

Plenum 1, Teil 3 : Architekturen

H.Pfeiffenberger, AWI

- **Requirements**
- **Harvesting versus Metasuche**
- **OAI-PMH, Dublin Core**
- **Data- und Service-Provider**
- **Umfeld an Informationssystemen**

Agenda

- Verständigung auf eine gemeinsame Basis
 - Gemeinsames **Verständnis der Begriffe** und Standards !
 - Notwendigkeit der **Abstimmung** in bestimmten Details
 - Nutzen für alle Zentren: **Schnelleres Lernen**
 - Setzt **Offenheit**, gerade über eigene Fehler, voraus
- Einbettung in andere Aktionen der Zentren
 - Kosten / Nutzen; Nachhaltigkeit

Zielsetzung (1)

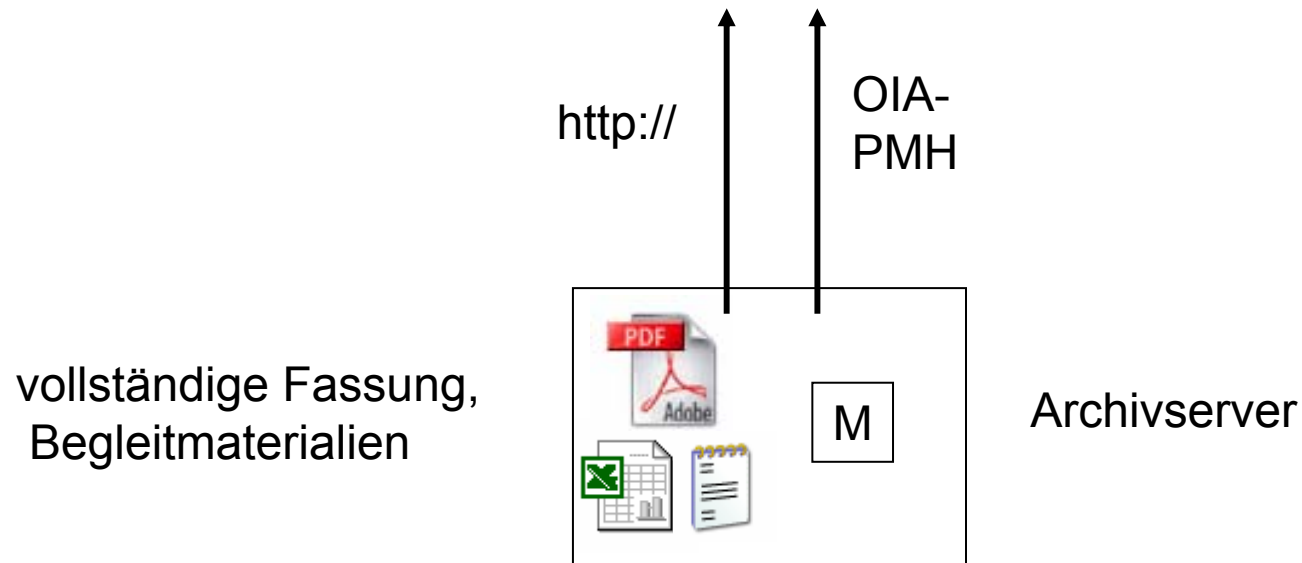
- Laut Implementierungsplan, AP2:
innerhalb 12 Monaten Postprints
bereitstellen
 - so, dass dabei Nutzen
 - für Wissenschaftler,
 - für die Zentren
entsteht (Akzeptanz)
 - unter Beachtung der Nachhaltigkeit
(keine Insellösung im Zentrum; Betriebskosten)

Requirements (1)

- Berliner Erklärung – zur Technik:
 - Eine **vollständige Fassung** der Veröffentlichung samt aller zugehörigen Begleitmaterialien wird ... in einem **geeigneten elektronischen Format** auf mindestens einem **online zugänglichen Archivserver** mit geeigneten **technischen Standards (wie die von Open Archive)** hinterlegt und damit veröffentlicht. Der Archivserver muss betrieben werden von einer wissenschaftlichen Institution oder Gesellschaft ...
(Hervorhebungen H.P.)

