Data Management Hub

Data is the basis of empirical research. Authentic, findable, accessible, interoperable and reusable data is paramount to enable the replication of scientific research and to empower open science.

We know first hand the challenges faced by the research community with scientific data management. Of researchers in Ireland, 80% will not be able to provide the data 3 years after completing the research [1] of the scholarly papers published, 30% are never opened by anyone (never mind being referenced) [1]. In 2015, for example, over 200 researchers repeated 100 published studies to see if the results of these could be reproduced (only 36% could) [2]. *This leads to bad science in limbo.*

At present there is no complete data management solution on the market for the management and preservation of research data. Available solutions are not designed to manage scientific data (Github, Dropbox); they are generally centralised or they focus on small data (Open Science Framework). Raw data is still saved on CDs, DVDs, USB sticks, portable drives and then forgotten somewhere. Institutional repositories only cover the 'last mile' of data publishing and are often seen by researchers as an additional burden because they do not fit into the workflow. Researchers struggle with data management.

It doesn't need to be this way. *Today we have technologies that allow us to build reliably distributed secure data infrastructures which can preserve authentic research data and empower research replication for future generations of students.*

We asked scientists what they needed for effective data management: Open Source, no vendor lock-in, easy to use, part of the data workflow, easy data sharing, with commenting (peer review), with versioning, authentic, secure, neutral and distributed (no single point of power over the system), searchable and easy to fit into the *Open Access* landscape.

Data Management Hub (DaMaHub https://damahub.com/) is a distributed platform for scientific data workflow that allows scientists to organise and share their data and their results in an easy to use, secure and reputation-building way.

DaMaHub extends the use case of existing open-access repositories to data and to the whole research science lifecycle. It will make scientific data and articles FAIR - Findable, Accessible, Interoperable and Reusable.

- *DaMaHub* is a solution that preserves the authenticity and accessibility of scientific data by employing openly licenced technologies, namely IPFS [3] and Blockchain [4,5].
- *DaMaHub* will empower findability, accessibility, interoperability and reusability of scientific research data [6].
- *DaMaHub* enables the replication of scientific studies for future generations of researchers [7].

How does DaMaHub work?

We believe that to make science open, a data management solution has to fit into the daily data workflow of a researcher. Therefore *DaMaHub* is a combination of a distributed Open Science platform together with a local client on the researcher's computer. The local client allows data to be shared privately and securely with colleagues, as well as providing data versioning and effortless archiving.

For the researcher *DaMaHub* is, at first sight, just another directory on the local computer. You can save files in it, and *DaMaHub* takes care of the versioning in the background. Sharing access is as easy as sending an email. The system will prompt the user for additional metadata to help the collaborator to get started with the data. If required, the complete power of a state-of-the-art versioning system is available to researchers, making it easy to manage different versions and follow up changes or detect differences.

With *DaMaHub* it will also be easy to open up the research data, which will only require a few clicks. The decentralised network, based on a distributed file system (IPFS) [3] and blockchain technology [4,5], will take care of Open Data publishing, archiving and interaction with data on a worldwide scale. Social sharing, peer reviewing and re-use of open research data will be built into this system.

We already have a basic demonstration of the technology available on a trial basis at https://kubrik.io/demos/odr/ (password-based access for uploading available on request).

DaMaHub has the potential to be as transformative for scientific work as Github has been for programming and will generate a similar level of interaction with open research data.

Our team

Denis Parfenov, CEO, MBS (International Business), MA (International Relations). Member of Open Data Governance Board, Open Knowledge Ambassador for Ireland [8].

Dr Ingo Keck, CTO, Dr. europ. Computer Science, Dipl. physicist. Has 10 years' experience as a researcher, system architect and full stack coder [8].

Notes

[1] Statement by Prof. Dr. Mark Ferguson, Science Foundation Ireland, August 2016
[2] 'Incentive malus: Poor scientific methods may be hereditary" via The Economist www.economist.com/news/science-and-technology/21707513-poor-scientific-methods-may-be-hereditary-incentive-malus [cited 22 September 2016]
[3] Interplanetary File System https://ipfs.io
[4] "Introduction: What Is A Blockchain?" Ingo Keck. 12 January 2017. [cited 23 January 2017]
https://medium.com/@ingokeck/introduction-what-is-a-blockchain-f22510332bff#.ramz7ngqj
[5] "All you need to know about blockchain, explained simply." Rosamond Hutt. 17 June 2017. [cited 23 January 2017]
https://www.weforum.org/agenda/2016/06/blockchain-explained-simply/
[6] Since July 2016 every application under Horizon 2020 program must include Data Management Plan - H2020 Program Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 3.1., 25. page 8, August 2016
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf [cited 23 January 2017]
[7] All publicly funded research outcomes (data and articles) must be made as Open Access under Horizon 2020 rules
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [cited 23 January 2017]
[8] Our team https://kubrik.io/about/team/

DaMaHub(https://damahub.com) is a project by Kubrik Engineering(https://kubrik.io) in collaboration with Open Knowledge Ireland(https://openknowledge.ie) under guidance from internationally recognised advisory board and with advice from IPFS core team.

Contact us via email denis@kubrik.io or twitter @KubrikHQ