Author(s): Korhonen, Ilari; Nurminen, Miika

Title: Status and Prospects of Kanki : An Open Source Cross-Platform Native iRODS Client Application

Year: 2016

Version:

Please cite the original version:

Korhonen, I., & Nurminen, M. (2016). Status and Prospects of Kanki : An Open Source Cross-Platform Native iRODS Client Application. In iRODS User Group Meeting 2016. Proceedings (pp. 69-74). iRODS Consortium. http://irods.org/wp-content/uploads/2016/12/irods_ugm2016_proceedings.pdf

Status and Prospects of Kanki: An Open Source Cross-Platform Native iRODS Client Application

Ilari Korhonen*, Miika Nurminen
IT Services, University of Jyväskylä
PO Box 35, 40014 University of Jyväskylä, Finland
ilari.korhonen@icloud.com, miika.nurminen@jyu.fi

ABSTRACT

The current state of development of project Kanki is discussed and some prospects for future development are laid out with reflection on the results of the research IT infrastructure project at the University of Jyväskylä. Kanki is a cross-platform native iRODS client application which was introduced to the iRODS community at the iRODS Users Group Meeting in 2015, and later released as open source. A total of 9 releases have been made, from which the latest 6 have been available in addition to the source code as pre-built binary packages for x86-64 CentOS Linux 6/7 and OS X 10.10+. The Kanki build environment at the University of Jyväskylä is running out of Jenkins continuous integration for both previously mentioned platforms utilizing disposable containers instantiated from pre-built Docker images for Linux builds. This provides an excellent framework for (regression) testing of the client suite. The immediate goals of development include: stability, testing, ease of install and use, and a complete iRODS basic feature set for graphical icommands alternatives. The prospects for more advanced future development include: a fully extensible modular metadata editor with pluggable attribute editor widgets, a fully extensible modular search user interface with pluggable condition widgets, data grid analytics, and visualization with VTK integration.

Keywords

Research data, iRODS, client software, graphical user interface, continuous integration, research support services.

INTRODUCTION

About a year ago in 2015, the cross-platform native iRODS client application Kanki was introduced to the iRODS community at the 7th Annual iRODS Users Group Meeting [1]. The Kanki iRODS client features a responsive UI with native look & feel to the desktop enabled by the Qt[1] framework, integration to Kerberos authentication with the option to use iRODS 4.x SSL secured connections. Metadata management features include a schema definition language, field validators, and type-specific display views and filtering. The client application is targeted towards researchers of various disciplines as well as other interest groups utilizing or curating research data (e.g. librarians). Users can utilize the full power of an iRODS data grid complete with powerful data management functions via its intuitive user interface. Kanki was released as open source with a 3-clause BSD license in GitHub[2] in September 2015. Kanki has been used in pilot projects at the University of Jyväskylä and has attracted interest from other research institutions as well.

In this paper, we focus on the recent development efforts of the Kanki iRODS client, as well as reflection on the research IT infrastructure development project that enabled the development. The paper is concluded with development ideas and a possible sketch on the role of Kanki and iRODS as part of the research support services at the JYU.

---

*Present address: PDC Center for High Performance Computing, KTH Royal Institute of Technology, SE-100 44, Stockholm, Sweden

[1] `http://www.qt.io/`

[2] `https://github.com/jyukopla/kanki-irodsclient`

iRODS UGM 2016 June 8-9, 2016, Chapel Hill, NC

KANKI DEVELOPMENT

Kanki is being developed as a cross-platform project with a single C++11 code base. Currently Kanki builds successfully on Linux (Red Hat and Ubuntu tested and documented) and Mac OS X 10.10/10.11. The multi-platform portability of the source is enabled by the portability of the Qt framework and the C++11 standard library.

Build Process

Currently the multi-platform build is being done via the Qt `qmake` utility, which generates a build environment to be executed with GNU Make. There is a simple build script in the GitHub repository called `build.sh` which builds the source package on Linux and OS X. The build script can be provided the location of the Qt framework as well as some other arguments. Help can be printed out with the `-h` switch.

