

Next Generation Repositories -

den Mehrwert von Repositories für die
wissenschaftliche Kommunikation erhöhen

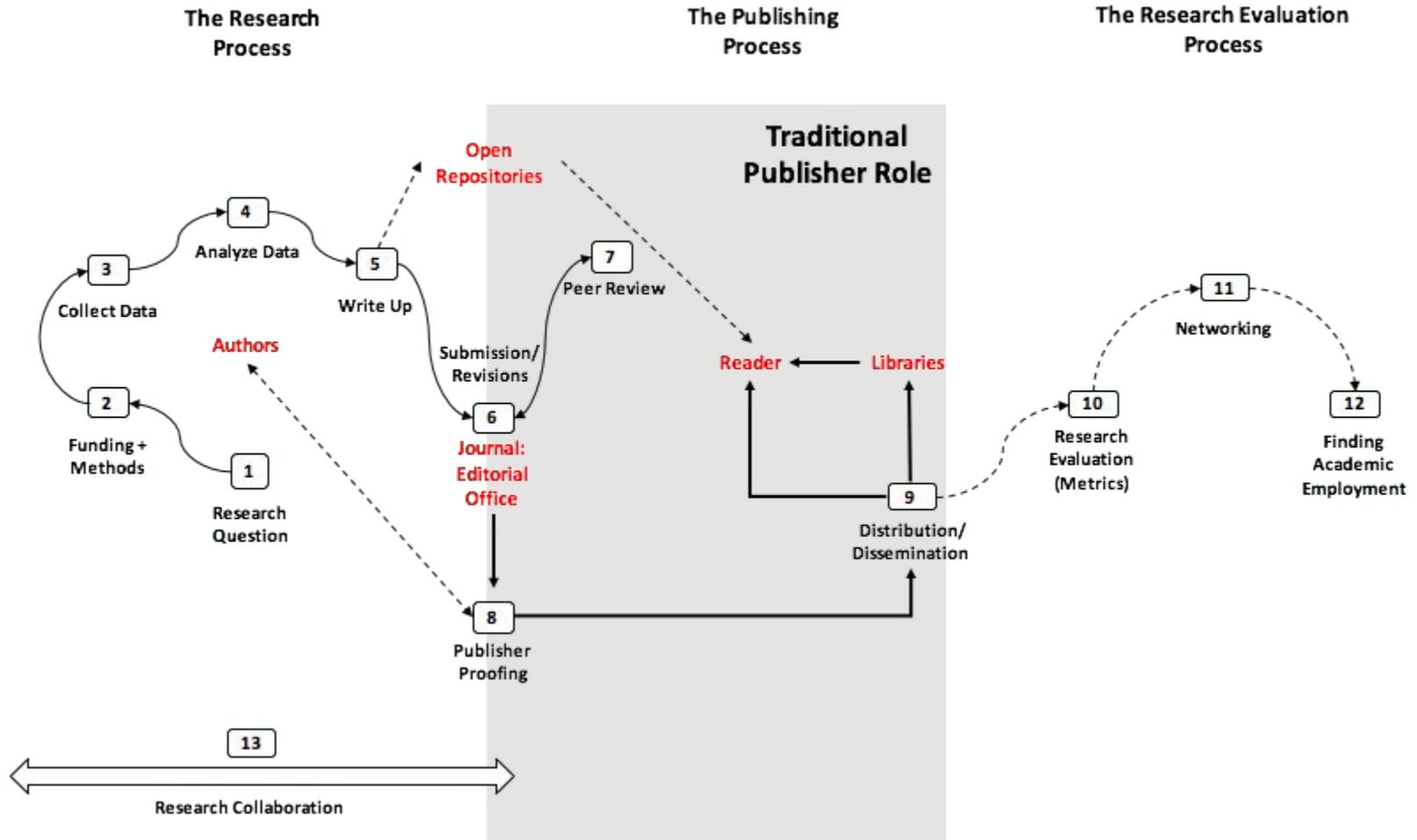
Vitali Peil, Jochen Schirrwagen
UB-Kolloquium Wissensinfrastruktur, 18.05.2018

Agenda

- Motivation
- Herausforderungen für heutige Repositorien
- COAR Next Generation Repositories
- ausgewählte Anwendungsfälle

