1.1.1 Generic Features

There are several desired features for the high-level characteristics of a DRN.

- **API Access:** A DRN should be accessible via an API (Application Programming Interface).
- **Automation:** Users want a certain level of automation, with files saving and syncing automatically, automatic updates, automatic extraction of metadata etc.
- **General Features:** Users want a DRN to be simple, scalable, transparent and enable them to access all their work in one place.
- **GUI:** Users want a graphical user interface to work with so they don't have to understand the inner workings of a DRN to use it.
- **Localisation:** Users want to be able to use a DRN in their own language working on their own timezones.
- **Remote Access:** Users want to be able to access a DRN remotely from other devices.
- **Synchronisation:** Users want to be able to work synchronously and edit / work on multiple documents / datasets at once.

1.1.2 Notebooking Features

There are a number of desired features that either exist as part of generic notebooking software or would naturally extend onto notebooking software.

- **Content Support:** Scientists want to be able to record and write up their work using a wide range of different content types (e.g., free text, numbers, photos, diagrams, graphs etc – the full list can be found in Appendix **Error! Reference source not found.Error! Reference source not found.**). A majority of these will be part and parcel of using generic notebooking software such as OneNote.
- **Interaction and Access:** In addition to supporting interfacing with paper, scientists want to be able to interact with a DRN in multiple ways including handwritten notes, using digital pens, via voice etc. A DRN should also be accessible via a range of devices including mobile phones and tablets, and be available offline.
- **Link to other files:** Some files might exist that are too large to incorporate in a document or set of notes e.g., large datasets or code scripts. In this instance users need to be able to link to these external files.
- **Organisation / Reconfiguration:** Users want to be able to organise and reconfigure their work, so it can be re-ordered to be viewed by person, experiment, date etc.
- **Paper Integration:** A big reason why scientists still don't digitise all of their work is due to the affordances of paper. Paper still remains easier to scribble notes on, and for many it remains the natural easy way of making notes. Even through a global pandemic this hasn't changed, scientists still make notes on paper and the ones they deem important enough will make it into an electronic format. We need to ascertain the current methods of interfacing with paper (e.g., hybrid devices, scanning paper notes) and ensure that these are available for users. Innovation on extracting information from unstructured data e.g., paper records, images and annotated charts is currently being undertaken by researchers, including members of the Frey group at the University of Southampton working on the Data Revival project[1]. This includes creating chemistry-based contextualisation around extracted chemical names, formulas and chemical structures.

[1] https://labs.futureworlds.com/data-revival/

- **Referencing & Literature:** Scientists want access to literature via their DRN. In some senses these features are covered by referencing software as they require the capability to link to literature and cite literature sources. However, there is also a deeper requirement to be able to extract data from legacy literature, and import/export literature references directly from a DRN. Users also want to be able to cite data and code in their work.
- **Word Processing:** One of the reasons that scientists continue to use software such as Microsoft Word and OneNote to record their notes is due to their superior word processing capabilities. Our ELN landscape assessment concluded the most widely supported file formats for including in the ELN are Microsoft Excel and Word, and often the ELN is able to process these files to convert them to the ELN format or enable the user to edit the content within the ELN, A DRN platform should interface with existing notebooking software such as these programs in a similar fashion.

1.1.3 Data Features

Data is at the cornerstone of scientific research, and as the amount of data generated (predominantly electronically but sometimes not) increases, researchers have ever growing needs for different digital tools to manage their data.

- **API Access:** A DRN should interface with APIs so users can extract the data they need.
- **Data Access:** There needs to be the concept of different access control levels with different user permissions. This is common with a lot of the software that would be integrated to create a DRN, but needs to be taken into account for any new features or programs that don't currently facilitate this ability.
- **Data Conversion:** This is a key feature required by scientists. Given the plethora of different data formats available for representing data in the physical sciences, and given how many pieces of software use different formats and potentially only facilitate import/export in certain formats, the capability to convert between formats is vital. There is no universal format converter for the physical sciences, so existing tools would need to be considered and a new one developed to enable this feature.
- **Data Exchange:** A DRN should facilitate data exchange, this may be through a data exchange platform.
- **Data Integration:** Scientists often need to integrate datasets; they need tools to enable the collation and integration of data.
- **Data Management:** Poor data management is one of the major issues with the current state of digitisation of physical sciences research. Scientists need a platform to record, manage and curate their data.
- **Data Quality:** Users need to be able to quality check their data, and there needs to be the notion of categorising data as "trusted".
- **Data Retention:** Data needs to be preserved for the required period of time, and there also needs to be a mechanism to remove obsolete data.
- **Data Security:** Any DRN would need to be offer secure storage, backup and archiving features.
- **Data Standards:** Standards are a key issue with respect to digitising scientific research. Ideally a DRN would be configured to work with required domain data standards, and would have features to ensure data integrity and consistent use of standards. This realistically means that we still have a requirement for standardised file formats that work across different domains and pieces of software.