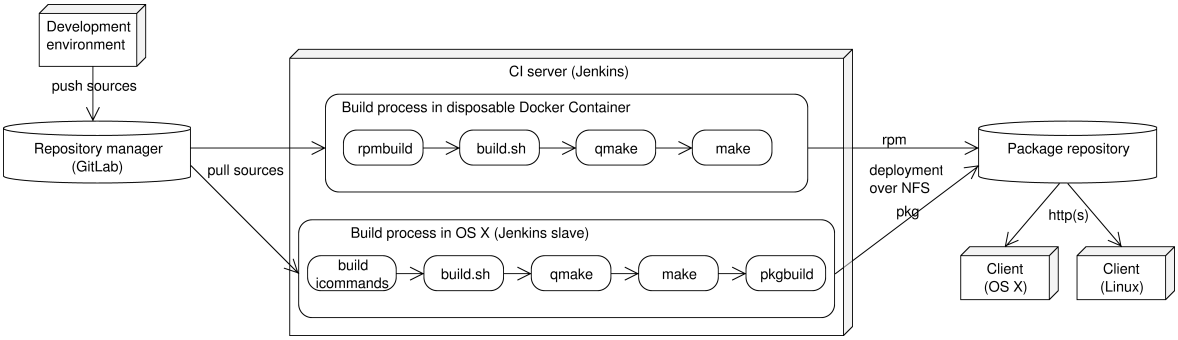

Figure 1. Illustration of the kanki-irodsclient build process.

Since the iRODS project is migrating its build environment into CMake (which also enables very convenient VTK/Qt linkage), work is currently in progress to switch the build environment to CMake. There is an experimental build environment which builds via CMake in the `develop` branch of the GitHub repository. A shadow build via CMake can be done as follows:

```
$ git clone https://github.com/ilarik/kanki-irodsclient.git -b develop
$ mkdir build_kanki-irodsclient; cd build_kanki-irodsclient
$ cmake -DCMAKE_PREFIX_PATH:PATH=/Users/tiilkorh/Qt/5.5/clang_64/lib/cmake ../kanki-irodsclient/src
-- The C compiler identification is AppleClang 7.3.0.7030031
-- The CXX compiler identification is AppleClang 7.3.0.7030031
-- Check for working C compiler: /Applications/Xcode.app/Contents/Developer/Toolchains/XcodeDefault.xctoolchain/usr/bin/cc
-- Check for working C compiler: /Applications/Xcode.app/Contents/Developer/Toolchains/XcodeDefault.xctoolchain/usr/bin/cc -- works
-- Detecting C compiler ABI info
-- Detecting C compiler ABI info - done
-- Detecting C compile features
-- Detecting C compile features - done
-- Check for working CXX compiler: /Applications/Xcode.app/Contents/Developer/Toolchains/XcodeDefault.xctoolchain/usr/bin/c++
-- Check for working CXX compiler: /Applications/Xcode.app/Contents/Developer/Toolchains/XcodeDefault.xctoolchain/usr/bin/c++ -- works
-- Detecting CXX compiler ABI info
-- Detecting CXX compiler ABI info - done
-- Detecting CXX compile features
-- Detecting CXX compile features - done
-- Configuring done
-- Generating done
-- Build files have been written to: /Users/tiilkorh/tmp2/build_kanki-irodsclient
$ make -j 8
[ ... lots of nicely formatted CMake build output ... ]
[100%] Linking CXX executable irodsclient.app/Contents/MacOS/irodsclient
[100%] Built target irodsclient
```

The build environment for the Kanki iRODS client at JYU is running out of Jenkins CI on two slaves, one with Docker capabilities and another slave running OS X. Linux builds are currently being executed in disposable containers instantiated from pre-built Docker images in a Jenkins slave with Docker. The OS X builds are being done against a prebuilt (another Jenkins job) irods-icommands distribution for OS X from the `4-1-stable` branch of the iRODS GitHub repository. Instructions for the OS X build have been published in the iRODS blog[3].

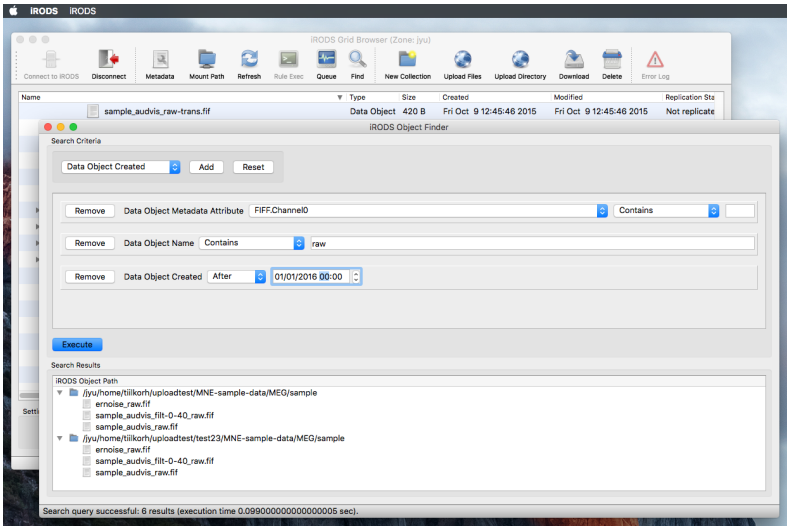

Figure 2. Object Finder component.