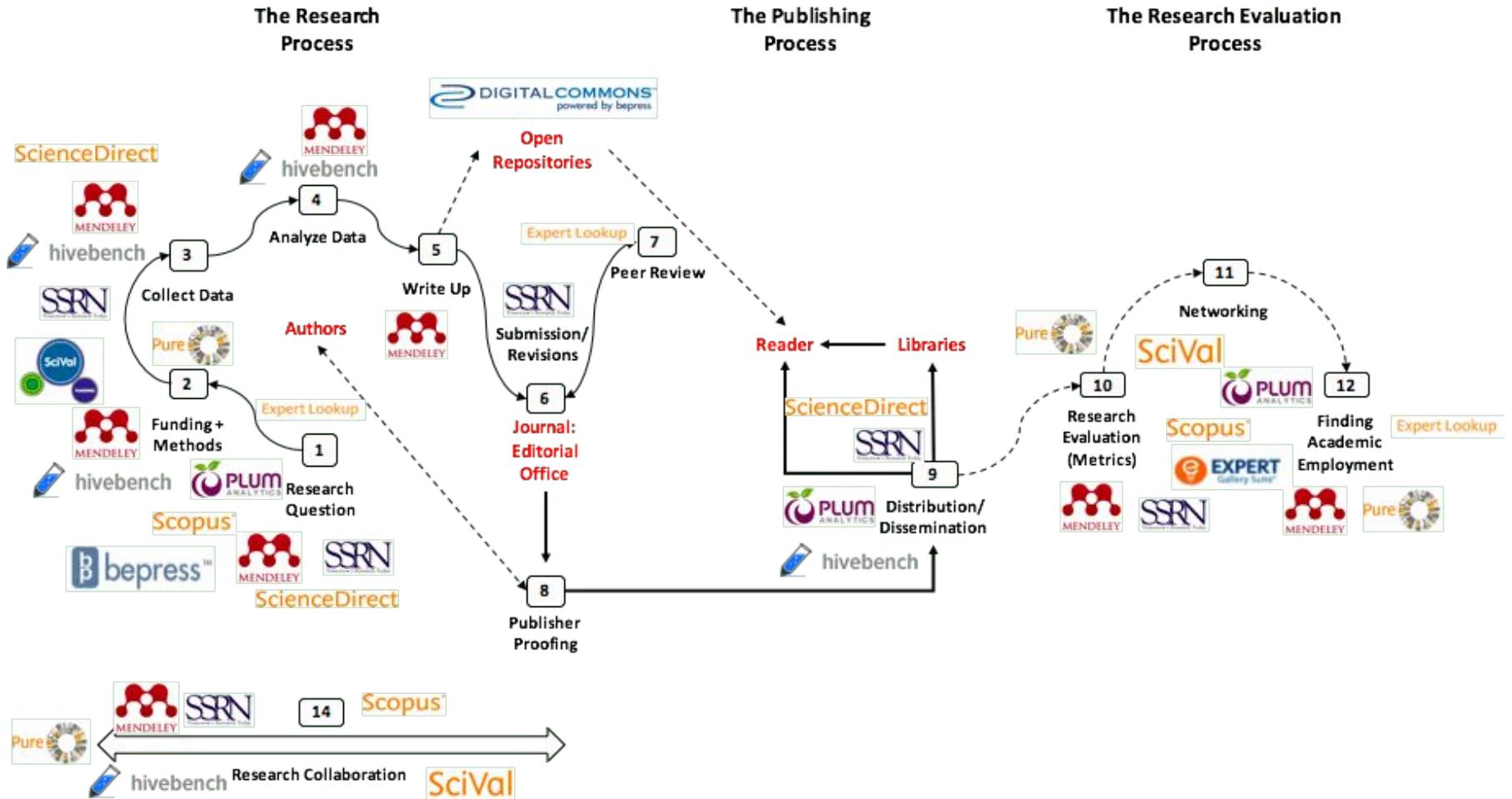
Motivation - oder wem gehört die Infrastruktur?

The Academic Knowledge Production Process



Motivation - oder wem gehört die Infrastruktur?

The Academic Knowledge Production Process



Herausforderungen für Next Generation Repositories:

- bessere Integration in den Forschungskreislauf
- bessere Interaktion durch Mehrwertdienste
 - Verlinkung, Versionierung, Peer-Review, Annotation
- zentriert und fokussiert auf publizierte Forschungsergebnisse als Web-Ressourcen
- Unterstützung moderner Web-Technologien als Grundlage für globale webbasierte Repository-Dienste
- Erarbeitung von User-Stories, Anwendungsfällen und Empfehlung möglicher Standards und Protokolle

