DMA ST1
seminar series
# Yves Kemp / Kilian Schwarz

DMA ST1 – seminar series

- Requirement analysis and review of existing tools via DMA ST1 seminar series

- Selected highlights from the seminar series so far will be presented in the following slides

Seminar presentations so far

- Kilian Schwarz, GSI: XrootD - an overview 2020-11-30

- Tigran Mkrtchyan, DESY: dCache - Scientific Storage 2020-12-14

- Sergey Yakubov, DESY: Managing metadata for Photon Science experiments 2021-01-11

- Thomas Gruber, HZDR: How can an ELN help to manage metadata of experiments? 2021-01-25

- Oliver Knodel, HZDR: The Future Metadata Management of our Research Project Lifecycles at HZDR 2021-02-08

- Jürgen Hannappel, DESY: Status and overview on data acquisition and analysis at Petra-III and FLASH at DESY 2021-05-31 Meeting notes

- Martin Gasthuber, DESY: ASAPO - new system to support stream based data analysis for photon science experiments, September 05, 22

- https://confluence.desy.de/pages/viewpage.action?pageId=183150136
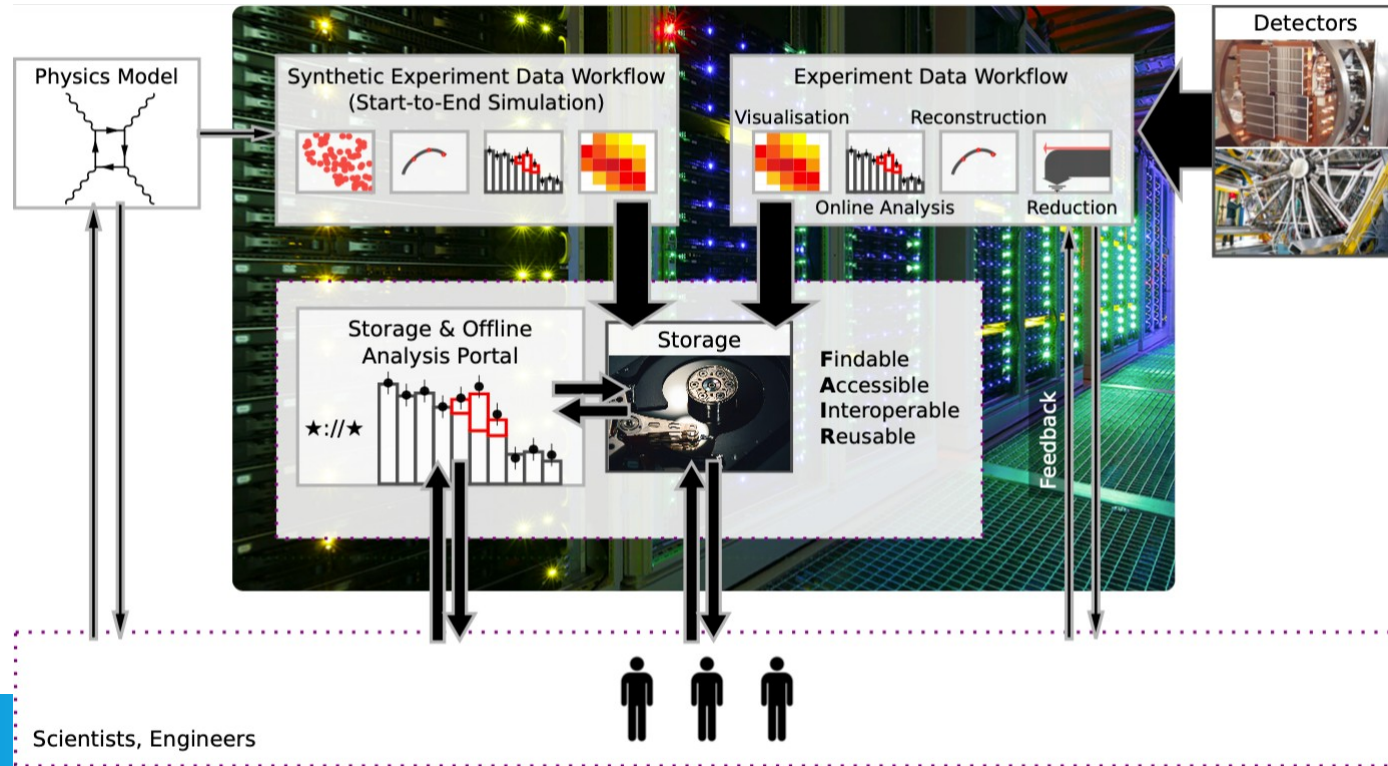
Seminar topics covered and missing