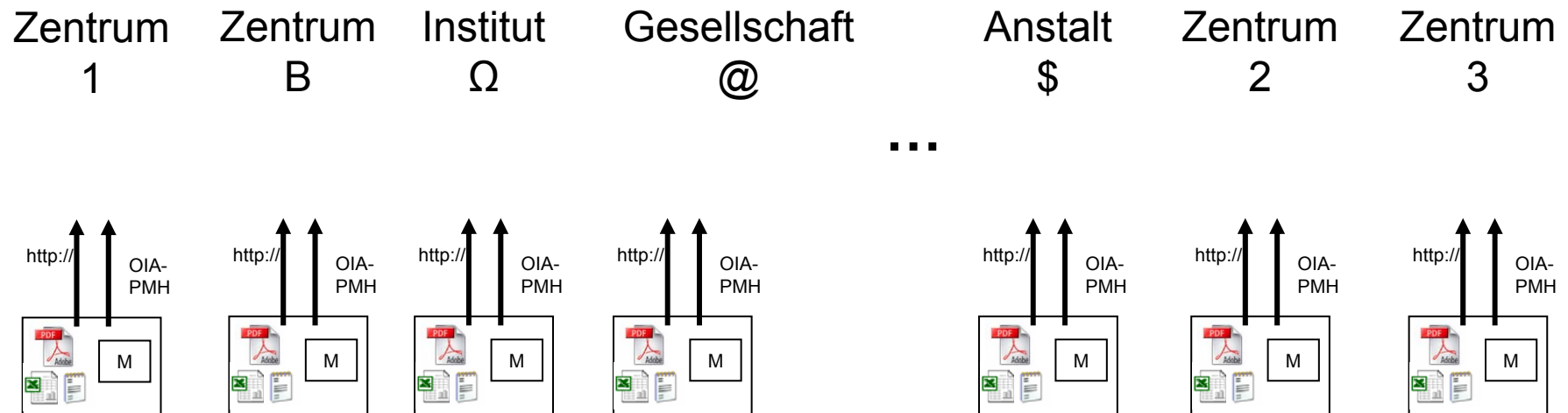
Elemente (1)

- „geeignetes Format“ : heutzutage PDF (und die Materialien?)
- „Open Archives“ :
OAI-PMH (Protocol for Metadata [Harvesting](#))



Architektur (1)

- Föderation Helmholtz-Gemeinschaft
 - Keine Hierarchie
 - Unabhängige Inhalte
 - Unabhängige Implementierung (System, Deadline)



