form, but you have to agree with your team on the key elements. We recommend using an online drafting platform to make it easier to collaborate on the plan and then upload the draft on figshare.

Thursday afternoon has been reserved for working on the project plan. Wi-Fi will be available, so we encourage you to contact any of your colleagues that were not able to attend, whether by skype, email or telephone.

In preparing a rough draft of the plan, we devised 4 key steps:

1. Define goals and objectives
2. List and describe anticipated deliverables
3. Create the schedule and milestones
4. Estimate resources involved and funding required
5. Ultimately, having all this information, you should be able to write a clear and concise summary

Step 1: Goals and Objectives

This might be the step that you take last, but some of the goals and objectives have probably already been set even on IdeaScale.

Before even thinking about the goals and objectives, remember that a project is successful when the needs of its stakeholders are met. Hence, identifying the stakeholders would be key for this step. These could be champions supporting your project, users of your solution, trainers, data managers, data scientists, perhaps repositories that you are working with etc. Below we included a research data lifecycle to help you identify where these stakeholders sit and how you can get them involved. You can flesh these out further within your comms plan.

Each project stakeholder that you list would then have different needs and would also involve different risks.

Here you would brainstorm and then try to prioritise these needs, hence setting the goals and objectives.

These objectives should be SMART, some guidance is available here:
http://www.jisc.ac.uk/fundingopportunities/projectmanagement/planning/objectives.aspx
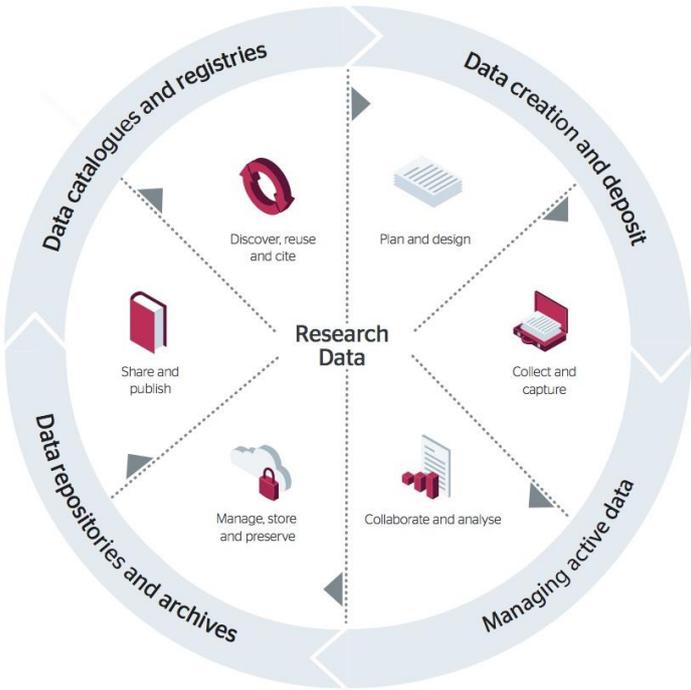
The research data lifecycle

Step 2: Deliverables

Now that you have your goals and objectives, you can write out a list of things that you and your team could deliver.

This would include tangible deliverables (reports, code, standards, software enhancements etc.) and less tangible knowledge and experience (use cases, interview outcomes, recommended solutions, technical advice etc.) that you hope to build and share. We expect that all projects will define and describe the use cases related to the idea/solution proposed.

You should have a few categories of deliverables based on Research Data Spring's timeline:

» General/long-term project deliverables
» Deliverables for the first 3 months
» Deliverables for next 4 and 6 months

Next to each deliverable you can set indicators of success.

You should also include an early **risk analysis** and the **mitigating actions** that you would be taking for the relevant deliverables.

You may find it easier to set out a table for this with the type of outcome and a brief description as well as an anticipated date, indicators of success and any risks involved.

Step 3: Schedule and Milestones

Once you have the aim, goals, objectives and deliverables, it is time to brainstorm a set of tasks that will be required to carry out the objectives and produce the deliverables. Think about the amount of effort involved and who will carry out the task, i.e. the capacity available. You should then be able to work out a rough delivery date and set a few milestones.

You should have an overall plan for the whole duration of the project with an estimated timeline, however be sure to focus some effort in identifying the tasks and milestones for the first three months in a bit more detail. These would be the key elements that would be funded in the first instance.

Step 4: Resources

In the previous steps, you should have been able to provide rough estimates of the effort involved in carrying out the project. It is time to put that all together and evaluate the financial, technical and human resources required.

Financial

List as many of the costs that you think would be involved based on the deliverables; rough estimates are fine at this stage. Please note that all travel and subsistence expenses should be included in the cost estimations (i.e. please include travels for next workshop, and any other travels and accommodation you may require for the duration of the project)

Technical

List how the project will follow best practice for technical development, any specific technologies or approaches that will be adopted, rough estimates for data storage, computing power and other technical requirements

Human

List all required skills and roles required to carry out the project. Identify by name, the individuals and organisations that are collaborating or have a leading role in the project. They must commit time and have the capacity to work on your project.