11 Anwendungsfälle für Repositories

	Anwendungsfall
1	Exponieren von Identifiern
2	Lizenz-Deklaration auf Ebene der Objekte / Ressourcen
3	“Discovery” durch Navigation
4	Interaktion (durch Annotation, Kommentare, Peer Review)
5	Resource Transfer
6	“Batch Discovery”
7	Ereignis-/aktivitäts orientierte Nachrichtendienste
8	Nutzeridentifikation
9	Nutzerauthentifizierung
10	Veröffentlichung standardisierter Nutzungsstatistiken
11	Langzeitarchivierung

Data Mining & autom. Klassifikation

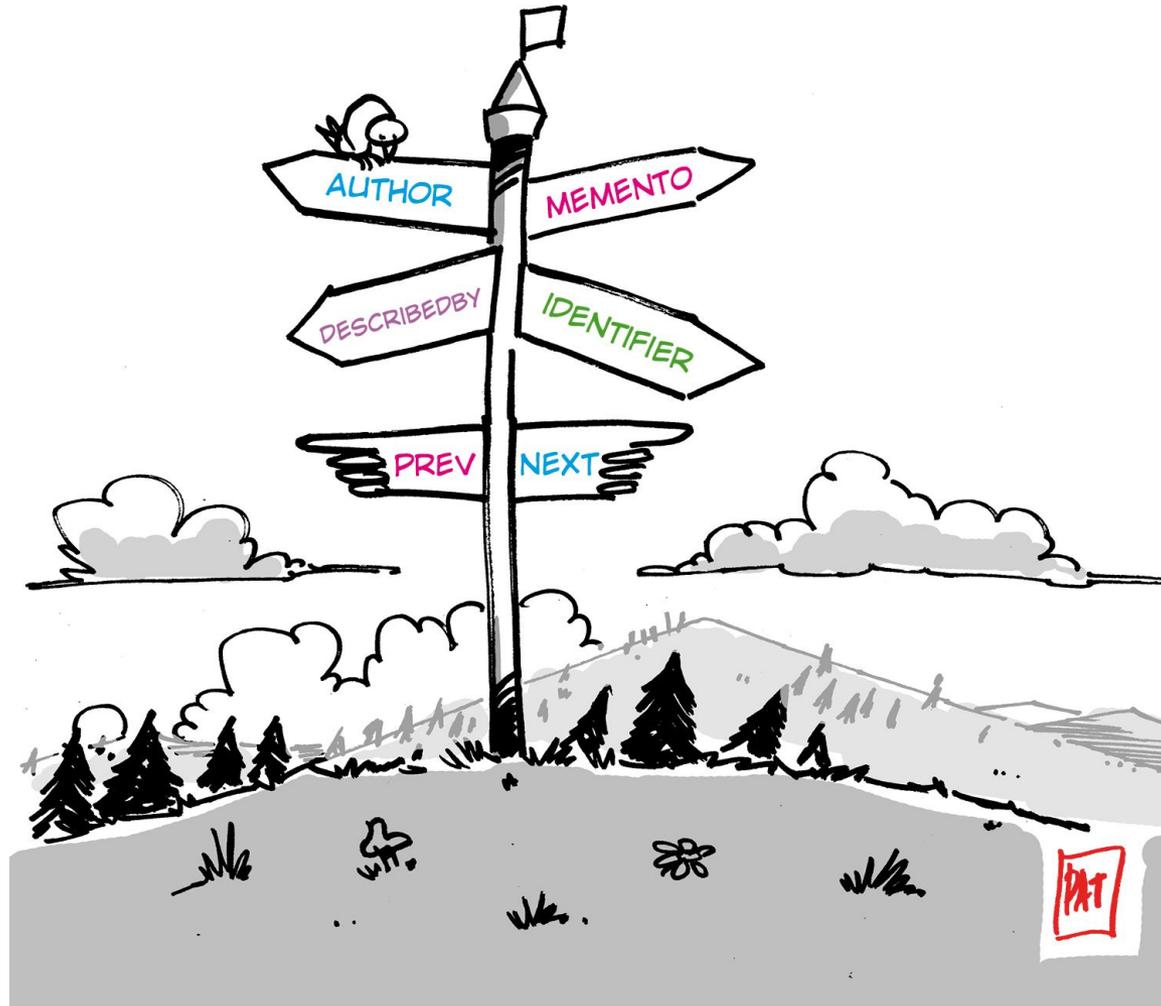
Wie findet man die Volltext-Links in einem Repository, um die Volltexte dann zu verarbeiten?

