

Setting up a National Research Data Curation Service for Qatar

Arif Shaon
Qatar National Library

Armin Straube
Qatar National Library

Krishna Roy Chowdhury
Qatar National Library

Abstract

Over the past decade, Qatar has been making considerable progress towards developing a sustainable research culture for the nation. The main driver behind Qatar's progress in research and innovation is Qatar Foundation for Education, Science, and Community Development (QF), a private, non-profit organization that aims to utilise research as a catalyst for expanding, diversifying and improving the country's economy, health and environment. While this has resulted in a significant growth in the number of research publications produced by Qatari researchers in recent years, a nationally co-ordinated approach is needed to address some of the emerging but increasingly important aspects of research data curation, such as management and publication of research data as important outputs, and their long-term digital preservation. Qatar National Library (QNL), launched in November 2012 under the umbrella of QF, aims to establish itself as a centre of excellence in Qatar for research data management, curation and publishing to address the research data-related needs of Qatari researchers and academics. This practice paper describes QNL's approach towards establishing a national research data curation service for Qatar, highlighting the associated opportunities and key challenges.

Received 23 February 2017

Correspondence should be addressed to Dr. Arif Shaon, Qatar National Library, Doha, PO Box. 5825, Qatar. Email: ashaon@qnl.qa

An earlier version of this paper was presented at the 12th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution 4.0 International Licence. For details please see <http://creativecommons.org/licenses/by/4.0/>



Introduction

The provision of research data curation services for the State of Qatar is part of the portfolio of Qatar National Library (QNL)¹. The library was originally established as a central library in 2008 before it was officially launched in November 2012 as a national library – a non-profit organization under the umbrella of Qatar Foundation for Education, Science, and Community Development (QF)².

The main goal of the library is to support Qatar on its journey from a carbon-based to a knowledge-based economy by providing resources and services to students, researchers and the community in Qatar. This is supported through three main functions of QNL: a national library for Qatar, a research library to support education and research at all levels, and a public library to serve the metropolitan area of Doha. QNL is also providing a range of services to other Qatari institutions.

As a research library, QNL aims to establish itself as a centre of excellence for research data management, curation and publishing to help Qatari researchers and academics to address their research data-related needs. Research data curation is therefore a key element of QNL strategic objectives and is offered as part of QNL's core services to the wider research community in Qatar. QNL plans to underpin the overall curation service with a national infrastructure for research data curation for Qatar to ensure wider efficiency.

QNL also recognizes that developing a national infrastructure for research data curation requires a robust and sustainable service model involving collaborative efforts with its stakeholders, along with the use of extensible and future-proof technical solutions. The wider goal is therefore to develop a sustainable information ecosystem for managing and disseminating all research data in Qatar.

This practice paper describes the journey of Qatar National Library (QNL) towards establishing a national research data curation service for Qatar, highlighting the associated opportunities and key challenges.

Research Landscape in Qatar

Qatar has a developing research landscape with a number of significant achievements in multiple disciplines over the past decade. In astronomy, research projects involving the scientists at the Qatar Environment and Energy Research Institute (QEERI)³, a member of QF, led to the discovery of five exoplanets that bear the country's name (Qatar-1b to Qatar-5b) in June 2016 (Alsubai et al., 2016). Also in particle physics, Qatar-based researchers collaborated with CERN⁴ to upgrade the Large Hadron Collider in 2013 (QF Telegraph, 2013).

The main driver behind Qatar's progress in research and innovation is the country's vision of becoming an international research centre – this vision is an integral aspect of the strategic objectives of Qatar Foundation, the governing body for QNL as an institution. QF aims to utilise research as a catalyst for expanding and diversifying the

¹ Qatar National Library: <https://www.qnl.qa>

² Qatar Foundation: <https://www.qf.org.qa>

³ Qatari Environment and Energy Research Institute: <https://hbku.edu.qa/QEERI>

⁴ CERN: <https://home.cern>

country's economy, enhancing educational offerings and developing areas that affect the community, such as health and environment.