Communications

List all stakeholders who need to be in the loop about the project and how they will receive this information.

Microsoft Research is supporting the Jisc Research Data Spring project. Should successful project teams require cloud resources, Microsoft will provide resources on **Microsoft Azure** to enable rapid prototyping, development, and deployment in the cloud. This includes compute, storage, databases and web hosting, as well as advanced services such as machine learning and big data analytics. They hope to help move efforts forwards, providing maximum agility and flexibility to research teams. This is part of our Azure for Research programme - more details at **http://azure4research.com/** or contact **Dr Kenji Takeda**. It is a choice the successful project teams can make, but we're really pleased Microsoft have offered this as part of their Research Programme.

Step 5: Writing the summary

Once you have discussed and brainstormed steps 1-4, the summary should be a piece of cake. The purpose of the summary is to provide a clear and concise description of your project so that it actually invites the reader to find out more about what you are hoping to achieve. As a disclaimer, we may post the summary on our website. But you do not need to worry about it too much or try to make it perfect at the workshop, focus on the other sections as they are more critical in terms of attaining funding. Although, a good executive summary can be the base of your pitch. Some might call it an elevator pitch.

Additional elements to consider

Depending on where you think these fit best, you may also think about the implications that the following elements would have on your project:

» Strategy and methodology of the overall approach
» Related projects
» Critical local, national, global initiatives and standards used
» Constraints to the project
» Intellectual Property Rights (all Jisc funded projects must share outcomes with the community) and what will happen to the project outputs/outcomes at the end of the project
» Sustainability plans

Pitching

Building the pitch

We have allocated 4 minutes for pitching and an additional 4 minutes for Q&A. We will have 4 sessions with 7 pitches each.

There are a few things to consider when preparing the pitch. In the business world, having a crystal clear value proposition is critical, because that will let people know very briefly why they would invest in your idea. For Research Data Spring, the same would hold. But what is the value proposition? That is not necessarily just the money that you can make. Essentially, if you can briefly summarize:

> The problem your idea is solving
> The quantified benefits that it brings
> The reasons why researchers or stakeholders in the research data management lifecycle would use your solution

then that will bring out your value proposition and you will definitely gain the attention of the judging panel and of any users that would want to join you in your quest.

Provided you have 4 minutes to impress the judges, your language is critical. Try to be concise, feel free to use metaphors if that will get your point across easier, and be very specific, though not too technical.

Running the risk of repeating myself, I want to emphasize that you have to set an achievable target within your project. A long-term target containing the outcomes and deliverables of the project as a whole, and shorter-term targets that would cover firstly the 3-month development phase, and secondly the remaining two development phases. We expect some projects would run for less than the entire duration of Research Data Spring and do not exclude that some may be longer. If your believe your project will take longer than the three development phases, have some ideas for sustainability in the back of your brain, just in case the judges will be curious about your plans.

In business, entrepreneurs are asked to have a valuation of their start-up. As we all know, this is not a real science. Some businesses are over- and some are under-valued, but they all have some rough estimate of their costs and benefits. Having those briefly outlined in your pitch would be an added value (no pun intended). I have mentioned about some of the costs that would need to be considered in the project requirements section, please do refer back to that. In terms of benefits, these can be tangible and less tangible and most likely expressed via use cases and the number of groups or institutions that would benefit from your solution. Be precise and to the point!

Finally, your plan (that is your deliverables, milestones, indicators of success) would form the basis of your pitch. You would have spent an afternoon developing it and discussing with your team and you would want to show some of the hard work that you have done. Of course, there will not be enough time to go over all the details, so you can submit the rough draft separately (or add some extra slides to have them ready if there will be any questions).

That brings me to the logistics – we will be providing a template PowerPoint deck with some headings and pointers. By no means is this a fixed, please feel free to adjust and update as you desire – if you decide to use a PowerPoint presentation (or any online presentation modules). The key reason for having a template is to make

sure it will be compatible with the computer that will be projecting it. Please upload these to **figshare** before the pitch and tag them as Research Data Spring, Sandpit1.

In short, your pitch would need to have the following elements:

» Title of your idea and aim

» Team – who is it that will work on the idea primarily? Who is ready to pilot the solution?

» Scope – how does your idea fit within the scope of Research Data Spring?

» What gap is being filled and why this will work (vs existing solutions if any)?

» Estimated impact, benefits and/or return on investment

» How is your solution sustainable?

» Outputs and milestones – what are your final long-term outputs and timeline, how and what are you planning to achieve during the first 3 months and the subsequent development phases?

» What are your indicators of success?

» Funding

Judging panel

We have created a panel of 6 experts that will be hearing the pitches, following up with questions and shortlisting the ideas. The criteria have been created as a direct reflection of the elements that we are asking to be included in the pitch.

We will be keeping strictly to the time to allow all teams to pitch before exhausting the judges!