

OSIRI:

Proposal for Open Data release of part of ZEUS data

What is OSIRI? (Osiris = god of the dead, the underworld, and **rebirth**)

Open Science Incubator for Research and Innovation

pilot proposal to be submitted to EU/ATTRACT call, total of
~1 FTE-year in phase 1, much more in phases 2 and 3 if successful



~half (1/2 FTE-year) “for ZEUS data” in phase 1 as pilot for a more
extensive project also involving other experiments

- **propose to release 2005 data (only), and 2005 inclusive DIS MC (only)**
- **propose to reformat ZEUS data to be compliant with data format of an ongoing experiment** for which Open Data exist (CMS nanoAOD format)
 - > will allow to do similar analyses on Open Data sets of all participating experiments **with the same analysis setup**

Previous experience with Open Data

Releasing data as Open Data means giving up both control over their use by outsiders and responsibility for results produced by such outsiders.

Experience with CMS public data since 1st release in 2014 (meanwhile half of original CMS Run 1 data + relevant MCs have been released):

- **so far 2 physics papers published** (on jet substructure, by group of theorists at MIT, in Phys. Rev. and Phys. Rev. Lett.).
- **more coming.**
- also usage for **machine learning** (e.g. Yandex), **event display algorithms**, and for **educational purposes**.
- **no problems with “crackpot” claims occurred** (as some had feared).
- voluntary support to users provided via `opendatasupport@cern.ch`
- usually, it is in the self-interest of outsiders to interact with some members of CMS on an ad-hoc basis.
- **responsibility for publication lies in the hands of authors and journal referees only.**

Special rules for collaboration members

Releasing data as Open Data means giving up both control over their use by outsiders and responsibility for results produced by such outsiders.

Example: CMS data preservation, reuse, and Open Data policy:
<http://opendata.cern.ch/record/411>

CMS members have to ask the collaboration for permission before becoming an author of a non-CMS paper from Open CMS data.

(Internal CMS document linked to Open Data portal for CMS members)

-> could decide on a similar rule within ZEUS

Rules for ZEUS Open Data use by ZEUS collaboration members remain entirely under control of the ZEUS collaboration!

Proposal for ZEUS Open data in context of OSIRI

release 2005 data (no STT -> simpler, most published ZEUS dataset)

+ inclusive DIS MC, latest CN version only (v8) -> few Tb

make available through Open Data portal at CERN and/or DESY (involve -L-, -IT-)

provide detailed public description of content and metadata for 2005 data on release portal: data format, calibration, DIS triggers, luminosity, etc.

-> needs ~half FTE-year of expert (ex-ZEUS member) person power
(full data + MC would need ~2 FTE-years -> outside of current financial scope)

-> funding request via EU/ATTRACT project in context of wider scope

known example use cases include (+ presumably plenty of as yet unknown ones):

- prestudies by EIC community
- limits on Lorentz invariance violation from inclusive DIS

In case of success, people can still contact us (ZEUS collaboration) 'as usual' to **get access to full data and MC within and under the control of the ZEUS collaboration**

Request and Question to ZEUS:

Kindly request ZEUS to agree to Open Data release of part of ZEUS data (~20-25%, e.g. 2005 data) in case the EU/ATTRACT proposal gets funded

close of EU call: October 15; internal preselection at CERN (involvement of CERN SIS) after September 15, no formal internal deadline at DESY

In the data release process, ZEUS would in any case be

- involved in the process of the final selection of the MC and data subset(s) to be released
- consulted concerning the content of and form in which the metadata would be provided on the portal

Would ZEUS as a collaboration wish to be involved as an official partner of the project (via DESY)? (rather than “consultant only”)