To achieve this goal, QF inception in 2006 the Qatar National Research Fund (QNRF)⁵, the leading institution in Qatar dedicated to funding research in the national interest. From a wider viewpoint, the QNRF has been funding research both locally and internationally to foster a sustainable research culture in Qatar. In 2008, QNRF's mission received significant impetus as the Amir of Qatar, His Highness Sheikh Hamad bin Khalifa Al-Thani, issued Decree-Law No. (24) Of 2008, which raised Qatar's annual investment in research and development to 2.8 percent of total annual government revenue (Cecchine et al., 2012). In 2012, the Qatar National Research Strategy (QNRS)⁶ was published to provide further assertion to Qatar's vision to become "a leading center for research and development excellence and innovation". The QNRS report was later revised in 2014 with a specific focus on Qatar's Grand Challenges in Water Security, Energy Security, Cyber Security and Health, and has been underpinning QNRF's mission as the leading funding body for research in Qatar.

The result of QNRF's effort has been a steady increase in the number of QNRF-funded projects with the overall count of Qatar-based research publication amplifying (approximately 2,500 publications disseminated via Qatar Research Portal⁷) over the recent years. In general, the QNRF partners in government, education and business, as well as other QF members, have all been active participants in producing wider and even richer results, in order to establish a sustainable research culture in Qatar.

To assess progress toward developing a Qatari research culture, QNRF has established metrics for measuring a number of key characteristics, including the number of publications produced by QNRF-funded research projects and the efficacy of the underlying infrastructure provided by QNRF. The latter focuses mainly on availability of sufficient financial resources managed by a visionary and efficient governing body. However, data on researchers, research capacity, ongoing research projects and research outcomes are currently collected individually at scattered research institutions throughout Qatar and in various formats. Disparate data storage and management practices exist at Qatar Foundation's three research institutes, Qatar Science and Technology Park (QSTP)⁸, QNRF and QF's publishing service QScience⁹, as well as Qatari universities and colleges. This also concerns government institutions, including the Ministry of Public Health¹⁰ and private entities that are engaged in conducting research.

The present situation of fragmented and scattered sources of research data and information that is currently going on in Qatar may limit utilisation of available national research potential, and thereby allowing research duplication and reduced research focus in areas of high priority in accordance with the QNRS. This also poses challenges in generating nation-wide reports, budgets and statistics on research related matters.

In view of the aforementioned challenges, and as implemented by some of the international funding bodies, such as NSF¹¹, on which QNRF is modelled, Qatar needs a nationally co-ordinated approach to addressing the emerging but increasingly important

⁵ Qatar National Research Fund: <https://www.qnrf.org/en-us/>

⁶ Qatar National Research Strategy: <https://www.qnrf.org/en-us/About-Us/QNRS>

⁷ Qatar Research Portal provides access to the outcomes of QNRF-funded research to advance the dissemination of knowledge to all members of the community. See: <http://qnrffeddiscovery.qnl.qa/qnrf-home>

⁸ Qatar Science and Technology Park: <https://qstp.org.qa>

⁹ QScience: <http://www.qscience.com>

¹⁰ Ministry of Public Health: <https://www.moph.gov.qa/pages/default.aspx>

¹¹ National Science Foundation, USA: <https://www.nsf.gov/>

aspects of research projects, such as publication of data as important research outputs, research data management and long-term preservation, i.e. the key elements of research data curation.

Promisingly, QNRF does recognise the need to address research data curation at a national level, and has commenced a collaboration with other partners, in particular with QNL to develop effective solutions. Early positive indicators of QNRF's effort include the 2016 workshop series on research data management, jointly organised by QNRF and QNL, and attended by over 200 researchers from various academic, government and private institutions.

