ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-MO2AO01

FACING THE CHALLENGES OF EXPERIMENT CONTROL AND DATA MANAGEMENT AT ESRF-EBS

J. Meyer, S. Debionne, W. De Nolf, S. Fisher, A. Götz,
M. Guijarro, P. Guillou, A. Homs Puron, V. Valls
ESRF, Grenoble, France

Abstract

In 2020, the new European Synchrotron Radiation Facility Extremely Brilliant Source (ESRF-EBS) took-up operation. With the much higher photon flux, experiments are faster and produce more data. To meet these challenges, a complete revision of data acquisition, management and analysis tools was undertaken. The result is a suite of advanced software tools, deployed today on more than 30 beamlines. The main packages are Beamline Instrument Support Software (BLISS) for experiment control and data acquisition, Library for Image Acquisition 2 (LIMA2) for high-speed detector control, the ESRF Workflow System (EWOKS) for data reduction and analysis workflows, and Daiquiri the web GUI framework. These developments provide a solid foundation to tackle current and future challenges at the ESRF and to continue providing cutting-edge capabilities for synchrotron research.

BLISS is programmed in Python, to allow easy sequence programming for scientists and easy integration of scientific software. BLISS offers: Configuration of hardware and experimental set-ups, a generic scanning engine for step based and continuous data acquisition, live data display, frameworks to handle 1D and 2D detectors, spectrometers, monochromators, diffractometers and regulation loops.

For detectors producing very high data rates, data reduction at the source is important. LIMA2 allows parallel data processing to add the necessary compute power (CPU and GPU) for online data reduction in a flexible way.

The EWOKS workflow system can use online or off-line data to automate data reduction or analysis. Algorithms are wrapped as a workflow module and can be combined with other processing modules. Workflows can run locally or on a compute cluster, using CPUs or GPUs. Results are saved or fed back to the control system for display or to adapt the next data acquisition.

ESRF-EBS

The ESRF-EBS is the first fourth generation high-energy synchrotron with a Hybrid Multi-Bend Achromat (HMBA) lattice [1]. The new X-ray beam is 100 times more brilliant and coherent than before, Fig. 1. For the construction the ESRF shut-down in December 2018. The new storage ring was built within the existing infrastructure and reduced the energy consumption by 30%. In September 2020 the new synchrotron took-up operation. With its much higher photon flux, experiments can be conducted faster and more in-situ experiments are carried-out.

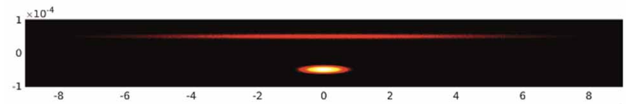

Figure 1: View of the X-ray beam before (top) and with EBS (bottom).

THE CHALLENGES

To cope with the new specifications of faster data acquisition, higher data rates and the increasing need for online data reduction, a large part of the ESRF software for beamlines had to be modernized after 30 years of continuous operation.

The main challenges to be addressed are:

- Data acquisition rates in the kHz and up to the MHz range.
- To limit the infrastructure cost and impact on the environment, the raw data to be stored must be reduced as early as possible during experiments.
- Standardization of all experimental raw and meta data for generic visualization, processing, automation and data curation.
- Automated online data analysis for immediate feedback of experiment results and avoid storing useless raw data.
- Experiment registration with all raw data, meta data and processed data produced. Remote users should have immediate access to experiment results.

The EBS shutdown was a unique opportunity to re-write the main data acquisition software, prepare for faster detectors with higher data rates, develop a workflow system for online data reduction and data analysis and to modernize the GUI framework

A clean management of the data flow is required on all levels. All the different software tools must work seamlessly together (see Fig. 2). All data visualization, data saving, data analysis and experiment registration depend on the data produced by the acquisition system.

DATA ACQUISITION

BLISS: Scanning and Sequencing

The replacement campaign of the old SPEC-based data acquisition system, was started during the EBS shutdown in 2018 to the situation today, where 35 beamlines are running with BLISS.