New and Prospective Features

Specific needs not properly accommodated by other existing freely available solutions included the graphical iRODS search tool with arbitrary search criteria formation for data discovery, metadata schema management with visual namespace and attribute views, and readiness for metadata schema validation for data quality assurance.

As the client is intended to eventually serve as a bona fide alternative user interface to iRODS icommands – the reference user interface for iRODS. Implementing all of this functionality in a native "desktop" application will enable the users to harness the full power of iRODS with native application performance and easy-to-learn graphical user interface. The following features are considered to be developed further:

- drag & drop inside iRODS and between desktop and iRODS
- synchronization of iRODS collections with local paths
- editing of access control lists and groups for groupAdmin role users
- metadata editor schema management with validation
- rule engine queue management and rule exec interface
- data discovery i.e "Find" UI for arbitrary metadata search criteria execution (GenQuery) (see Figure 2)
- VTK[4]-based visualization tools for data grid analytics (e.g. object relations, see Figure 3)

[3] `http://irods.org/2015/10/native-gui-access-to-irods-on-a-mac-or-linux-desktop/`
[4] `http://www.vtk.org/`

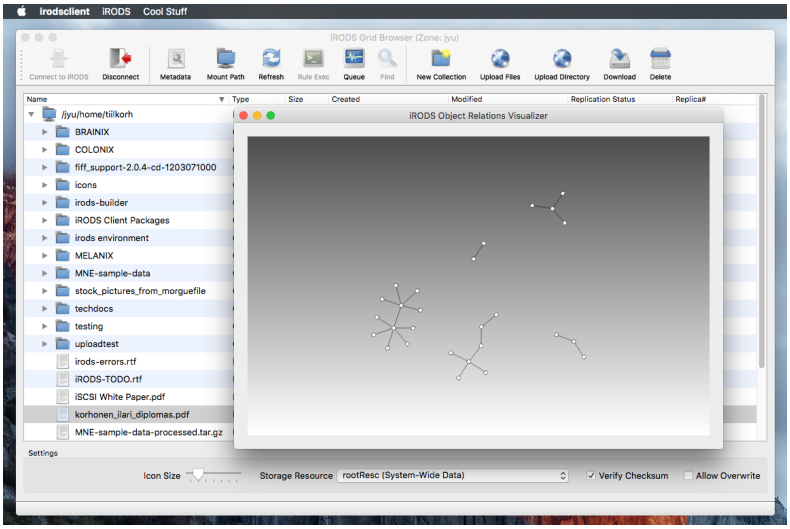

Figure 3. Experimental iRODS Object Relations Visualizer.

Windows still remains as an unsupported platform since iRODS 4.0 isn't Windows compatible at the time of writing this paper. With Windows support added to the iRODS codebase our client can be built on Windows as well.

REFLECTION ON THE RESEARCH IT INFRASTRUCTURE DEVELOPMENT PROJECT

Initial implementation of Kanki client has been done under the development project for research IT infrastructure and research data management, active in 2013-2015 at the University of Jyväskylä. One of the project goals was putting the university-level principles for research data management [2] (e.g. processes for handling research data, secure infrastructure for data storage and access, standard metadata descriptions, advancing open science) into practice. Dataverse[5] was first adopted, making it possible to publish datasets in citable format with on-line analysis capability [3]. iRODS data grid has been adopted with additional development of server-side iRODS modules for e.g. automatic metadata extraction and finally, the development of Kanki client. A recommended minimum metadata model has been developed with JYU Library, providing also training for research data management practices.

iRODS and Kanki client have been utilized in several pilot projects, including the Jyväskylä Centre for Interdisciplinary Brain Research[6], JYFL Accelerator laboratory[7], and datasets from the Department of Music. Various datasets – including video, audio, motion detection data, EEG measurements, and transcribed interviews – have been imported to the system. Metadata extraction functionality has been implemented for DICOM[8], EXIF[9], IPTC[10], and FIFF[11] formats. Faculty of Information Technology is evaluating the suitability of iRODS as the storage backend for Faculty's local computing cluster.