QNL Research Data Curation Service Model

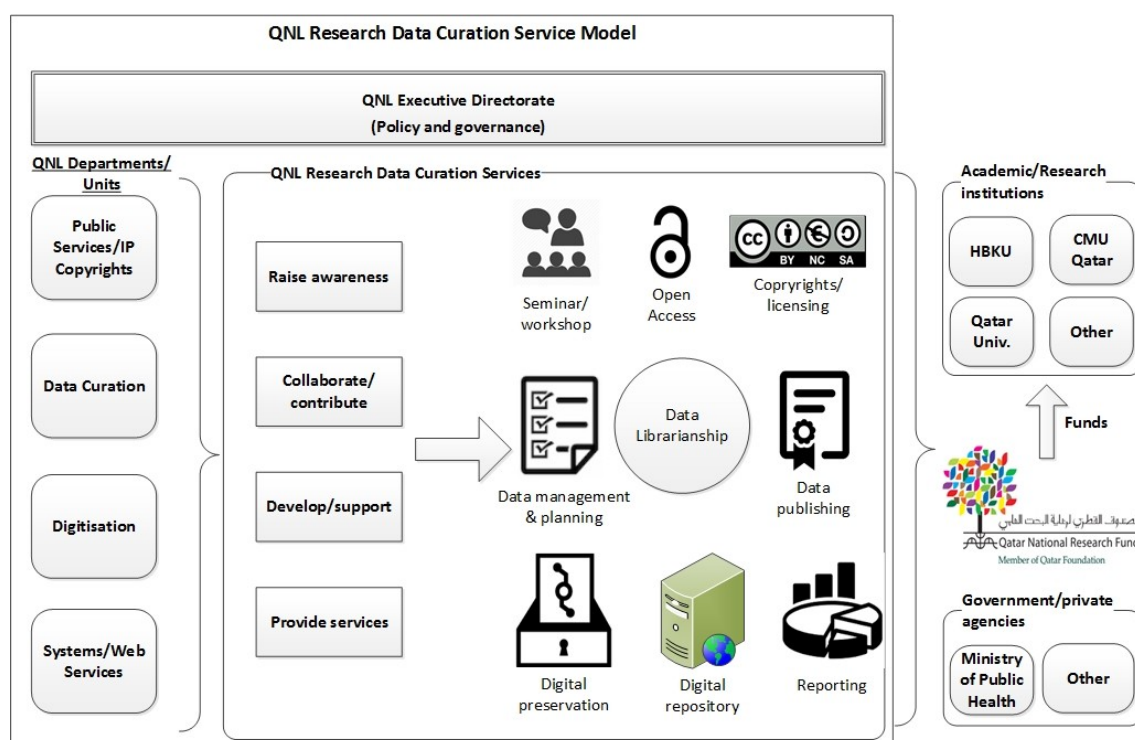


Figure 1. QNL Research Data Curation service model.

The QNL research data curation strategy is primarily focused on developing a robust and sustainable service model (see Figure 1) for research data curation in Qatar, involving collaborative efforts with its stakeholders along with use of extensible and future-proof technical solutions. This service model forms part of QNL's governance framework for its research data curation service by recognizing that the research data with long-term value produced in Qatar are an important part of the nation's heritage.

The wider purpose of this service model is to position QNL as a centre of excellence in Qatar for research data management, curation and publishing, envisioned to be achieved through the following key processes:

- Collaborate with leading research institutions in Qatar to raise awareness of the issues and challenges associated with research data management and publishing, such as copyright, licensing, intellectual property and ethics constraints;

- Establish a best-practice culture for research data management through multidisciplinary data librarianship within the wider research communities;
- Develop a national infrastructure for managing and disseminating all research data in Qatar. The infrastructure will include sustainable workflows for ingesting research data from various stakeholders into the QNL digital repository for long-term preservation. This will also help minimize the risk of losing research that are unique or difficult to reproduce;
- Maximise initial investments by funding bodies in research projects in Qatar by facilitating effective long-term reuse of the corresponding research outputs, including both the final and interim data results as well as the associated metadata; and
- Increase the overall benefits of research projects by facilitating long-term availability of research data that are important for developing social, health and environment-related policies or legislations in Qatar.

The remainder of this section describes QNL's approach to setting up the aforementioned processes and the associated challenges and opportunities.

Collaboration with Internal and External Stakeholders

As illustrated in Figure 1, research data curation at QNL caters to the need of several stakeholders:

- Qatar National Research Fund
- Local research institutions
- Local campuses of internationally renowned research universities
- Individual researchers
- Government agencies

As mentioned before, a key partner in the development of research data curation service at QNL is QNRF. The common aim is to retain publicly funded research data in Qatar and to make it accessible and reusable for future research. QNL has proposed to offer the preservation and access infrastructure and a joint working group explores opportunities to implement active research data management in QNRFs reporting cycle. While the options are being investigated, no formal decision has been reached yet and the retention of research data sets or the submission of data management plans is not yet mandatory.