19th Int. Conf. Accel. Large Exp. Phys. Control Syst. ICALEPCS2023, Cape Town, South Africa JACoW Publishing
ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-MO2AO01

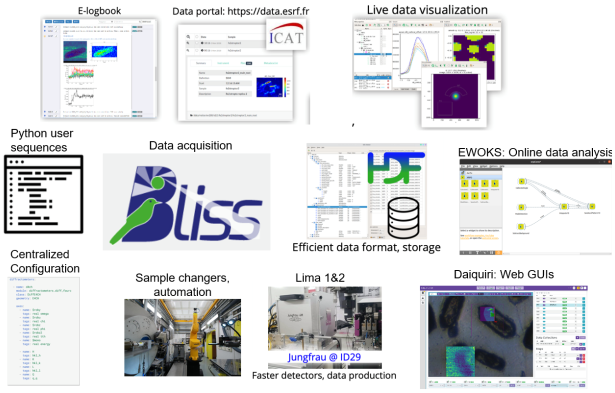

Figure 2: Software Integration.

BLISS [2] is a command line driven sequencer written in Python. As most scientific software is written in Python too nowadays (or can be called from Python), scientists easily integrate their software into the acquisition sequences.

The main points of the BLISS architecture (Fig. 3) are:

- A **generic scan engine** for step and continuous scans. The same concepts for any data acquisition and coherent data management. The use of trajectories and HKL space [3] is possible with all scans.
- **Decoupling of data acquisition from data saving and analysis**. Allows higher acquisition speed without blocking. Any data consumer uses the same API.
- **Easy configuration of hardware and experimental environment**. YAML configuration files for easy editing and maintenance. Measurement groups to switch between predefined acquisitions set-ups on the fly.
- **Coherent storage of all acquired data** at high speed and for large data volumes. All data of a dataset, sample or proposal is saved as a HDF5 data tree that fits the ESRF data policy.
- **Live data display** of all acquired data, Fig. 4. Immediate visibility of acquisition results for the user.

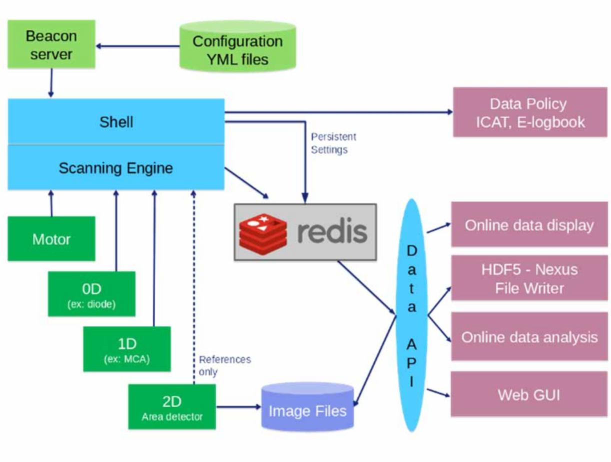

Figure 3: BLISS Architecture.

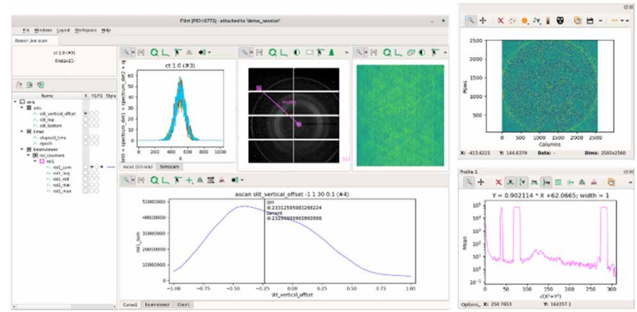

Figure 4: Online Data Display (Flint).

BLISS is designed to control synchrotron beamlines and includes several internal hardware abstraction frameworks to standardize the control of complex instruments, Fig. 5.

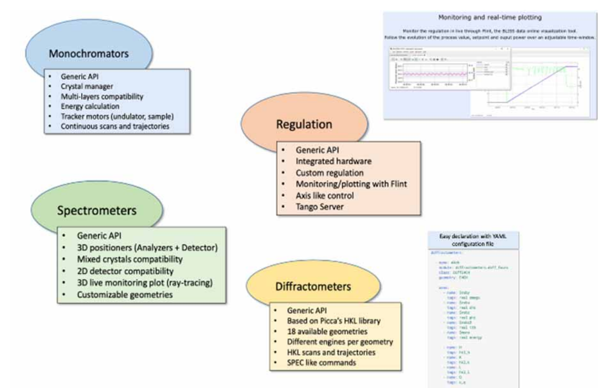

Figure 5: Bliss Internal Frameworks.