- Research data management, WAN transfer, lab notebooks

- Data acquisition, data analysis, metadata management

-

- Online analysis & data reduction, analysis frameworks

- Cluster and dynamic/on demand file systems, dynamic caches

- Portals, Jupyter, AAI

- Detector simulation

- ML/AI methods

- Usage of HPC centres, batch systems, Grid/Cloud

- Metadata standards
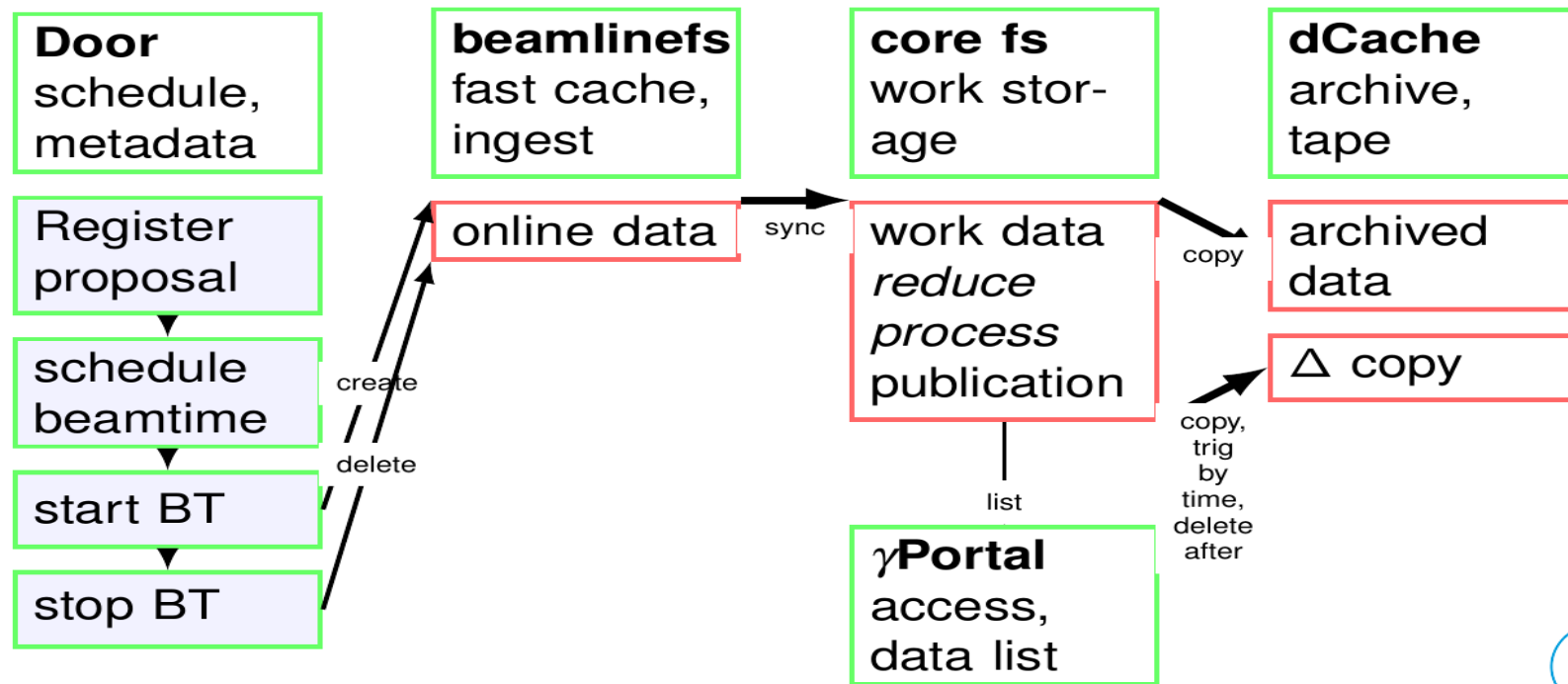
J. HannappelS. Dietrich, M. Gasthuber, S. Yakubov et al.   Status and overview on data acquisition and a   May 31, 2021   4 / 12

ASAPO (High Performance Online Data Analysis) – the middleware

> middleware for high-performance next-generation detector data analysis

- Provides API to inject data to the system - e.g takes care of the "first mile" between the experimental hall and the compute center (high-performance data transfer)