Architektur (2) Warum „Harvesting“

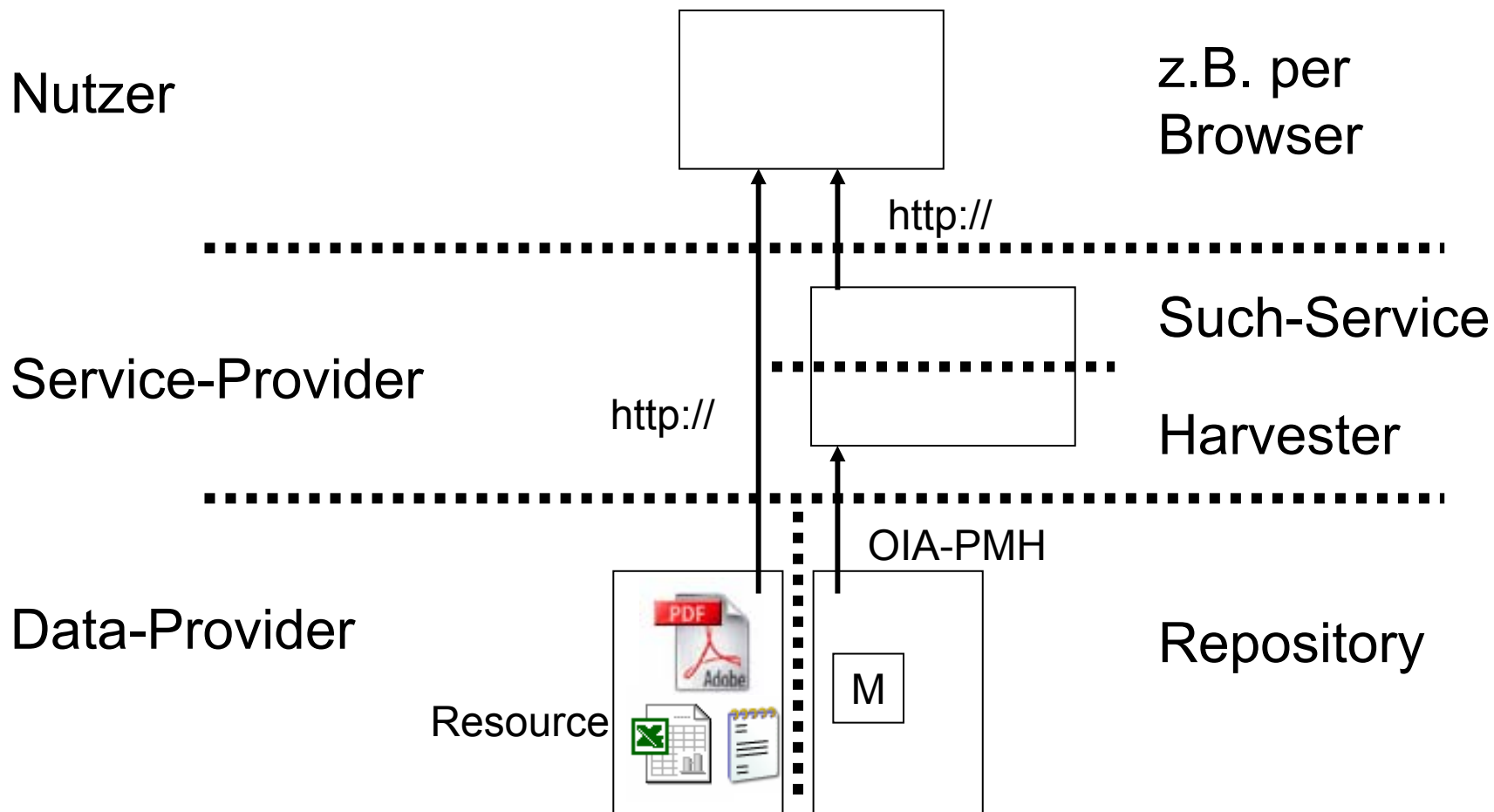
- Metasuche:
 - In der Bibliotheks-Welt ist eher [Z.39.50](#) bekannt und beliebt ;-)) - ein Metasuch-Protokoll
 - Funktioniert nur (hinreichend) gut, wenn **alle (!!)** abgefragten Suchmaschinen **erreichbar und schnell** genug sind.
 - Fraglich ist also, ob man damit über [alle \(relevanten\) Archive der Welt](#) suchen kann („Skalierbarkeit“)
- Harvesting :
 - Hält lokale Kopien (oder daraus gebildete Indizes)
 - Sucht nur auf der lokalen Kopie – **klare Verantwortung!**
 - Skalierbarkeit und Performance: Siehe **Google !!**

Architektur (3a)