Wo ist der Volltextlink?

```
<dc:type>doc-type:article</dc:type>
<dc:type>text</dc:type>
<dc:identifier>https://pub.uni-bielefeld.de/publication/1784184</dc:identifier>
▼<dc:identifier>
  http://nbn-resolving.de/urn:nbn:de:0070-pub-17841840
  </dc:identifier>
▼<dc:identifier>
  https://pub.uni-bielefeld.de/download/1784184/2314347
  </dc:identifier>
▼<dc:source>
  Krämer A, Herth D, Keyserlingk H von, et al. Non-responsiveness to hepatitis-B
  674.
  </dc:source>
  <dc:language>eng</dc:language>
▼<dc:relation>
  info:eu-repo/semantics/altIdentifier/doi/10.1007/BF01726924
```



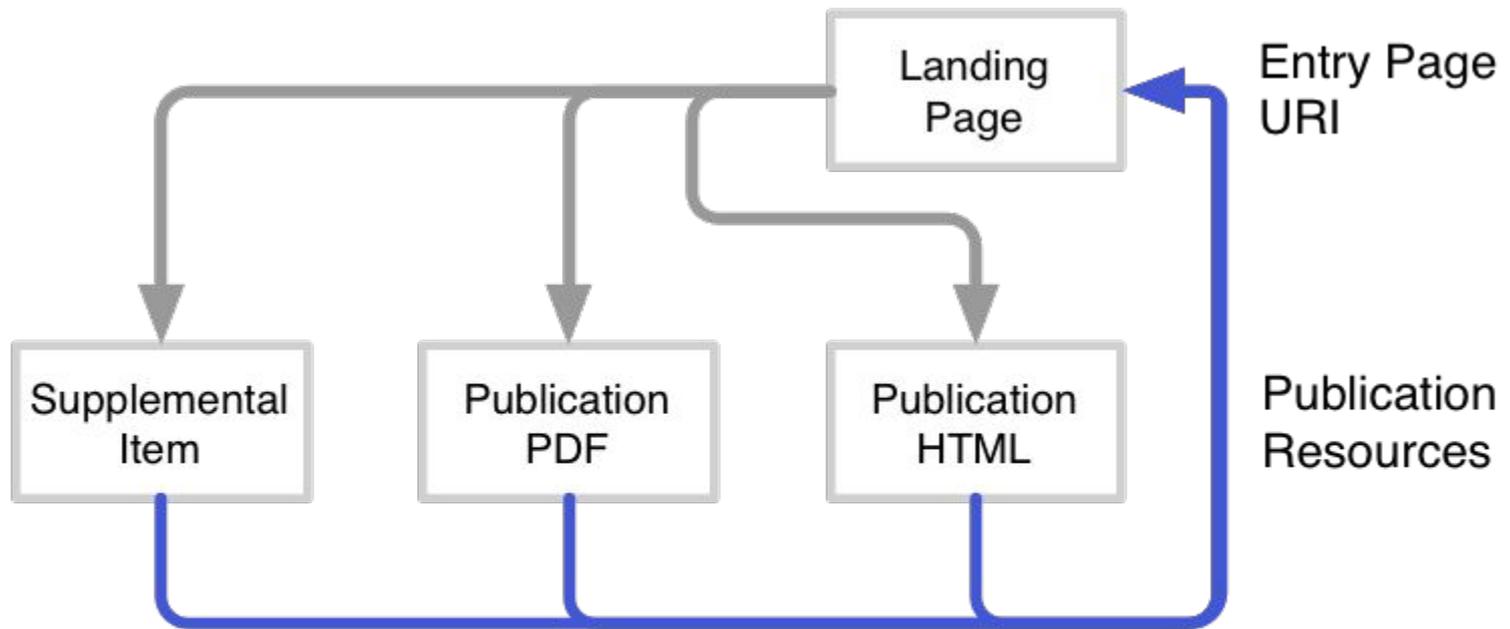
- Es existieren Links in den Metadaten, aber diese sind nicht (immer) semantisch ausgezeichnet.
- Lösung: Signposting



Quelle: Zeichnung erstellt von Patrick Hochstenbach (<http://signposting.org/>)

Signposting

“Signposting is an approach to make the scholarly web more friendly to machines. It uses Typed Links in HTTP Link headers as a means to clarify patterns that occur repeatedly in scholarly portals.”