- Provides API to retrieve data from the system - e.g. for data analysis synchronous (online) and asynchronous (offline) to data taking

> Basic characteristics

- Scalable (N sources, K network links, L service nodes, M analysis nodes)

- Highly available (services in Docker containers managed by Nomad/Consul)

- Efficient (C++, multi-threading, RDMA, …)

- Provides user friendly API interfaces (C/C++, Python, REST API)

- Runs on Linux/Windows/…
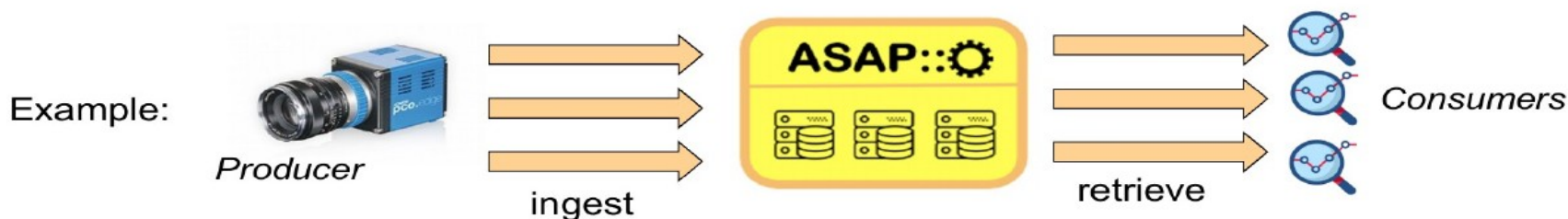
Example:

Producer   ingest   ASAP::⚙   retrieve   Consumers

dCache project

dCache.org
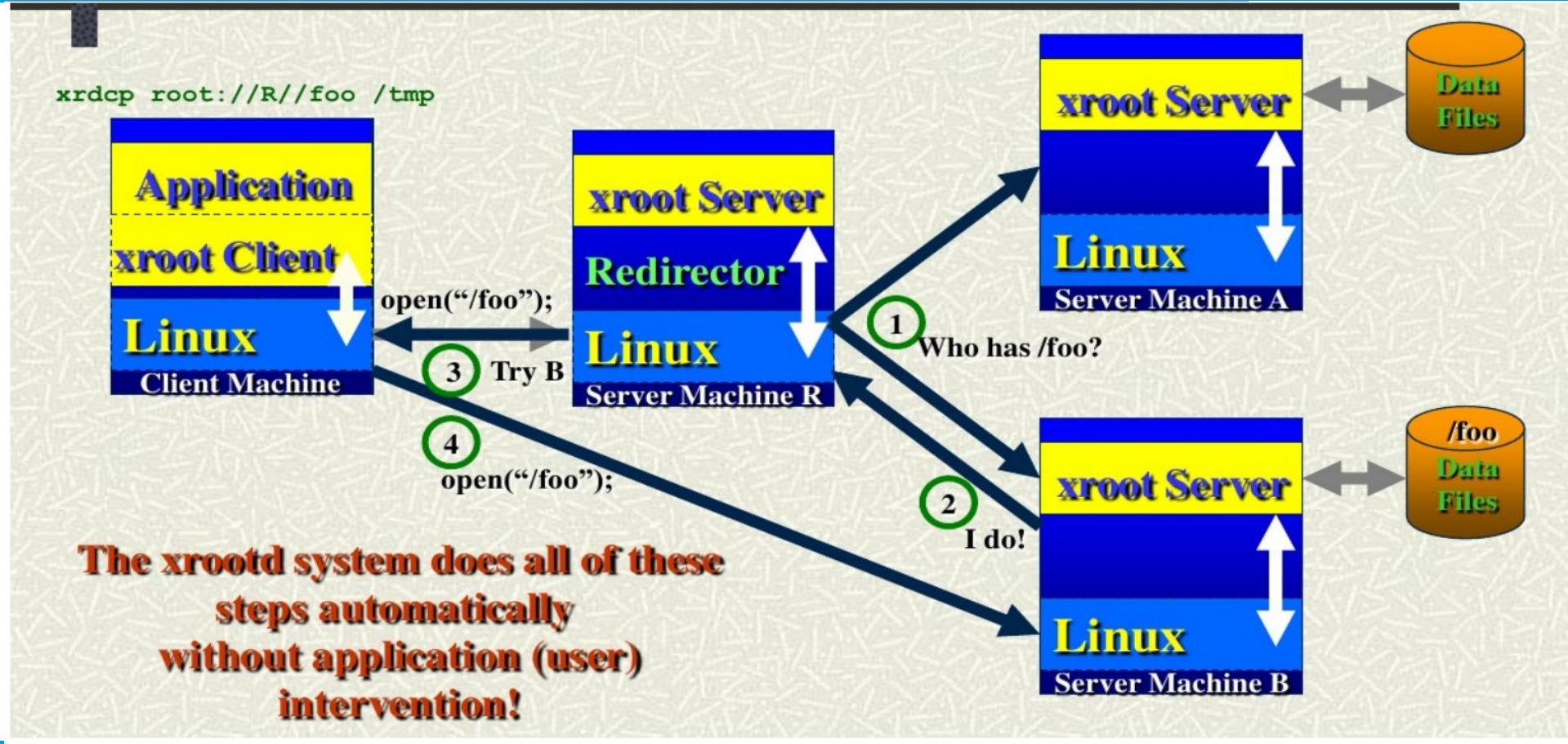
- Joined effort between DESY(2000), FNAL(2001) and NDGF(2006)