The data acquisition with high speed 1D or 2D detectors acquiring large data volumes is implemented in external frameworks, MOSCA [4] for 1D and LIMA and LIMA2 [5] for 2D detectors. Both frameworks are accessed via Tango servers with a standard API for all detectors of the same type. The corresponding BLISS controllers need to be implemented only once.

LIMA2: Distributed Acquisition for High Performance 2D Detectors

Since the start of the ESRF-EBS operation, the number and variety of high speed detectors has increased significantly. With acquisition speeds in the kHz range and raw data volumes reaching 10s of GB/s, the sequential acquisition framework for 2D detectors (LIMA) needed to be upgraded.

The new framework, LIMA2, implements several topologies to distribute acquired images for processing on multiple computers (see Fig. 6). The distributed architecture supports multiple receivers, ensuring scalability for very high data rates and low latency online data analysis. Early on the fly image processing (CPU or GPU) on the frontline data acquisition computers allows:

- Online data reduction for large data volumes
- Closed loop applications with feedback to the experiment
- Live visualization of results

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-MO2AO01

The new LIMA2 framework decouples the data acquisition and processing life cycle. A simplified single-command configuration and a high-level API hide the distributed process topology to all clients e.g. BLISS and MXCuBe [6] etc.

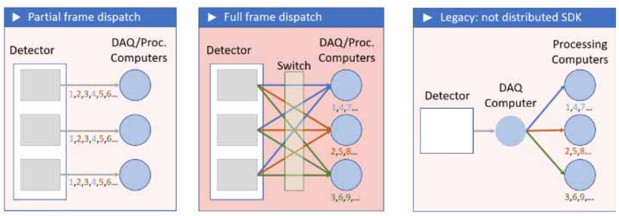

Figure 6: LIMA2 Data Distribution Topologies.

Figure 7 shows an example of an online data reduction pipeline for serial crystallography on the ID29 beamline. The detector is a 4M PSI-Jungfrau detector running at 1 kHz, producing 8 GB/s of raw data. The aim of the online reduced sparse data flow is to save only 400 MB/s i.e. 20 times less.

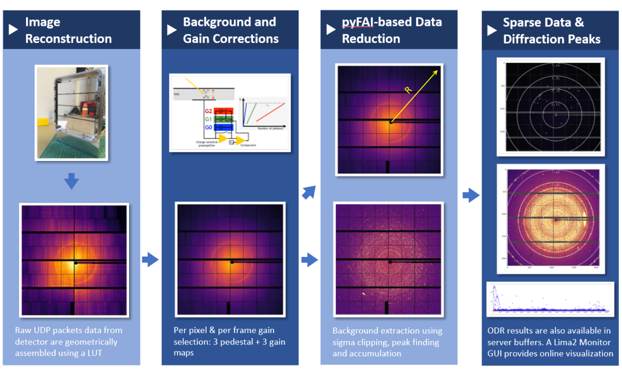

Figure 7: LIMA2 Online Data Reduction Example.

ONLINE DATA ANALYSIS

BLISS Data API

As already mentioned in the section about BLISS, data acquisition and data saving and analysis are decoupled. All access to the acquired data happens via the BLISS Data API, Fig. 8.

The API is separated into two parts. The low level streaming API offers direct access to all incoming data. The BLISS file writer (Nexus Writer) and the BLISS live data display (Flint) are clients of this API.

Because the lifetime of data stored in the Redis [7] in-memory database is limited, a high level API has been implemented to read live data as if it would be read from an HDF5 file. If the data is no longer available in memory, the API automatically reads from the HDF5 data file stored on disk.

Data analysis clients that follow the speed of the data acquisition can do analysis in real time. Clients which are slow, will also continue to run in the same way, but will switch to file access when data is no longer available in Redis or the LIMA image buffer.