Data- und Service-Provider

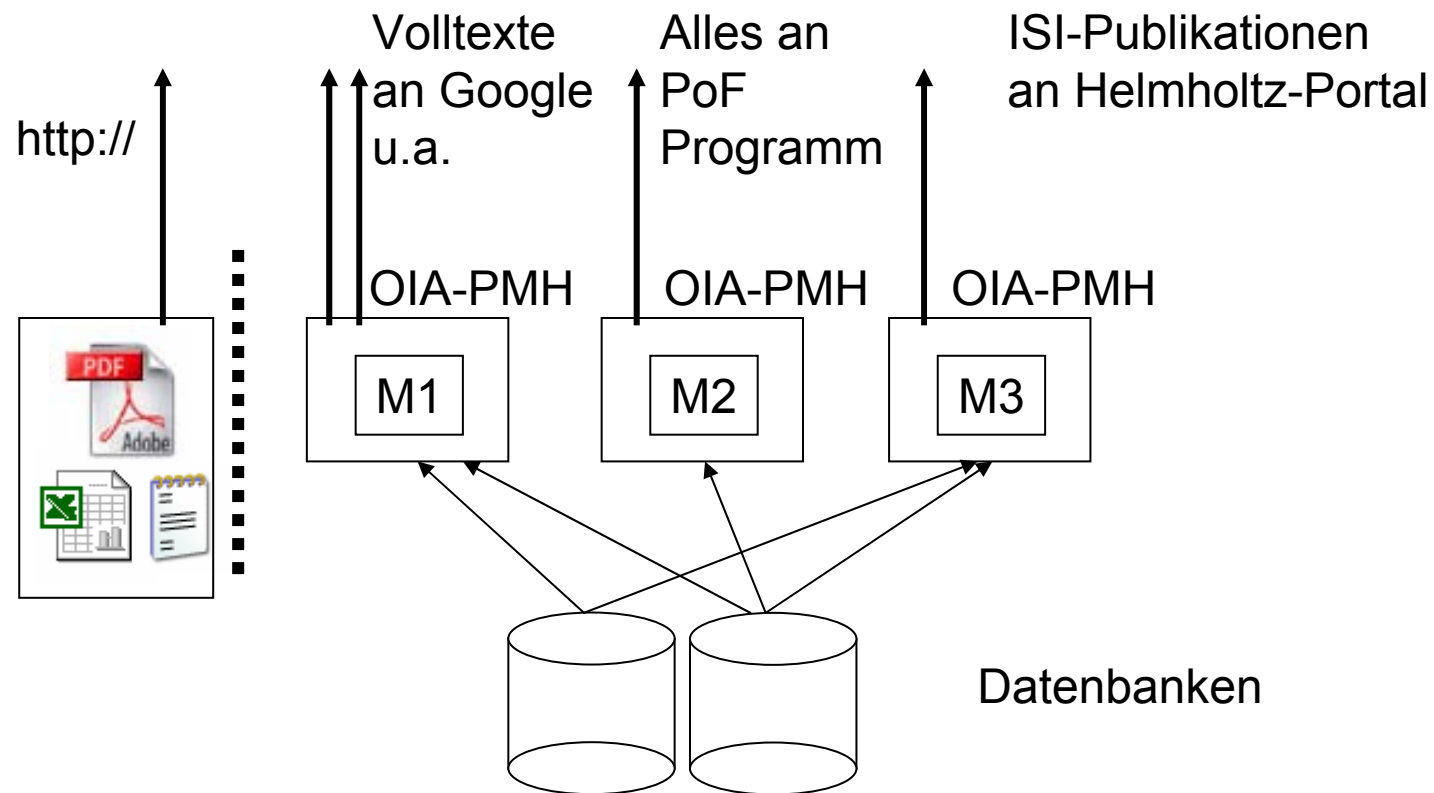
- In der Einleitung der Protokolldefinition von OAI-PMH v2.0 heißt es:
„There are two classes of participants in the OAI-PMH framework:
 - *Data Providers* administer systems that support the OAI-PMH as a means of exposing metadata; and
 - *Service Providers* use metadata harvested via the OAI-PMH as a basis for building **value-added services** „
- Im Dokument ist keine Rede von „**Suche**“

Architektur (3b) Data- und Service-Provider



Architektur (3c)

- Mehrere Sichten möglich durch mehrere OAI-PMH Data-Provider



abgeleitete Zielsetzung (2)

- Erhöhung des Impact !
 - Zumindest für Google (Scholar) sollten wir aktiv optimieren
 - Wir lassen uns von „jedem“ ab-ernten
 -
 - Ein „Helmholtz Open Access-Portal“ bringt für den Impact wahrscheinlich nichts
 - Programm- und andere spezielle Portale sind vielleicht für **interne Kommunikation** (der Programme) nützlich, damit aber außerhalb des Scope von AP2

Elemente (2)

Zugang zur „vollständigen Fassung“

- OAI-PMH spezifiziert, dass das Protokoll *selbst* die „Ressource“, die durch die Metadaten beschrieben wird, nicht kennt
- Auf die Ressource (z.B. den Volltext) soll z.B. mittels des **Dublin Core identifier** Elements (nicht etwa des OAI identifiers) verwiesen werden.
- Wir benötigen also noch (mindestens) einen weiteren Standard zum Zugriff auf die Ressource, etwa eine **bestimmte Form einer URL**

Elemente (3) OAI-PMH

- Ein Protokoll für das Harvesting, d.h. es „holt“ die (**Updates der**) **Metadaten** für komplette Kollektionen (**Sets**)
- Es ist ein „Simple“ Protokoll, wohl aus strategischen Überlegungen
- OIA-PMH ist agnostisch bzgl. des Metadaten-Schemas, verlangt **mindestens** die Unterstützung von **Dublin Core**, erlaubt aber mehrere Schemata