- Provides storage solution for scientific data

- Supports standard and HEP specific access protocols and specific authentication mechanisms

- Developed for HERA and Tevatron, used for LHC and others
  - WLCG, Belle II, LOFAR, CTA, IceCUBE, EU-XFEL, Petra3, DUNE, And many more ...

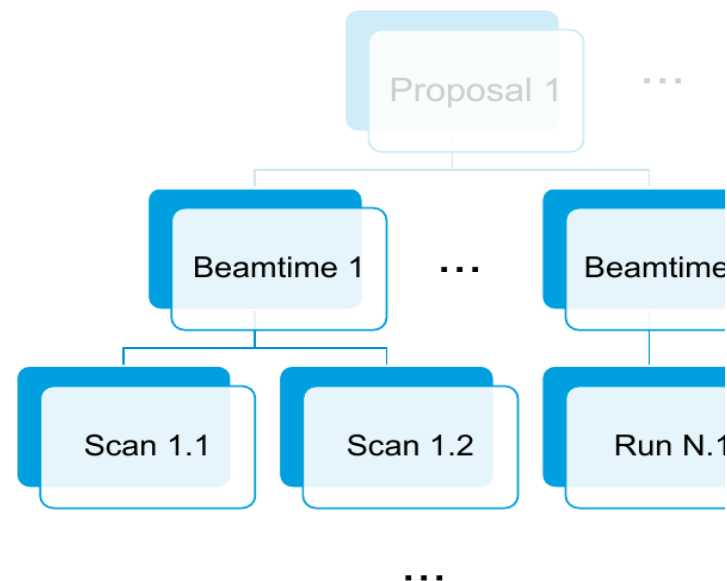Data management/dCache

ASAPM

Overview

- ASAP Metadata Service – provides a centralized and standardized way to store experiment-related information (metadata)
  - administrative / technical / scientific / custom in key-value form
  - logbook entries in free text form
- This information is stored in metadata collections
  - e.g. beamtime metadata
  - each collection entry can have subcollections, e.g. beamtime meta -> scan meta -> ...
- everything can be done via Web API
- Various modes to work with metadata in Web UI
  - collections view
  - beamtime view (with status)
  - logbooks view
  - other views can be implemented later (e.g. proposals, samples, ...)

DESY.

