

JCDL Workshop

Dr. Heinrich Stamerjohanns
Institute for Science Networking
at the
University of Oldenburg



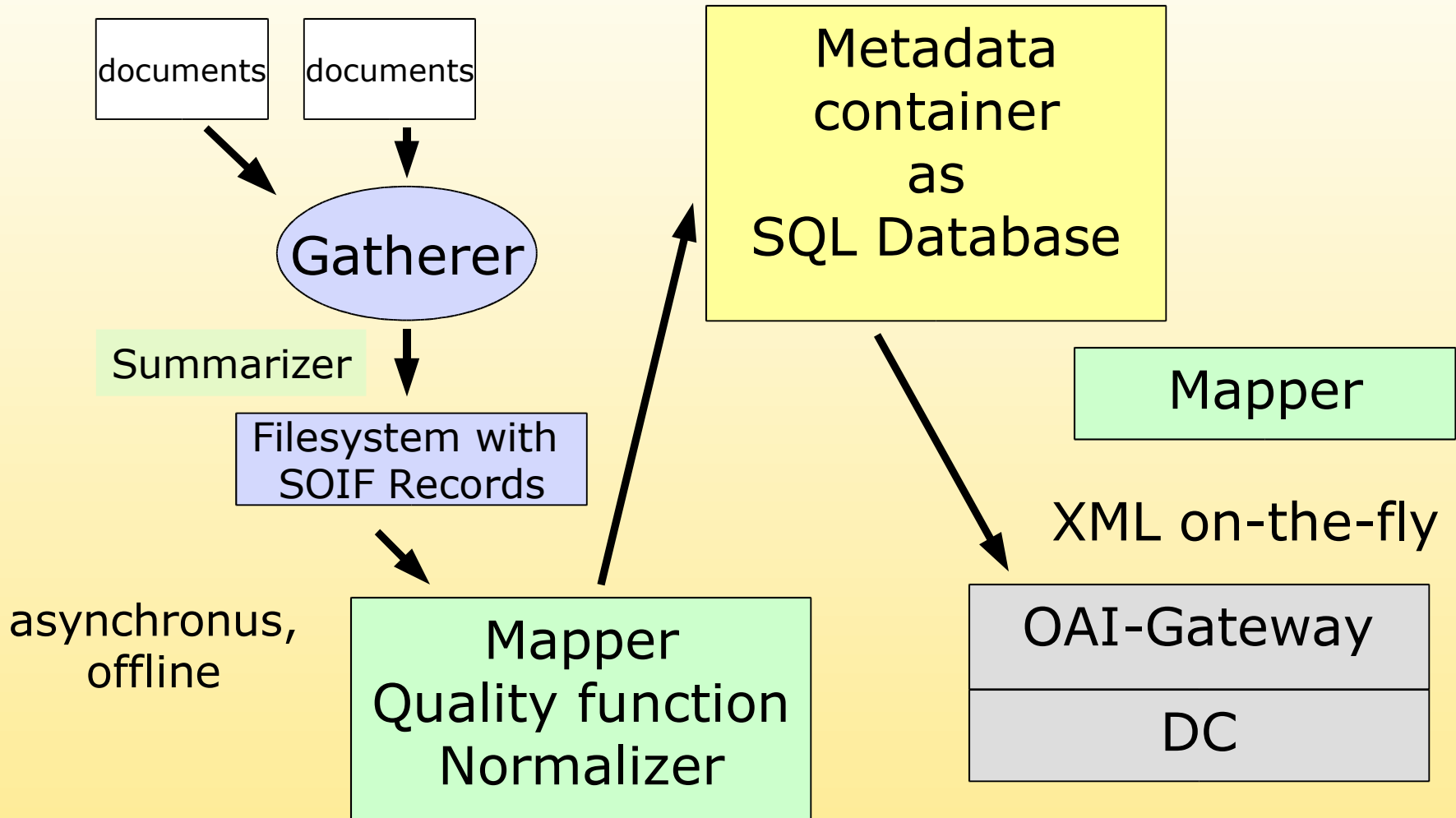
Institute for
Science Networking

Heinrich Stamerjohanns

Integrating self-archived articles into our OAI Data Provider

- self-archiving (on web servers)
- crawl, harvest articles by Harvester
- make searchable through unified search interface
- try to extract metadata and/or extend unstructured data by metadata
- approaches taken by Harvest, mnogosearch