- **Data Support:** A DRN should support data from other resources, and be able to deal with all different stages of data.
- **FAIR Data:** A DRN should support curating data to ensure it is FAIR, and publishing FAIR data.
- **Identifiers:** The use of identifiers is very important, not just for data but for all aspects of scientific research. A DRN would need to be able to link data together using standard identifiers (e.g., chemical identifiers) in addition to working with unique IDs for samples, experiments, researchers etc.
- **Provenance:** A DRN should contain or link to tools to capture provenance. Users want to be able to capture the provenance of their data, code and structures. Users also want to be able to see the algorithms/code that were used to create data or perform data analysis. It is also important to users to be able to track what happens to their work, it would be useful to be able to see what datasets and work get re-used and how.

1.1.4 Publishing & Sharing Features

One of the desired end products of a scientific research project is creating publications. Users want a DRN that will aid with this process:

- **Documentation & Instructions:** One of the key aspects of recording your work (in any fashion) is to document your work and provide instructions to others as to how to implement it or use it. A DRN needs to facilitate this.
- **DOIs:** We need better mechanisms to create and supply DOIs / permanent links for data.
- **Export:** Users want to be able to export all aspects of their work (including the entire notebook) in multiple formats.
- **Licensing:** There is a requirement to be able to create license for data, code and research prior to sharing it.
- **Open Access:** There is a distinct call for an increase on Open Access in the scientific community. Both in terms of work that would be shared via a DRN but also work that a DRN can access e.g., scientific articles and data.
- **Publishing:** One of the concerns of using ELNs has always been that there would be an unnecessary duplication of effort. Users are reluctant to spend a long-time curating notes and data when they have to then redo half the work to produce the final publication. There has long since been a call for a “generate report” or “generate thesis” button that would enable data and notes to be pulled together in at least a ‘semi-publication-ready' state.
- **Sharing:** One of the key reasons for using a DRN would be to collaborate with others. Users want to be able to share work and data at different stages with their collaborators, and share in a range of ways be that singular files, whole notebooks, datasets etc.
- **Social Media:** Some users want to be able to share updates about their projects via social media. There should be an easy method to do this from a DRN.
- **Researcher Attribution:** There is a call to be able to better attribute data, code and components of work to different researchers when creating publications such that everyone who worked on a project gets properly acknowledged as opposed to just the main authors.
- **Repositories:** Users want a direct link between their DRN and a repository such that they can upload and download data/code/files seamlessly.

1.1.5 Collaboration & Management Features

The global pandemic necessitated an increase in the use of management-esque software. There are a number of features of collaboration and project/team management that would be required of a DRN.

- **Auditing:** One of the features that users have cited that they need to be able to replicate in any digital system is the audit trail that is vital both from a health and safety perspective but also when it comes to claiming patents. Which arguably should be easier to achieve using digital tools, but it needs to be done in a user-friendly way. There still needs to be moderation levels whereby managers can review rate approve and reject plans, COSSH forms, processes etc. These processes will work to protect IP rights.
- **Comments:** Users want the capacity to comment on all aspects of work (data, code, notes) to collaborate with their colleagues on every level.
- **Notifications:** Users want to be notified for a range of different activities, e.g. comments, messages, updates to work, updates to tasks etc.
- **Subscribe:** Users want to be able to subscribe to each other's work, e.g., notebooks, projects etc.
- **Team Management:** Team management is vital to any group project. There are lots of project management programs out there (Teams, Trello, Notion etc) and these should be incorporated into the overall DRN such that large scale projects can be managed, tasks can be allocated and work can be organised appropriately.

1.1.6 Domain Based Features

The physical science domain contains a wide range of different disciplines, which within themselves (e.g., chemistry) have a large amount of subdomains. This means that there are a great deal of domain specific formats, standards, databases, software packages etc. A DRN would need to be able to interface with all of these.

- **Chemical / Molecules:** Many of the researchers in the physical sciences community work with chemicals and molecules, and a system that has an inbuilt understanding of molecules and chemicals is one of the key types of domain software that users flock to. There should be standard lists of reagents and reaction schemes, and it should be possible to import reactions into a DRN. Additionally, many users have cited that a system that could automatically recognise chemicals in their notes would be incredibly useful.
- **Default Lists:** Another desirable feature that users have noted is the inclusion of default lists of relevant domain-based items, such as a standard list of instruments and reagents, default values for fields, a global database of chemical values etc. There should be a way to interface with standard taxonomies and vocabularies as well.
- **Equipment Interface:** Users have mentioned a desire for more modern equipment. Obviously PSDI can't control what equipment scientists have to hand, however this does mean that any DRN should be equipped to interface with modern equipment as well as legacy equipment. As noted above there should be standard lists of instruments, including mechanisms to interface with the different data types that these instruments produce.
- **Experiment Planning / Recording:** One of the main things that scientists use their paper-based lab notebooks for is to plan and record their experiments. Therefore, it is unsurprising that they would expect this functionality to be replicated in a DRN. Users want to be able to add all the usual information to their experiment (plan, title, date,

conditions) and have requested more advanced features that would improve upon using a paper lab notebook. Users want to be able to copy and modify plans for future use, demonstrate the status of their experiments and mark them as completed when they are finished.