DMA ST1 – seminar series Metadata management

HZDR Data Management Strategy   **HELIPORT**
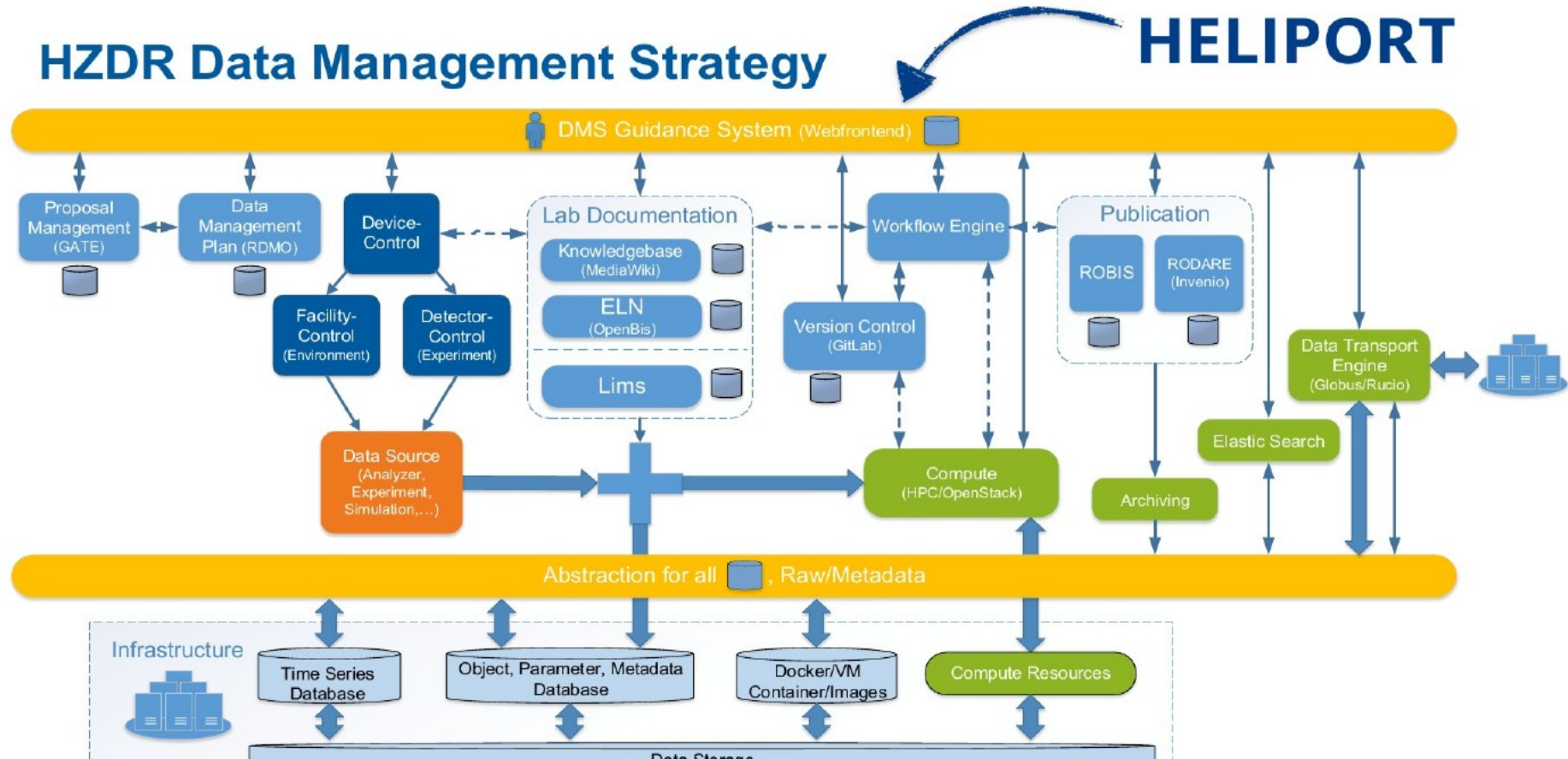
DMA ST1 – seminar series Electronic Lab Books

OpenBIS

DRESDEN concept   **HZDR**

Homepage: https://openbis.ch/

openBIS

Video tutorial, documentation and link to demo server

INVENTORY MANAGEMENT
Whatever your research field, keep track of all materials and samples used in your lab. Create experimental or computational procedures to share with all lab members.

LAB NOTEBOOK
Describe your wet-lab or computational experiments and link to the materials, samples and protocols stored in the inventory.

DATA MANAGEMENT
Make your data FAIR. Store all data connected to your experiments, of any size and format, via the web interface or using background upload scripts, depending on the data size.

BIGDATALINK
Use openBIS as metadata repository and use our git-like tool to link huge datasets across different locations.

INTEGRATIONS FOR DATA ANALYSIS
Use Jupyter notebook to analyse your data and store your notebooks in openBIS. Use the APIs to integrate openBIS in your workflow manager to extract data for large-scale analysis on a computer cluster and upload results back to openBIS.

IMPORT/EXPORT
Excel-compatible import and export formats for all tables. Export of lab notebook to text and word files.

ACCESS AND RIGHTS MANAGEMENT
Full control of user profile and access privileges.

AUDIT TRAIL
Keep track of all modifications entered in the database.

MODULARITY
Write your own plugins to extend existing functionalities or add new ones.