— rel="collection"
— rel="item"

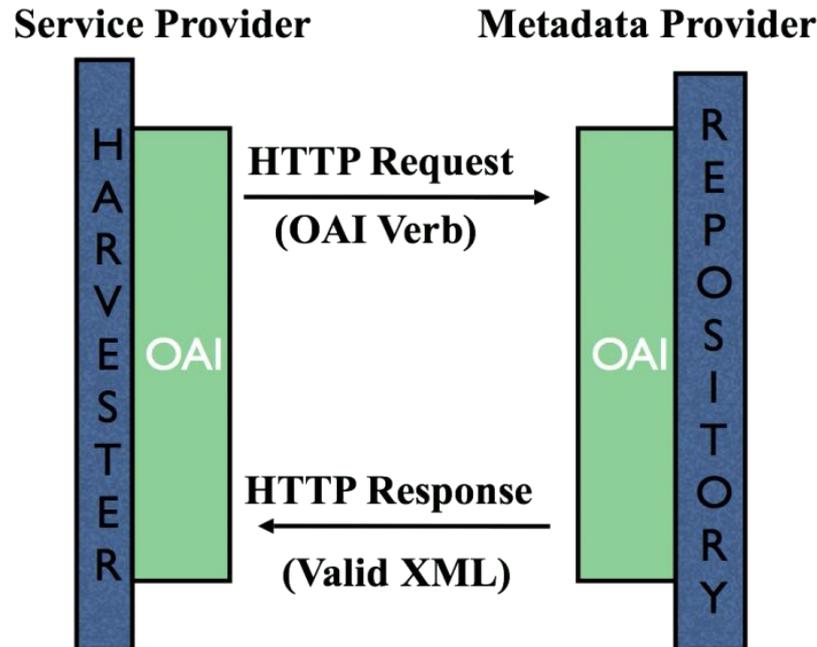
Live-Demo Signposting

- `curl -I "https://doi.pangaea.de/10.1594/PANGAEA.867908"`
- `curl -I "https://doi.pangaea.de/10.1594/PANGAEA.867908?format=zip"`

Batch discovery

- Wie erhält man sämtliche Dokumente aus einem Repository?
- Wie kann man Quellen unterschiedlicher Größe und mit unterschiedlichen Update-Intervallen synchronisieren?
-

Warum nicht OAI-PMH ?



- Repository- aber nicht Ressourcen-zentriert
- nur XML-Metadaten
- wiederkehrender Metadatenaustausch von Daten- zu Serviceprovider
- nicht für Content-Transfer
- technologisch veraltet