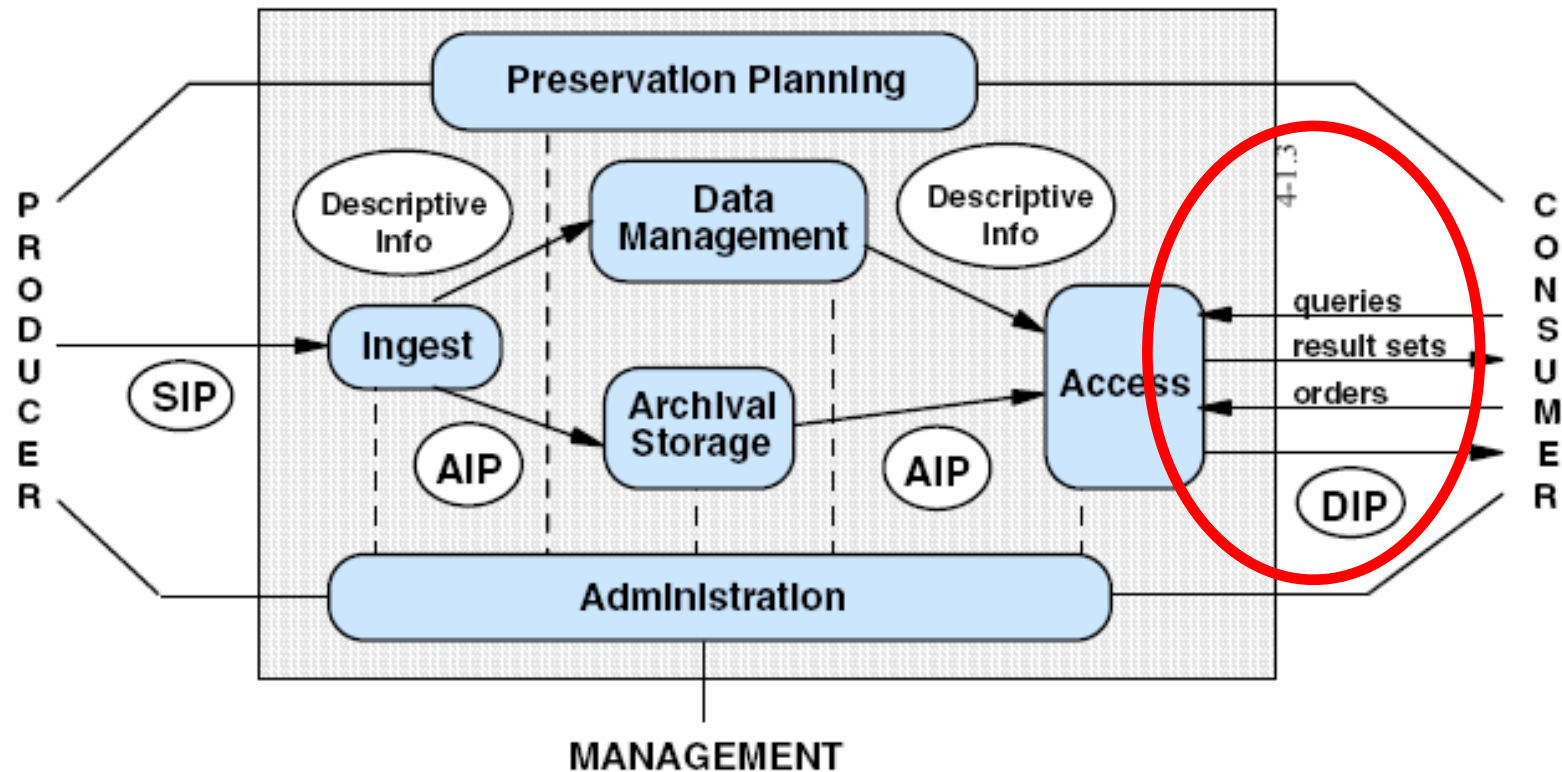
Abgeleitete Zielsetzung (3)

Nachhaltigkeit

- Alle Helmholtz-Zentren haben schon (?) „Data-Provider“, vielleicht **noch nicht** OAI-PMH-fähig ?
- „Einfache“ Lösungen:
 - „**staging**“ der vorhandenen Daten **in ein Open Source System**, das OAI-PMH-fähig ist
 - geht das auch inkrementell, also Updates??
 - **Vorhandenes System** um OAI-PMH Protokoll-Interface **ergänzen**
 - hat das Zentrum dieses Maß an Kontrolle über das vorhandene System?

Architektur (4)

- OAIS Functional Model



Zusammenfassung

- Die Forderungen der Berliner Erklärung geben uns eine klare Richtung für die **Architektur des Zugangs** zu unseren Informationssystemen
- Die Einzelheiten sind im Grunde einfach zu realisieren, **Probleme im Detail** sind durch gute **Information und Abstimmung** zu lösen
- Die Hauptfrage für jedes Zentrum wird sein: Machen wir bei dieser Gelegenheit (noch) einen **Anbau oder renovieren** wir?
- Ich hoffe, einen Anstoß zur **Diskussion** gegeben zu haben, gleich oder in der nächsten Session