Research efforts at QNRF are focussed on a limited number of focus areas, outlined in the National Research Strategy¹². This is an advantage for providing nation-wide coverage for research data curation.

Qatar hosts two domestic universities (Hamad bin Khalifa University¹³ and Qatar University¹⁴), and a variety of local research institutes both public and (semi-)private. QNL is actively raising awareness with those institutions that their generated research data is of high value. Because only a minority of these institutions has the capacity to build an in-house repository infrastructure, QNL is in a good position to offer its

¹² National Research Strategy: <https://www.qnrf.org/Portals/0/Download/QNRS%202014.pdf>

¹³ Hamad bin Khalifa University: <https://hbku.edu.qa>

¹⁴ Qatar University: <http://www.qu.edu.qa>

services in regard to the management of the data and as a repository for preservation and access.

A special case involves the local campuses of renowned foreign research universities, including US and UK-based universities that receive significant amount of funding from QF. Every year, QF reportedly allocates about US\$320 million to run six campuses of US universities¹⁵. Set up to jumpstart higher education and research on an advanced level, these dependencies of large institutions can rely on their home campuses and institutions to provide support with data management and repository services. Nonetheless, QNL advocates the retention of all research data funded by Qatar in the country to ensure access and usability.

Individual researchers are already aware of the need for good data management practises to support their research goals. Subscription numbers to QNL research data management workshops are very high. The majority of researchers at our workshops stated nonetheless that they are sceptical about publication of their research data, due to the effort involved, the little reward publication brings compared to traditional publications and the desire to exploit their research data exclusively. Transforming the awareness of the need to curate research data within institutions and among researchers into actual practice will be a key challenge of QNLs data curation service.

Government agencies are also increasingly aware of the value of their data. The Ministry of Information & Communications Technology (ictQATAR, 2015) has issued a data management policy¹⁶ that governs exchange of data between government agencies and with the general public. QNL was able to submit comments on the policy via a process orchestrated by Qatar Foundation, but it proved to be challenging to influence policy setting in government agencies or to advocate the need of research institutions in regard to government data. As part of the semi-governmental QF, QNL is not a government agency, even though it is funded by the government. Being part of the Qatar Foundation has otherwise proved to be very beneficial for the aim of establishing a national research data curation service. QNRF and a majority of research institutions in Qatar belong to Qatar Foundation as well, so establishing working relations is supported by the Qatar Foundation.

Internationally, QNL data curation team plans to collaborate with relevant initiatives, such as the Research Data Alliance¹⁷, to contribute to new innovations in research data curation and digital preservation as well as strengthening the QNL data curation service. This will position QNL as a leading institution in both data curation and digital preservation for Qatar and in the Gulf region.

Internally, the data curation team is part of QNL's Digital Initiative department and uses the libraries repository system and application development capacities. It is closely associated with other services hosted by QNL Digital Initiative, like the digital preservation service and the National Digitization Centre. The data curation team can also rely on the discipline specific knowledge of its subject librarians. This facilitates the adoption of an institutional strategy, rather than an isolated one, for the provision of efficient research data curation service.

¹⁵ See: <https://www.qnrf.org/en-us/Newsroom/In-the-Media/ArtMID/649/ArticleID/935/Science-Led-Economies>

¹⁶ IctQATAR data management policy:
http://www.motc.gov.qa/sites/default/files/data_management_policy.pdf?TSPD_101_R0=43985aa0282462f223571751c432b9a1xy600000000000000056cc0856ffff0000000000000000000005aa7dff00af9bc4ef

¹⁷ Research Data Alliance: <https://www.rd-alliance.org>

Training and Consultancy

QNL recognises the importance of incorporating research data management practice into the early stages of a research project lifecycle in order to facilitate production of good quality research data and associated metadata by that project. To that end, QNL organises training workshops on research data management plans for researchers who start new research projects. It is envisioned that dedicated workshops dealing with the publication and dissemination of research data sets will complement the training activities at a later stage. The workshops are strategically aligned with QNRFs funding cycle and are tailored to the need of specific disciplines, along with individual face-to-face consultancy with researchers as needed. As mentioned before, the 2016 workshop series on research data management, jointly organised by QNRF and QNL, was attended by over 200 researchers from various academic, government and private institutions working in health, social science, natural science and engineering. Each attendee was awarded a certificate of attendance.