PhysDoc as OAI Data-Provider

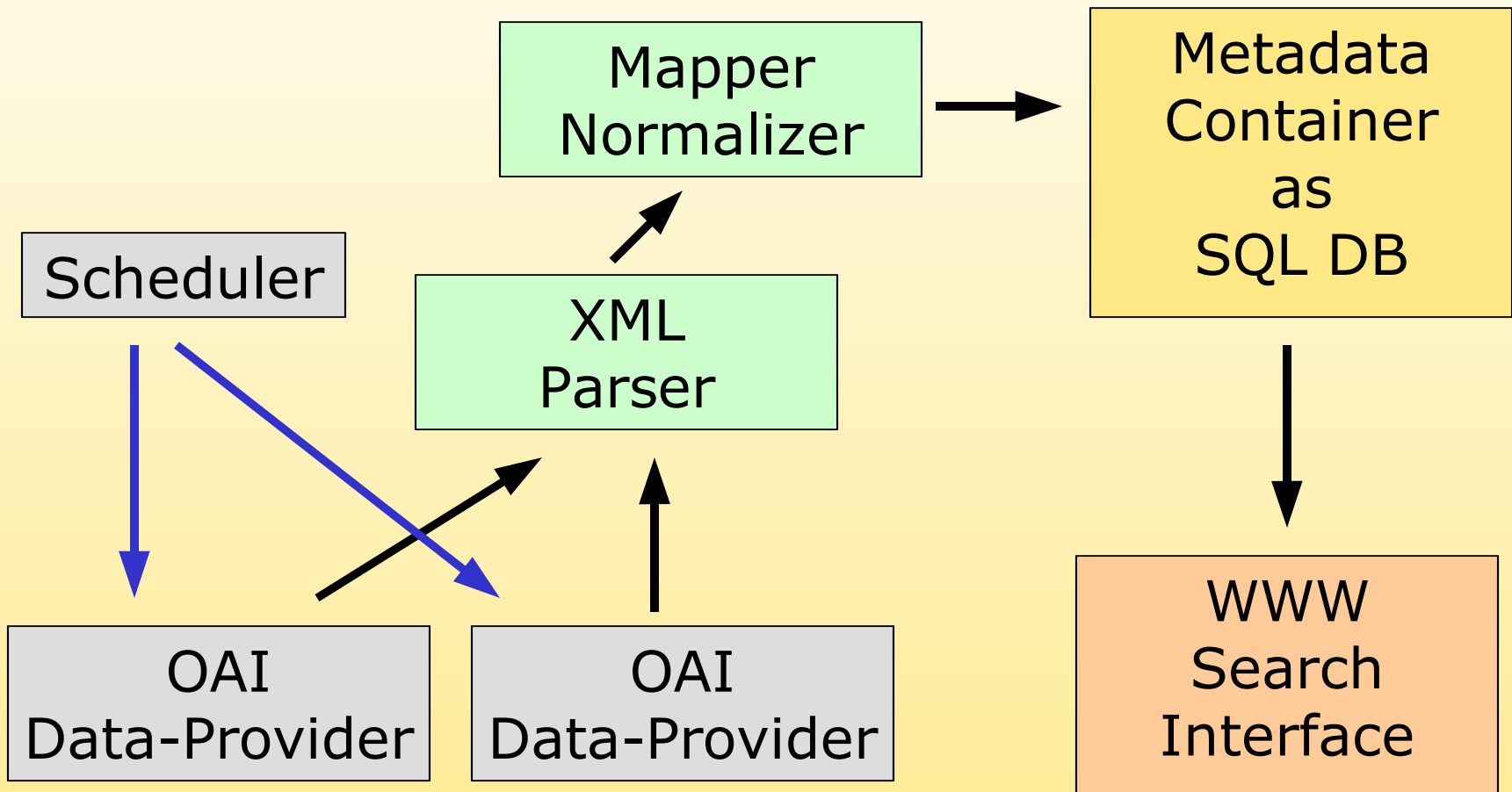


OAD

PhysDoc as Service-Provider

- articles are collected through OAI from various OAI Data-Providers
- other publishers are and will be incorporated through proprietary interfaces.