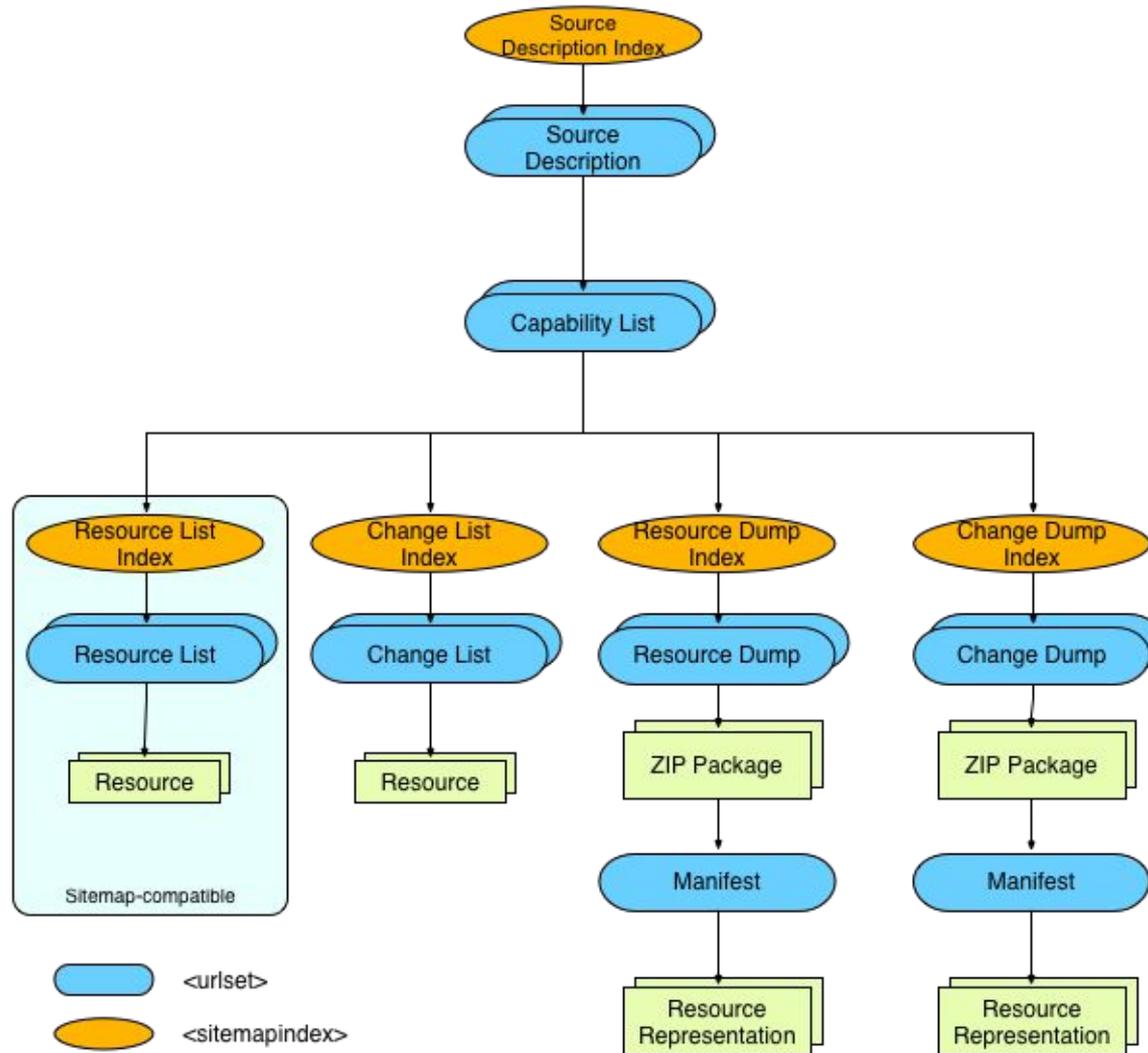


ResourceSync

- ein Framework für “Discovery”, Harvesting und Synchronisation von Web-Ressourcen
- ANSI / NISO Z39.99-2017 Standard
- entwickelt 2012/2013 mit Update in 2017 (v1.1)
- für beliebige mit HTTP URIs adressierbare Web-Ressourcen
- basiert auf und erweitert das Sitemap-Format
- folgt dem Ansatz des REST/HATEOAS (Hypermedia As The Engine Of Application State)
 - siehe <https://en.wikipedia.org/wiki/HATEOAS>