Although the number of attendees for the workshops was certainly encouraging and indicative of a positive step forward, the overall experience of the QNL data curation team interacting with the researchers identified the need to build sufficient data librarianship expertise within the team to cover all major areas of research in Qatar. The team also recognises the need to expand its capacity to support non-QNRF funded research projects.

In terms of opportunities, the workshops generated sufficient motivation from the researchers and QNRF as a funding body for embedding research data curation within QNRF-funded research workflows. Based on discussions around the approaches adopted by several foreign funding bodies, in particular NSF, the option to incorporate data management planning and training into QNRF grant application process was considered. For QNL, this would require further development of both its relationship with QNRF and its technical capacity.

Over the long term, delivering training and workshops on research data management to the wider research communities should establish QNL as a centre of excellence in this area for all Qatar-based research.

Central to the QNL research data curation service infrastructure is the QNL digital repository, which was set up in January 2016 to serve a variety of functions besides research data curation (See Figure 2). The overall repository architecture complies with the OAIS Reference Model (CCSDS, 2012), and is based on Islandora version 7.x-1.6¹⁸, an open-source and modular digital content management system, integrated with Fedora Commons (version 3.8.1)¹⁹ as the underlying repository technology.

Islandora provides a range of Drupal²⁰ modules that support research data curation. In particular, the repository will support assignment of Digital Objects Identifier (DOIs) to all digital objects, including research datasets to enable standard-based scholarly communication for the objects. In addition, the capacity of the Islandora modules can be enhanced – e.g. by integrating with external solutions – to accommodate future requirements, such as provision of a research data management planning function.

¹⁸ Islandora: <https://islandora.ca>

¹⁹ Fedora Commons: <http://fedorarepository.org>

²⁰ Drupal: <https://www.drupal.org>

An Extensible Technical Infrastructure

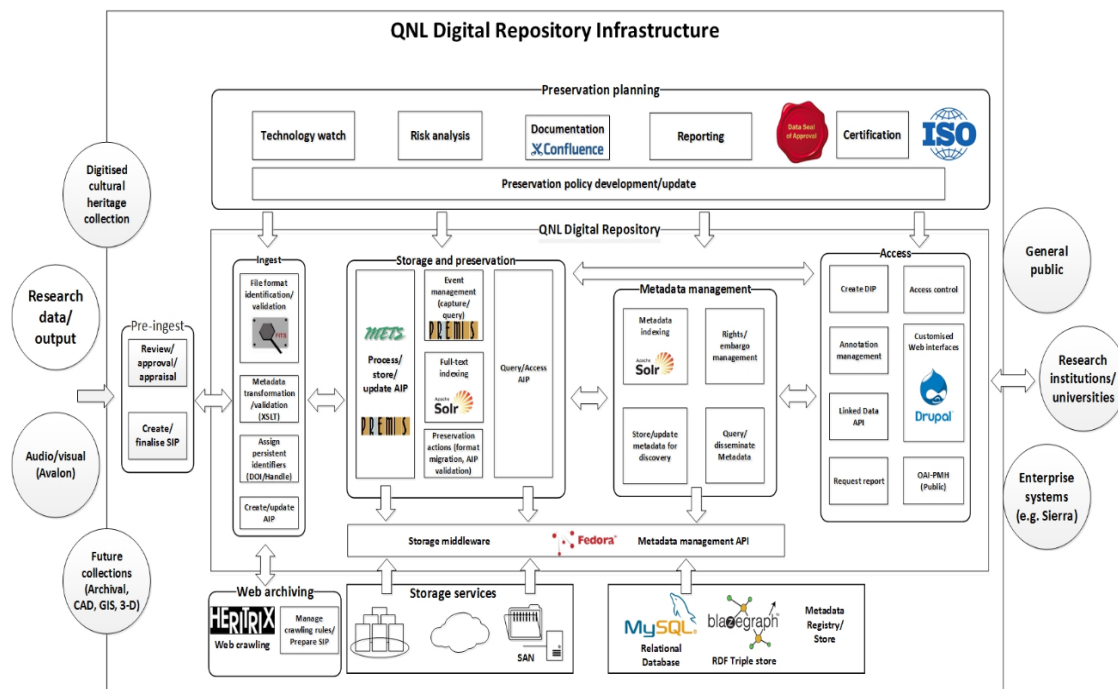


Figure 2. QNL digital repository infrastructure (Straube et al., 2016).