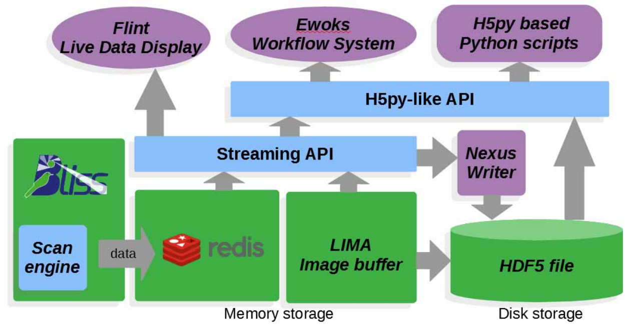

Figure 8: Bliss Data API.

EWOKS: ESRF Workflow System

Today the data acquisition system is able to cover the needs of ESRF-EBS in the kHz range and higher. But with experiments executed in hours instead of days and much larger data volumes produced, scientists are faced with problems transferring and analyzing the data in their home institutes. Online data reduction and analysis become a necessity to reduce the time between measurement and scientific publication or to make scientifically relevant decisions during the experiment.

The ESRF workflow system EWOKS [8] was designed to execute tasks in a fully automated manner when the data acquisition is running and when manually triggered by the user. Results can be stored as processed data of the experiment or can be fed back to the data acquisition system for display or to control the next acquisition.

Today, 25 beamlines are running workflows to process their data in different scientific domains like tomography, spectroscopy, fluorescence, SAXS/WAXS, dark field microscopy, diffraction and structural biology.

Workflows can be described as the flexible chaining of processing tasks, Fig. 9. The processing code is written by scientists and integrated to the workflow system by application experts. Afterwards the user can chain the different processing tasks graphically.

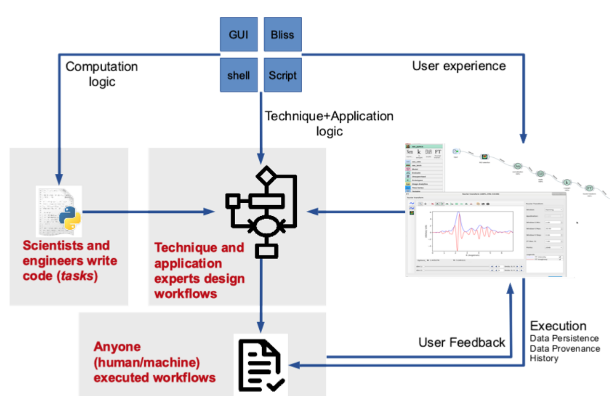

Figure 9: Workflow Definition.

The acquisition system can trigger any EWOKS workflow and survey its execution status (see Fig. 10). Complex results are stored in Redis and can be used for display or any further action.

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-MO2AO01

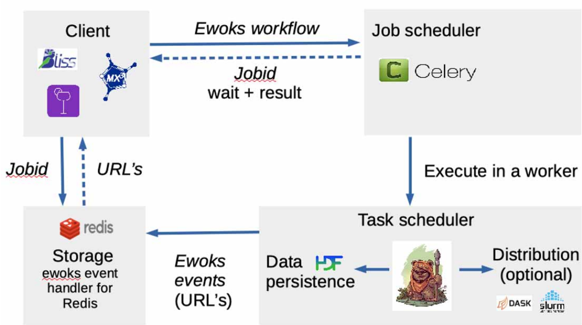

Figure 10: Automated Workflow Execution.

Examples of workflows in operation are the online tomography volume reconstruction (Fig. 11) or the EXAFS online visualization in scientific parameter space (Fig. 12).

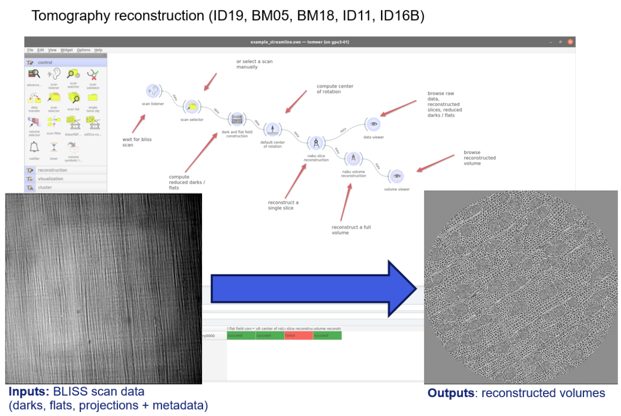

Figure 11: Tomography online volume reconstruction.