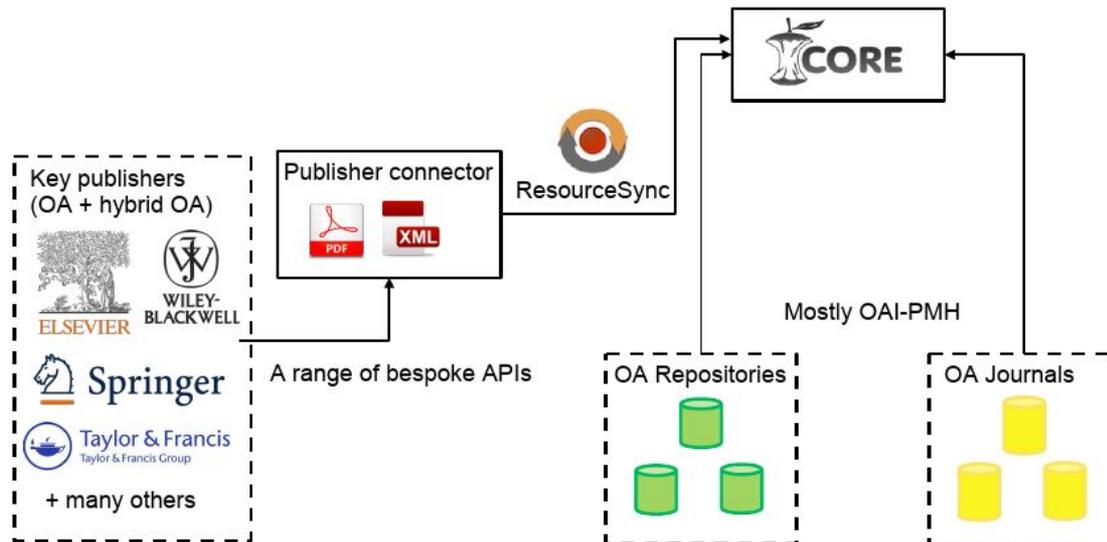


ResourceSync Framework Struktur



Anwendungsfall: Synchronisation Metadaten und Volltexte am Beispiel CORE-UK

- skalierbar
- unabhängig von bestimmten Metadatenformaten



Demo

- <http://publisher-connector.core.ac.uk/resourcesync/>

Repository-Nutzungsstatistiken

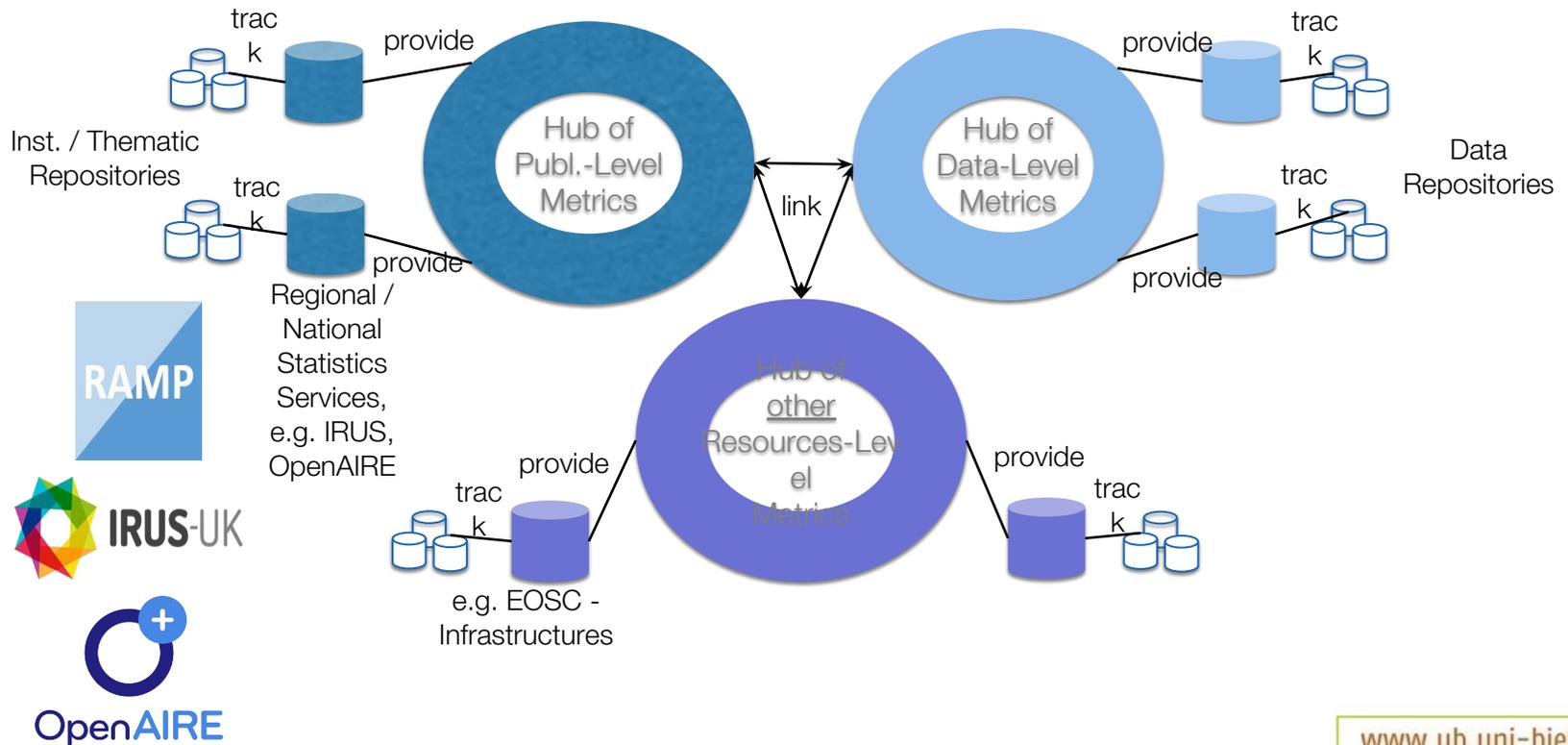
- Warum?
 - Anreize für Autoren
 - Anhaltspunkt für die Akzeptanz / Nutzung der Plattform des Repositorienbetreibers
 - Beantwortung von Fragestellungen wie Vergleich der Nutzung OA vs. nicht-OA Version eines Artikels
- können zeitnah erfasst, geloggt, verarbeitet werden
- COUNTER ermöglicht standardisierte, verlässliche Nutzungsstatistiken
- offener Zugang zu aggregierten Nutzungsdaten
- Komplement zu klassischen Metriken
- Aber: Herausforderungen bzgl. Vergleichbarkeit zwischen Nutzungsdaten