The digital repository is integrated with the QNL state of the art, highly robust and secure storage infrastructure. Storage capacity is as follows (January 2017):

- Disk Capacity – 1.35PB (1,378.3 TB)
- Tape Capacity – 7.6 PB (7,780 TB), with a robotic tape library scalable to 56 Drives, 252 MB/sec – 50.8 TB/hr and 3000 Tape Slots, 8TB/tape – 24 PB

Since QNL is at the beginning of offering its services, some of the technical requirements and workflows are still under development and no datasets have yet been formally published. So far the systems supports bit stream preservation for any content and additional support for a variety of image and text based objects, audio-visual content and web archives.

Developing a technical infrastructure to support data curation at a national level is expected to face challenges. For QNL, a key challenge is to design and implement efficient workflows for supporting curation of “active” research data – i.e. the data produced or sourced by researchers that contribute to the production of the final dataset. In addition, integrating the QNL data curation service with discipline-specific research workflows to enable seamless data ingest experience to researchers will need to be supported. This requires building sufficient technical expertise, including software development, within QNL to cover all aspects of the development of the QNL research data curation service.

In terms of opportunities, QNL is expected to have a unique set of use cases to contribute to development of Open Source curation tools, including the Islandora modules.

Timeline for QNL Research Data Curation Service

QNL has been making steady progress towards setting up a national research data curation service for Qatar over a two-year timeline (see Figure 3). The timeline has been kept strategically short to ensure tangibility of the goals, achievement of which will determine future extensions as deemed appropriate by the QNL directorate.

Since the data curation unit was formed in October 2015, the focus has been to establish the technical infrastructure, particularly the digital repository to support preservation and dissemination of digital objects. As the initial phase of the digital repository development is nearing completion in the first quarter of 2017, the focus will shift to establish workflows with key stakeholders for acquiring research data for preservation and publication using DOIs. The effort to organise workshops and training sessions on research data management will continue throughout 2017 and is expected to strengthen the relationship between QNL and QNRF. Another key milestone for 2017 is to acquire the Data Seal of Approval²¹ certificate for the QNL digital repository, as this will help promote the repository as a trusted repository to the wider research communities in Qatar.