Both workflows are used as immediate result feedback for decision taking and as first results for further analysis. Thus speeding-up the data validation and data analysis process for the beamline scientists.

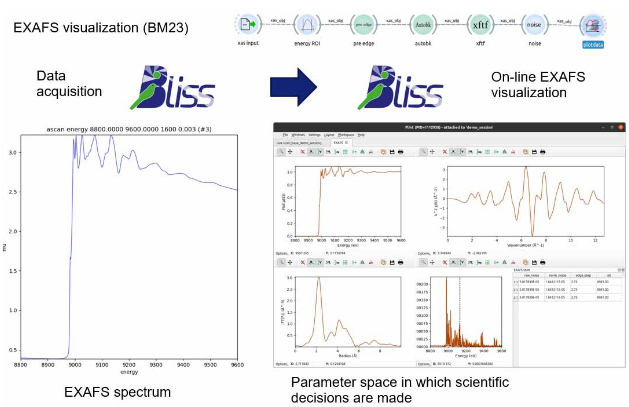

Figure 12: EXAFS Plotting During Acquisition.

EXPERIMENT GUIS

Daiquiri: Web GUI Framework

The first web GUI for control of protein crystallography experiments at the ESRF was MXCuBe3 [6] in 2017. Based on the good experience and the strong trend of developing towards graphical user interfaces as web applications, the decision was made to invest in web development.

The daiquiri framework [9] was designed to work on top of BLISS for control and monitoring but can also handle EWOKS tasks and access Tango servers, Fig. 13.

The access to BLISS is based on the BLISS REST API and a web terminal offering the possibility to integrate a command line interface to BLISS as part of a web GUI.

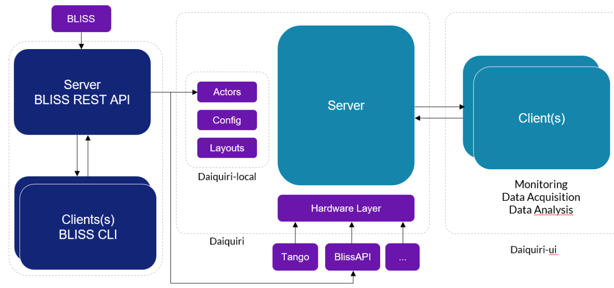

Figure 13: Daiquiri Architecture.

A good example of a complex experiment GUI, is the tomography GUI (see Fig. 14). It offers the user a view of the sample, the sinogram during the data acquisition and a reconstructed slice for data validation. Slice reconstruction is executed by an EWOKS workflow. The rotation axis can be modified directly on the reconstructed slice for the next data acquisition.

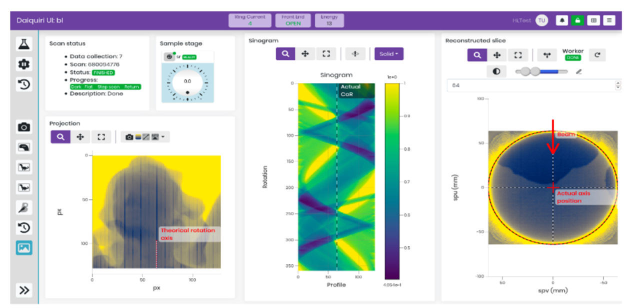

Figure 14: Tomography GUI.

Another example is the GUI for milli X-Ray Fluorescence (mXRF) for the quantification of element concentrations in samples (see Fig. 15). Acquired mXRF maps are superimposed on the visible-light image of the sample. On the bottom left mXRF map, a smaller map was selected and acquired with higher resolution. On the top right mXRF map, a series of POIs have been selected for the collection of mXAS spectra.