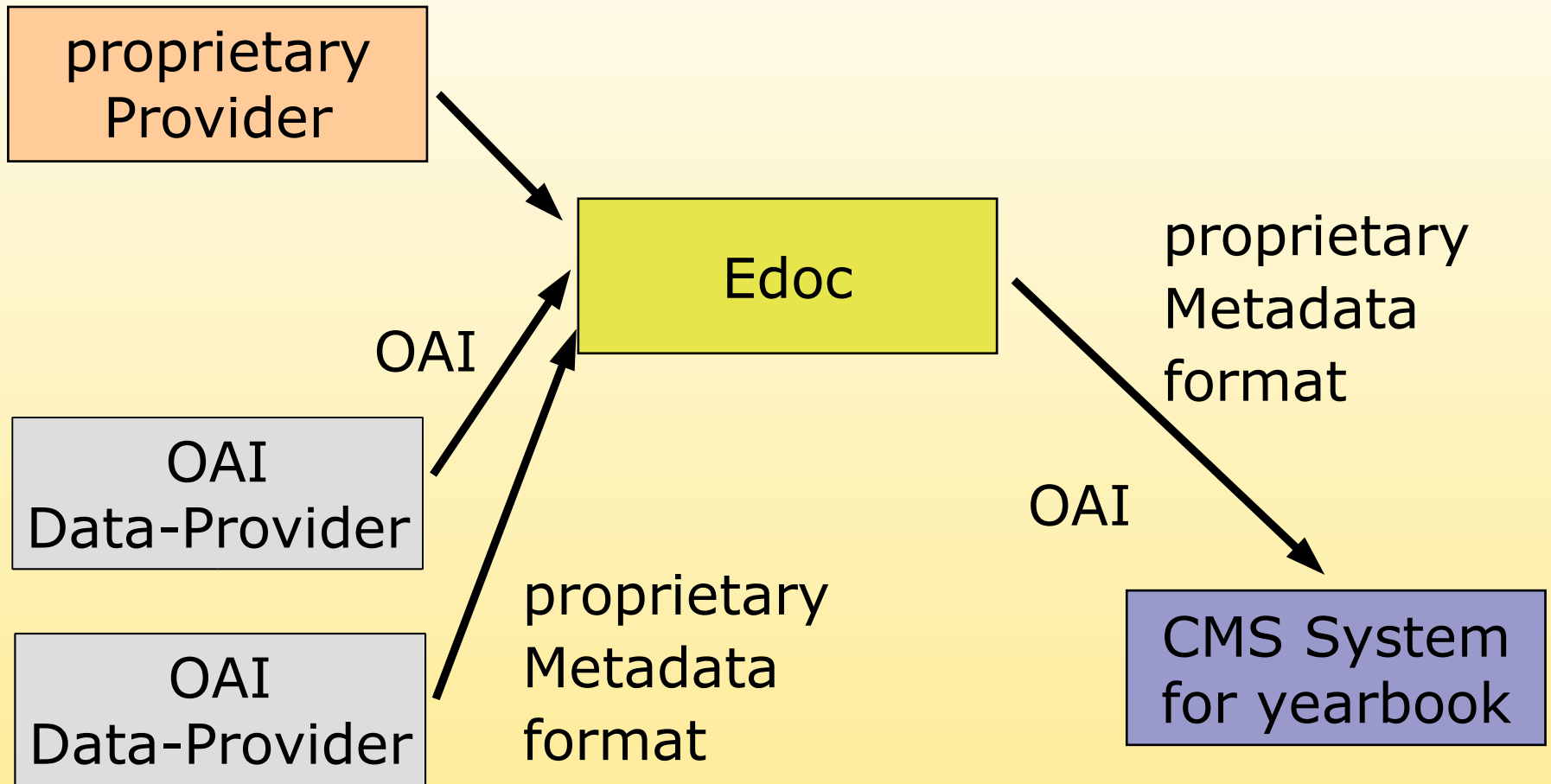
PhysDoc as Service-Provider



Edoc at Max Planck Society

- Max Planck Society consists of many research institutes which focus on specific research fields
- Edoc acts as central document repository for articles published by members of Max Planck Institutes
- documents (objects) can be stored directly or objects can be imported from local repositories
- also used to create yearbook of publications made by members

Edoc as intermediate system



DINI

- German Network Initiative
 - *Deutsche Initiative für Netzwerk Information*
 - many scientific institutes and university libraries are members
- propagates OAI through tutorials

Lessons learned from tutorials

- technical details not important for most people (samba...)
- many implementations exist, but people have problems to apply them to existing solutions
- discussions in tutorials always led to
 - organizational problems
 - metadata problems
 - myth of Dublin Core as the only possible format

PMH ist not just replication

- obeying the protocol is not enough
 - xml encoding (and cleaning) of existing content
 - metadata format conversion
 - character format encoding
 - especially (limited) conversion of markup and word format to utf8

SP has to look at content...

- cannot “trust” the incoming metadata
 - normalization necessary
 - metadata is formally correct, but lack of shared semantics
 - simple examples
 - DC.language "deutsch" → "ger"
 - DC.date "1.02.1999" → "1999-02-01"
- manual selections of DP necessary
 - registering of DP outside protocol
 - no shared set semantics

Conclusion

- technical framework has been established
 - update policies, interoperability of large numbers of open archives
- People trying to apply PMH need support for
 - organizational issues
 - incorporating their community metadata format besides Dublin Core as least common denominator
 - conversion issues

Thank you

- OAI am Institute for Science Networking,
Oldenburg:
<http://physnet.uni-oldenburg.de/oai/>
- stamer@uni-oldenburg.de