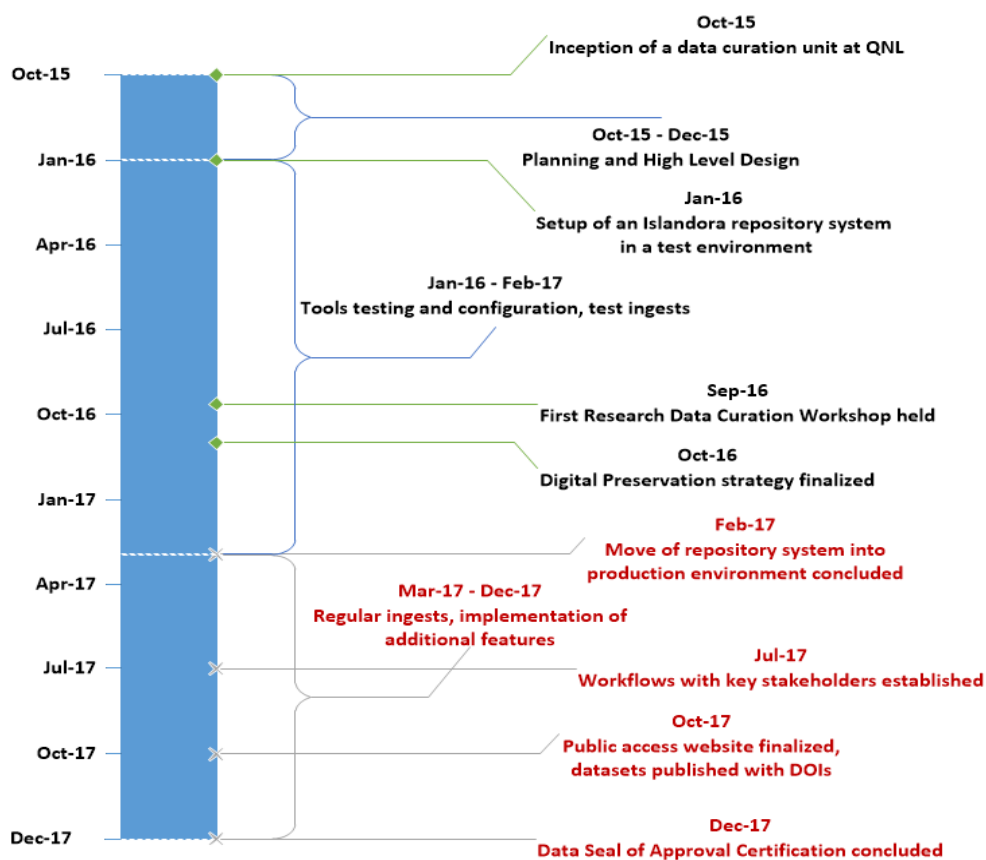


Figure 3. Two year timeline for developing QNL Research Data Curation service.

²¹ Data Seal of Approval: <https://www.datasealofapproval.org/en/>

Conclusion

The development of the QNL research data service is underway. The digital repository system is being developed to offer preservation and dissemination services to a greater variety of research data. This is complemented by a wider support model for Qatar-based research, being established through collaboration with key stakeholders. Importantly, progress in raising awareness through workshops and customised training sessions with the researchers is also being made.

The library, however, recognises that setting up a national service is a monumental task and substantial work is still remaining before success can be measured. Nevertheless, QNL has taken confident steps over the past two years in addressing, where feasible, key challenges and utilising opportunities through a conceptually sound service model. The library is currently focusing on the implementation of the service model, and thereby making further progress towards developing a sustainable research data service for Qatar.

References

- Alsubai, K. A., Mislis, D., Tsvetanov, Z. I., Latham, D. W., Bieryla A., Buchhave L. A., Esquerdo G. A., Bramich D. M., Pyrzas S., Vilchez, N. P. E., Mancini, L., Southworth, J., Evans, D. F., Henning T. & Ciceri, S. (2016). Qatar Exoplanet Survey: Qatar-3b, Qatar-4b and Qatar-5b. *arXiv:1606.06882v3 [astro-ph.EP]*. Retrieved from <https://arxiv.org/abs/1606.06882>
- Cecchine G., Darilek, R. E., Harrell, M. C., Mattock, M. G., Culbertson, S., Ortiz, D. S., Maynard, N. C., Balkovich. S. R. B. E., & Nardulli, B. R. (2012). *Sustaining the Qatar National Research Fund*. Report by RAND-Qatar Policy Institute, 2012. Retrieved from http://www.rand.org/content/dam/rand/pubs/technical_reports/2012/RAND_TR1125.pdf
- Consultative Committee for Space Data Systems. (2012). *Reference model for an Open Archival Information System (OAIS)* (Magenta Book CCSDS 650.0-B-1). Retrieved from <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- Ministry of Information & Communications Technology. (2015). *Data Management Policy*. Retrieved from http://www.ictqatar.qa/sites/default/files/data_management_policy.pdf
- QF Telegraph (2013, November). Qatar's Potential Praised During CERN Workshop. *Qatar Foundation Telegraph, issue 19 November 2013*. Retrieved from <http://www.qf.org.qa/content/qf-telegraph/issue-98/qatar-potential-praised-during-cern-workshop>

Straube, A., Shaon A., Abo Ouda, M. (2016). Digital Preservation with the Islandora Framework at Qatar National Library. *Proceedings of 13th iPRES conference 2016 in Bern, Switzerland*. (pp. 270-271) Retrieved from https://ipr16.organizers-congress.org/frontend/organizers/media/iPRES2016/_PDF/IPR16.Proceedings_4_Web_Broschuere_Link.pdf