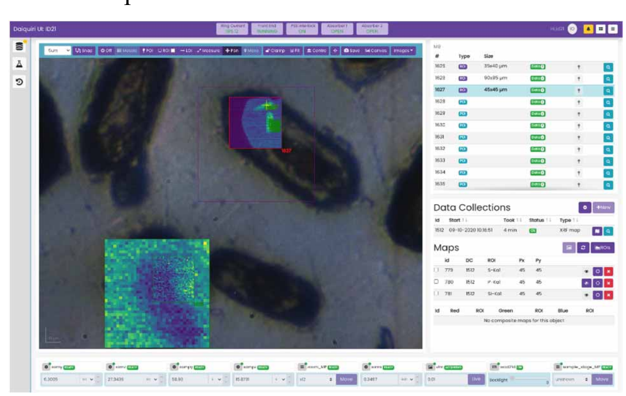

Figure 15: mXRF Mapping GUI.

CONCLUSION

With BLISS and LIMA2 we have a flexible and performant data acquisition system. We are able to integrate high performance detector systems, further speed-up experiments and implement on-the fly data reduction and low latency processing.

The standardization of data APIs and standardized data and meta data in HDF5 produced by the acquisition system allow automation for visualization, saving and further processing by the EWOKS workflow system or any other processing tools.

The definition and evolution of EWOKS workflows for different science domains will allow more and more beamlines to visualize processed data from experiment on the fly or to reduce the raw data storage. Veto systems to avoid storage of useless raw data or to feedback results to the running acquisition become very easy to implement.

With BLISS, LIMA2, EWOKS and Tango data reduction and analysis are possible at all possible acquisition speeds. The door is wide open for complex experiments.

For example, a user group working on X-ray reflectometry on the ID10 beamline recently used all the tools described above to profit from AI/ML algorithms to close the loop during the experiment [10] (see Fig. 16).

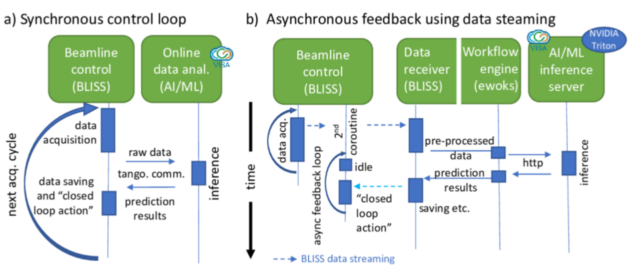

Figure 16: Autonomous experiments enabled by machine-learning-based online data analysis (diagram from [10] with permission of L.Pithan).

The Daiquiri web GUI framework opens the door not only to remote operation, but also to a closer integration of acquisition with other tools developed for experiment management. Those tools, for example electronic logbooks or the ESRF data portal, are all implemented as web applications and can run in the same browser or even in the same Swindow.

If you want to try and play, a BLISS demo session is available under [11].

ACKNOWLEDGEMENTS

Many thanks to the members of the ESRF software group for the development of all the different software tools and to all the beamline scientists for their feedback, patience and bug reports since the EBS start-up.

REFERENCES

[1] P. Raimondi *et al.,* "Commissioning of the hybrid multibend achromat lattice at the European Synchrotron Radiation Facility", *Phys. Rev. Accel. Beams,* vol. 24, p. 110701, Nov. 2021. doi:10.1103/PhysRevAccelBeams.24.110701

[2] BLISS, https://bliss.gitlab-pages.esrf.fr/bliss/master

[3] HKL space, https://en.wikipedia.org/wiki/Miller_index

[4] MOSCA, https://gitlab.esrf.fr/bcu/mosca/mosca.git

[5] LIMA2, https://limagroup.gitlabpages.esrf.fr/lima2

[6] MXCuBE3, https://mxcube3.esrf.fr/

[7] Redis, https://redis.com

[8] EWOKS, doi:10.5281/zenodo.6075054

[9] S. Fisher, M. Oscarsson, W. De Nolf, M. Cotte, and J. Meyer, "Daiquiri, a web-based user interface framework for beamline control and data acquisition", *J. Synchrotron Radiat.*, vol. *28, no. 6, pp. 1996-2002,* Oct . 2021. , doi:10.1107/S1600577521009851

[10] Linus Pithan *et al.*, "Closing the loop: Autonomous experiments enabled by machine-learning-based online data analysis in synchrotron beamline environments", *J. Synchrotron Radiat.* accepted Aug. 2023. doi:10.1107/S160057752300749X

[11] BLISS demo, https://bliss.gitlab-pages.esrf.fr/bliss/master/bliss_demo.html