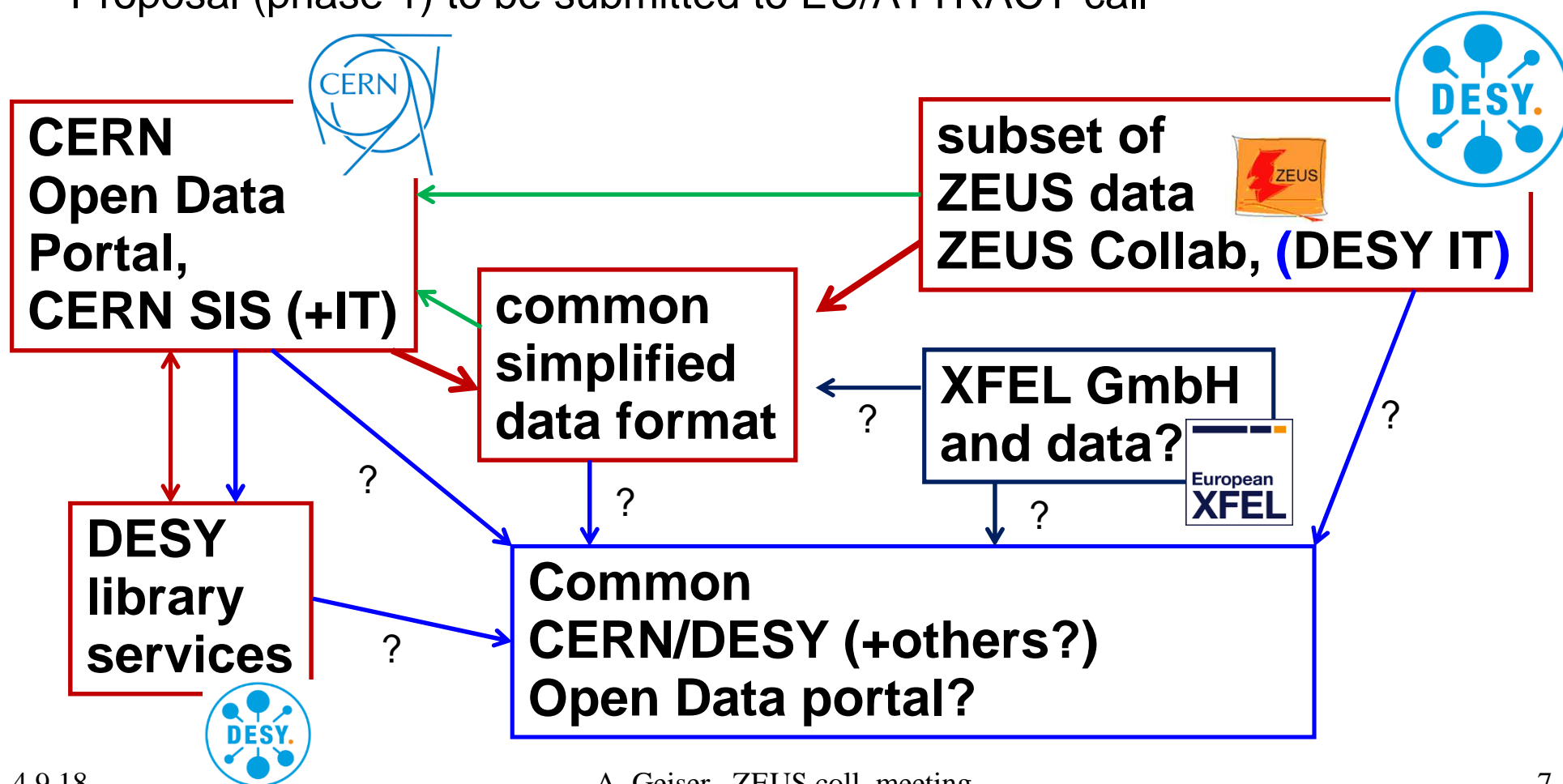
-> would ensure more direct influence on the work done, paired with an expectation that at least a bit of additional internal (volunteer) person power would be provided

Backup

Potential structure of the OSIRI project

Open Science Incubator for Research and Innovation

Proposal (phase 1) to be submitted to EU/ATTRACT call



Deadlines and Status

www.attract-eu.org

(CERN and XFEL GmbH are program coordinators)

17 M Euros to fund 170 `disruptive' projects (phase 1), i.e. 100k Euros/project

Significantly more funds/project for subset of these planned for phase 2

Call opened: Aug. 1, 2018

ZEUS collaboration meeting (approval?) Sept. 4, 2018 (tbc)

Deadline for internal CERN submission
for all projects involving CERN Sept. 17, 2018
(no internal deadline at DESY)

Close of EU call: Oct. 31, 2018

Start/duration of project (if accepted): ~ April 2019, 1 year

Thoughts about simplified data format

