Goethe University of Frankfurt Manages Spectroscopy Data with LOGS

Version 2

Challenge

The BMRZ[1] at the Goethe University Frankfurt lacked a unified repository for spectroscopy data, making it very difficult to find measurement files for validating publications or evaluating scientific theories. Important metadata such as experimental parameters were not available with measured data, and after scientists left their research group, the whereabouts of data files themselves were lost.

Solution

The BMRZ now manages spectroscopy data with LOGS[2], which automatically uploads data from the spectrometer, ensures easy data access, and helps comply with funding body regulations for long-term data retention. Metadata is saved and available centrally, keeping data reproducible and reusable.

The Center for Biomolecular Magnetic Resonance (BMRZ[1]) is an infrastructure facility for research involving high-end nuclear magnetic resonance (NMR) and electron paramagnetic resonance spectroscopy (EPR) at the Goethe University of Frankfurt. The research is dedicated to the elucidation of structure and functional mechanisms of biomolecules involving RNA and RNA-protein complexes, large soluble protein complexes and membrane proteins.

Managing research data from seven groups is a huge challenge

The Center's seven groups share about 20 spectrometers, including cutting-edge high field liquid and solid-state NMR and EPR spectrometers. Managing the research data generated by all of these groups is a huge challenge. "Everyone uses their own system to store spectroscopy research data," said Dr. Martin Haehnke, who handles IT administration and equipment monitoring in Prof. Dr. Schwalbe's group. "Researchers pull the data from the spectrometer to their desktops, but everyone uses different conventions. When people leave, we archive their hard disks in a box. We have a lot of boxes now, and it's really challenging to find the right data when you need it for a publication."

Dr. Burkhard Endeward, a senior scientist in Prof. Dr. Prisner's group, had a similar story to tell. He developed his

[1] Center for Biological Magnetic Resonance – http://www.bmrz.de
[2] LOGS data repository – https://logs-repository.com

own software for generating electronic documents from spectrometer data, but it was tough to get other people to use it because they couldn't address bugs on the fly.

"There are basically two ways to track your data: a paper notebook or electronic files. Saving the spectra is hard enough, and everyone has their own system, but there's often a huge gap describing the sample, experiment, and parameters," said Burkhard. "Worst case, you want to write a grant proposal and can't find the data, so you have to do it all over again."

> "If someone asks you in 10 years how you did your research, you want to be able to answer, not only to comply with funding regulations but also to protect your reputation."
>
> Dr. Burkhard Endeward, senior scientist in Prof. Dr. Prisner's group

Funding bodies require reusablity and reproducibility of data, with ten year data retention

The groups usually get a majority of their funding from DFG[3] grants and other funding bodies. For some time the DFG has been requiring that primary data as the basis for publications shall be securely stored for ten years in a durable form . More recently, the requirement bar was raised even more: Data should be made accessible at a stage of processing that allows it to be reused by third parties. This entails saving experimental parameters to ensure the reproducibility and reusability of data. Each grant application requires the principal investigator to sign a contract to this effect. Without a central repository for spectroscopy data, the noted requirements are virtually impossible to fulfill.

The university had recently experienced a public DFG inquiry into a different department, where the data couldn't be presented to satisfaction. "If someone asks you in 10 years how you did your research, you want to be able to answer, not only to comply with funding regulations but also to protect your reputation," said Burkhard.

Groups didn't want to waste time on a homegrown solution

Burkhard had been thinking about rebuilding his documentation system from scratch when he first heard about LOGS, SIGNALS's data repository solution for spectroscopists. "I wanted to add a search function to my software but I didn't want to waste weeks on coding. It just distracts from our research and then becomes very time consuming to look after. LOGS is a really good alternative that is a lot easier to administer and maintain."

Collaborating with SIGNALS[4], Burkhard gave LOGS developers feedback about native data formats, so that many fields would be automatically populated and searchable. "It really helped that SIGNALS's scientists understand spectroscopy so you don't have to explain why a certain data field is important," he said.

"When we looked at LOGS, it quickly became clear that we wouldn't want to build a solution ourselves," said Martin. "You don't realize how complex the problem is until you start really thinking about it. LOGS extracts spectral data and tracks samples, experiments, and parameters."

Automatic upload ensures spectra are centrally stored

One of the concerns was whether researchers would adopt the solution. To help promote usage, LOGS automatically uploads all spectra from the spectrometers. Users then just have to claim their data in the queue, which they can simply do from the web browser. "We no longer let people book new instrumentation time if they haven't claimed their data from the previous sessions," said Burkhard. "Researchers who don't want to use LOGS have to document their data retention another way and get a sign-off from the group leader, but only one person has chosen this path so far."

The auto-upload had another advantage, when Martin received a call late one evening. "A colleague thought he had lost his data because he had started another measurement without saving the spectra first, and the spectrometer had overwritten the previous data. He was very happy to hear that LOGS

[3] German Research Foundation (http://www.dfg.de/en)
[4] http://logs-repository.com

had already uploaded his dataset," said Martin. "Otherwise, this simple mistake could have thrown him back weeks. Protein samples degrade within a couple of days and he didn't have the instrument booked for another week. He would have had to go back to square one to create the sample."

Groups decide on their optimal workflow

Each group's LOGS admin can choose how to best mark up data within LOGS for the purpose of the team and provide guidance to the rest of the group. Martin's team uses tags to indicate sample name, nuclear isotope markers and solvent type.

Burkhard chose a slightly different path. He optimized an experimental workflow where he enters the sample number at the instrument before starting the measurement. This automatically connects the experiment with the correct sample entry when the data is uploaded to LOGS, saving a lot of time later. He also adds experimental parameters, e.g. the time window for PELDOR experiments, to a custom field in LOGS, which helps him when analyzing data. While he previously cross-referenced the lab book volume and page number with other notes, Burkhard now just notes down the unique LOGS ID.

> "Searching for spectroscopy data has become a lot easier. We have thousands of data sets. All the spin labeling is now well documented, and it's become easy to find the data I need."
>
> Dr. Burkhard Endeward, senior scientist in Prof. Dr. Prisner's group

Searching for spectroscopy data has become much easier

"Searching for spectroscopy data has become a lot easier. We have thousands of data sets. All the spin labeling is now well documented, and it's become easy to find the data I need," said Burkhard. "Now we simply write on the grant proposals that we are using LOGS for long-term data retention, and we no longer have to worry about an audit."

LOGS has come a long way since the BMRZ first started using it. "The people from SIGNALS are very open to feedback, and they added things like experimental parameters and the correlation of samples and spectra based on our request," said Burkhard.

Going forward, Martin would like to lock down remote sFTP or SSH access to the spectrometers altogether, for the scientists carrying out measurements, so that data is claimed only through LOGS, by default. Scientists can continue downloading their measurements from LOGS, just like they used to from the spectrometer, and save it and proceed with it as they are used to. But additionally this workflow now consistently ensures that data is recorded and available in an extra layer of security, provided by LOGS.

Signals GmbH & Co. KG
Graf-von-Stauffenberg-Allee 83
60438 Frankfurt am Main
Germany

+49 (0)69 27147 442
info@signals.company
http://signals.company
http://logs-repository.com

SIGNALS