There have been multiple, partly concurrent efforts related to research data management and, more generally, research support services – most of these overseen at the JYU by Steering group for Research information systems[12], Steering

[5] https://dvn.jyu.fi/
[6] http://cibr.jyu.fi/
[7] https://www.jyu.fi/fysiikka/en/research/accelerator/
[8] http://dicom.nema.org/standard.html
[9] http://www.jeita.or.jp/cgi-bin/standard_e/list.cgi?cateid=1&subcateid=4
[10] https://iptc.org/standards/photo-metadata/iptc-standard/
[11] http://www.aston.ac.uk/lhs/research/centres-facilities/brain-centre/facilities-clinical-services/meg-studies/downloads/
[12] https://www.jyu.fi/hallinto/tyoryhmat/tutkimuksen_tietojarjestelmat_or/en

group for the Digitalisation project[13], and the Science council[14]. Nationally, Open Science and Research (ATT[15]) project has produced services and recommendations related to e.g. open research data. All departments of the University have provided documentation of their research infrastructures[16]. Procedures for digitising analogue content have been implemented by the University Printing Services. JYU is currently replacing the legacy research information system[17] with a new CRIS[18] (Current Research Information System). Jyväskylä University Library is planning a model for centralized open science services and promoting parallel publishing in a joint development project with University of Eastern Finland Library [4], expanding on highly successful centralized publication recording process [5] in JYU, realized by close collaboration between the Library, Research management, and IT Services. National Data Management Plan tool Tuuli[19] is being tested by various Finnish universities. After the adoption of the new CRIS, data management plans will be systematically collected as part of research project -specific records.

Despite the various efforts and activities in the IT Research infrastructure project, it has to be concluded that even though the implementation of Kanki client and the technical infrastructure can be considered successful, the wider goal of executing the principles of data management has turned out to be challenging. It has proved to be of considerable difficulty to advance standard data management practices when the research itself is done more or less independently of administrative processes, often using specialized tools and software for e.g. analysis on datasets. The majority of JYU researchers have not yet used iRODS or Dataverse. IT Services unit has sometimes been perceived as not sufficiently accommodating the needs of various research groups that may have differing requirements in terms of hardware, software, scalability, or access rights. The services have been improved during the project, but yet more steering may be needed related to job function scoping and e.g. pricing models – the cost of centralized storage has not been competitive compared to ad hoc solutions such as portable drives. Another identified challenge has been the lack of resources related to coordinating the activities of different actors related to research support services.

CONCLUSION

The development of Kanki client is still at relatively early stages. The solution has provoked interest from multiple research institutions and developer community seems to be building up in GitHub. The build process utilizing Jenkins and Docker seems promising considering regression testing and streamlined deployment, possibly utilizing Ansible in the server-side in the future. iRODS maintenance at JYU is being transferred from IT development services to storage services, thus making the iRODS environment as part of the standard service porfolio. Development of Kanki client continues as open source project. The software can be considered adequate in production settings, but is by no means "ready". Immediate goals of development include stability, testing, ease of install and use, and a feature set for graphical icommands alternatives. Concerning the usage related to different stages of research data lifecycle, longer-term development prospects include: a fully extensible modular metadata editor with pluggable attribute editor widgets, a fully extensible modular search user interface with pluggable condition widgets, data grid analytics, and additional VTK-enabled visualizations.

Even though iRODS or Dataverse have not yet been widely adopted by JYU researchers, the situation might change in the near future. It is likely that as funders like Academy of Finland or EU (Horizon 2020) start to demand open research data, repository services – be it university-specific, international (e.g. Zenodo, FigShare), or subject-based (e.g. Finnish Social Science Data Archive[20] – shall become more attractive. The general data protection regulation[21] in EU will considerably raise the requirements for all systems containing personal data – this will likely generate new use cases for iRODS.