- **Health & Safety:** A DRN should enable users to link to and incorporate health and safety data into their work. There should be an index of all of the COSHH materials and users should be able to link to COSHH forms in their experiments. Experiment records should be automatically populated with risk assessment information based on the chemicals that are being used, and these should be automatically flagged upon entry. Users have also noted that additions such as a risk phrase vocabulary would be incredibly useful.
- **LIMS/ELN:** Just because ELNs aren't overly popular in academia, doesn't mean that they aren't being used in industry, or indeed that some academics aren't using them. A DRN should be able to interface with an ELN if necessary, particularly if a user is collaborating with someone else using one. Additionally, there needs to be links with LIMS software as this is much more widely used across the physical sciences community. This would require a level of collaboration with these software vendors to identify common data standards for interfacing.
- **Link to domain-based databases:** There are many databases out there that are well used by the physical sciences community, therefore any DRN would need to interface with these.
- **Link to domain-based software:** There is a wide range of domain-based software available (as demonstrated in Section **Error! Reference source not found.**). It would be entirely unrealistic to re-build all of these software packages within one DRN platform, therefore a DRN would need to interface with these such that scientists working across different domains could still use their required software alongside their notes and data.

1.1.7 Coding Support

As the amount of data generated by the physical sciences community continues to increase, there has been a vast surge in using computational methods to handle and analyse this data.

- **Coding:** Support for coding, and recording the process of developing code within or linked to a DRN is essential. Software such as Jupyter notebooks have become incredibly popular as data and code can be stored side by side and users can collaborate on coding projects, and any viable DRN would need to interlink with Jupyter notebooks and other software that offers similar functionality.
- **Versioning:** Hand in hand with coding comes the notion of version control. In order to record the process of developing code, different versions of the code at different points needs to be stored and there needs to be a capacity to roll back to different versions. Ideally a DRN would interface with GitHub and other popular version control services.

1.1.8 Metadata, Semantics & AI

As the use of computational techniques to handle data has increased, there has been a growing desire to capture and store metadata, and to use intelligent technologies to improve data handling.

- **AI Tools / Integration:** Users have requested integration with ML packages to handle and analyse data, and the capacity to store ML training data.
- **Metadata:** There is a high requirement for capturing and supporting metadata about documents, datasets and scientific content. Digital tools can automatically capture vast

amounts of valuable metadata and the provenance trail, and where automatic capture is possible, they can prompt the user to add metadata as they go along, rather than forcing them into a burdensome process of metadata generation and curation at the end of the project when they are ready to move on.

- **Semantics:** There is a demand for the inclusion of semantics, for interoperability and to add context and meaning to the data captured in a DRN. Users want their data to be marked up/annotated semantically where appropriate with links to established ontologies and vocabularies. This will enable experiments, data, code etc to be classified, and will facilitate semantic search on concepts. This also enables projects and notebooks to be linked together, and for inferences to be made about similar work. Users have also requested inferences for the same molecules of reactions.

1.1.9 Searching

One of the key things that a computer can do immeasurably better than a human being is search efficiently through large quantities of data. Many of the features requested by users are related to search.

- **Domain Search:** Users also want to be able to do domain specific searches, such as searching by structure, reaction scheme, chemical etc.
- **Characteristics Search:** Users want to be able to search for datasets with specific characteristics.
- **Indexing:** Notebooks, data, notes, code all needs to be indexed such that it can be searched.
- **Keyword/Concept Search via Content Types:** Users want to be able to search through their data by keywords, and to search intelligently on concepts rather than just plain text (e.g., semantic search).
- **Literature Search:** Users want to be able to search through their work for inclusion of certain literature, access literature searches in bibliographic software, and search for similar types of literature.
- **Notebook Search:** Users want to be able to search through entire notebooks or collections of notebooks, including previous project notebooks.
- **Similarity Search:** Users want to be able to find similar content to their work, including similar data or experiments.
- **Visual Search:** Users want to be able to search by visual content such as images, they also want to be able to search for specific pieces of information in graphs.

1.1.10 Customisation & Extension

Notetaking is very personal, and yet within that personalisation lies different individuals' standard methods of doing work. Users need to be able to personalise their work but also create templates for their own methods.

- **Personalisable:** DRNs should provide a system where users can tailor them to their own specific research and domain. Linked to the notion of interfacing with domain-based software, users should be able to add plugins of their choice and they should be able to create their own plugins.
- **Templates:** Users want to be able to create their own templates for models, code, experiment plans, notes etc. They also want to capacity to share these templates and use those created by others.

1.1.11 Training & User Support

In addition to providing the features that users require, it is equally important to provide them with support to enable them to use a DRN in the most effective manner.

- **Training:** As with any new software system, training should be provided to users to demonstrate the different types of functionalities a DRN has to offer, and to show them the most effective ways of using one in their specific research.
- **User Documentation:** A DRN should come equipped with a decent level of user documentation to aid users in using the system.