---

[13] https://www.jyu.fi/hallinto/tyoryhmat/digitointityoryhma
[14] https://www.jyu.fi/hallinto/neuvostot/tiedeneuvosto/en/sciencecouncil
[15] http://openscience.fi/
[16] https://www.jyu.fi/palvelut/wolmar/palvelut/infrat
[17] http://tutka.jyu.fi/
[18] https://www.jyu.fi/yliopistopalvelut/str/erityistoiminnot/tietohallinto/cris-kayttoonottoprojekti
[19] https://www.dmptuuli.fi/
[20] http://www.fsd.uta.fi/en/
[21] http://www.consilium.europa.eu/en/policies/data-protection-reform/data-protection-regulation/

JYU Library and University Museum will form a new unit named *Open Science Centre* (OSC) as part of a more extensive organizational transformation – the University's structural development[22] – effective in 2017. The University is also mapping the present state of research support services and identifying potential development needs and gaps wrt. National Open Science and Research Reference Architecture[23]. The new organization might provide opportunities to further clarify service models and distribution of responsibilities – perhaps even establishment of new centralized functions. For example, there is an emerging need for new services to support open research *methods* (e.g. source code and notebooks [6]), and to perceive an expanding variety of tools related to different kinds or research workflows [7]. The long term goal for research IT infrastructure (at least for data-intensive disciplines) could be supporting fully reproducible research [8]: making the code and data available in a platform such that the data can be analyzed in a similar manner as in the original publication.

In the new organization, OSC is to take more comprehensive responsibility on coordinating research data management and digitisation activities. However, execution of the principles requires still substantial technical development, as well as architectural steering. CRIS has potential to be used as a (meta)data hub combining information about research infrastructures, projects, and outputs (e.g. publications and datasets), provided that sufficient resources are reserved for data integration. There is also a partial overlap between functionalities of the Institutional repository[24] and Dataverse. Further opportunities for integrating data or even establishing shared systems (cf. Tuuli DMP, National ORCID consortium[25]) by multiple universities should be explored to minimize duplicated work [9]. An important issue for the future is the acceptance of iRODS by the researchers. There is a growing need to support data management during the research life cycle as a whole [10] – we believe that iRODS is the principal enabling service to accomplish this, provided that it remains supported with sufficient advocacy and training.

REFERENCES

[1] I. Korhonen and M. Nurminen, "Development of a native cross-platform iRODS GUI client," in *Proceedings of iRODS User Group Meeting 2015*. The iRODS Consortium, 2015.

[2] A. Auer and S.-L. Korppi-Tommola, "Principles for research data management at the University of Jyväskylä," University of Jyväskylä, Tech. Rep., 2014. [Online]. Available: https://www.jyu.fi/tutkimus/tutkimusaineistot/rdmenpdf

[3] M. Crosas, "The Dataverse network®: an open-source application for sharing, discovering and preserving data," *D-Lib Magazine*, vol. 17, no. 1, 2011.

[4] P. Olsbo, A. Muhonen, J. Kananen, and J. Saarti, "Suomi rinnakkaistallentamisen mallimaaksi [Finland to become a model country for parallel publishing]," 2015, Portti. [Online]. Available: http://portti.avointiede.fi/tutkimusjulkaisut/suomi-rinnakkaistallentamisen-mallimaaksi

[5] M.-L. Harjuniemi, "Kirjasto yliopiston julkaisurekisterin ylläpitäjänä: kokemuksia keskitetystä kirjaamisesta [Library as maintainer of university's publication registry: experiences of centralized recording]," *Signum*, no. 2, 2015.

[6] A. Perrier, "Jupyter, Zeppelin, Beaker: The rise of the notebooks," 2015, ODSC blog. [Online]. Available: https://www.opendatascience.com/blog/jupyter-zeppelin-beaker-the-rise-of-the-notebooks/

[7] B. Kramer and J. Bosman, "101 innovations in scholarly communication - the changing research workflow," 2015, poster at FORCE2015 Conference.

[8] V. Stodden, F. Leisch, and R. D. Peng, Eds., *Implementing Reproducible Research*. Chapman&Hall/CRC,2014.

[9] M. Nurminen, "Preparing for CRIS: Challenges and opportunities for systems integration at Finnish universities," 2014, poster at Open Repositories 2014.

[10] T. Walters, "Assimilating digital repositories into the active research process," in *Research Data Management – Practical Strategies for Information Professionals*, J. M. Ray, Ed. Purdue University Press, 2014.

[22] https://www.jyu.fi/hallinto/strategia/strategiat/strategian_toimenpideohjelma_en

[23] https://avointiede.fi/viitearkkitehtuuri

[24] https://jyx.jyu.fi/

[25] https://tutkijatunniste.fi/