Herausforderung I - Aussagekraft

- absolute Nutzungszahlen über Downloads und Views haben keine besondere Aussagekraft
- D.h. Bezüge herstellen
 - durch Betrachtung des zeitlichen Verlaufs
 - zu Fachdisziplin und Größe der Community
 - durch Vergleich mit “gleichartigen” Publikationen auf der gleichen Plattform
 - durch Vergleich zwischen Publikationen eines Projekts
 - im Zusammenhang mit Nutzungsstatistiken verlinkter Forschungsergebnisse (Bsp. Publikationen - Forschungsdaten)

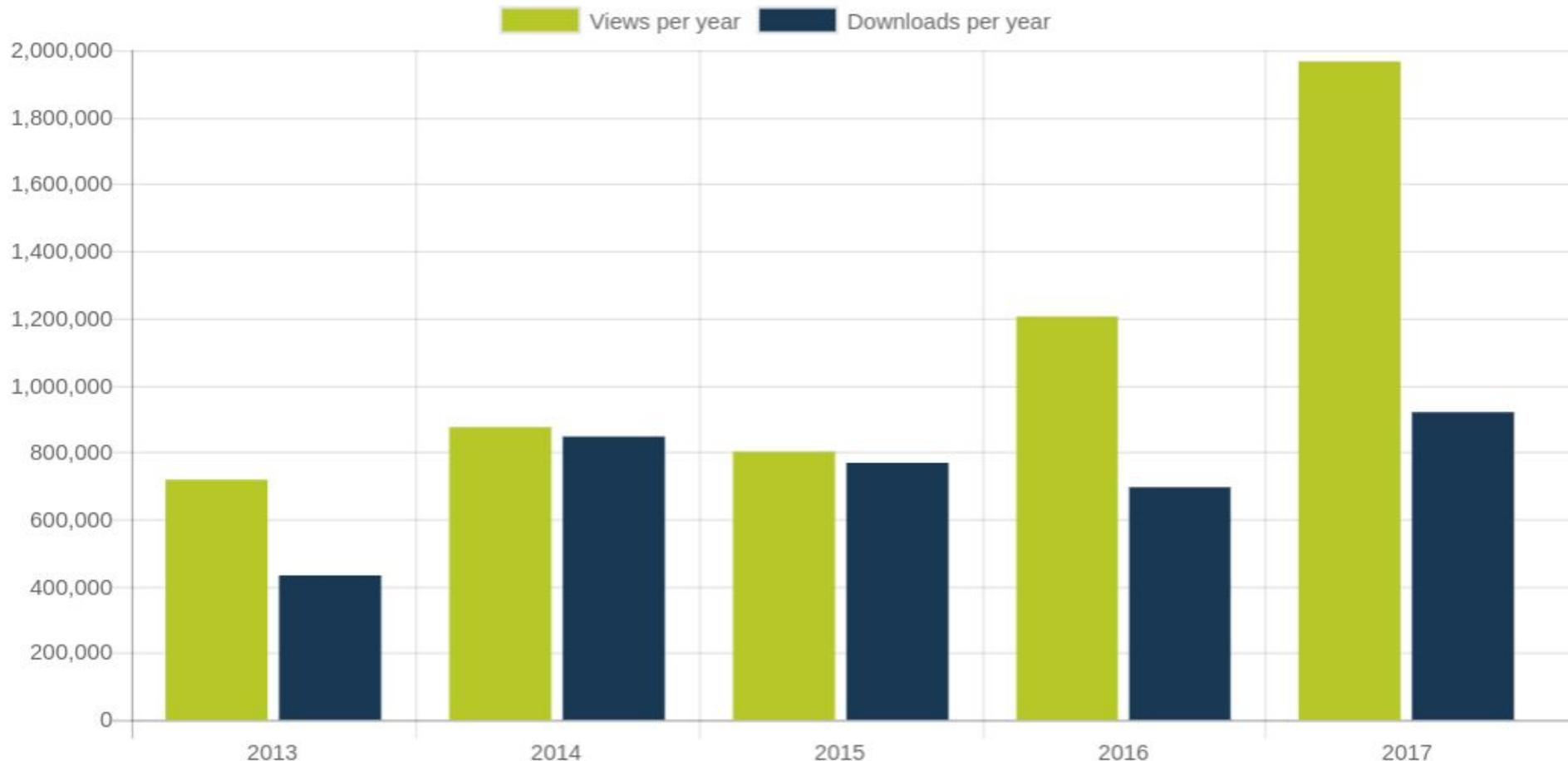
Herausforderung II - Zusammenführung

- Zusammenführung der Statistiken aus verschiedenen Quellen



Anhaltspunkt Nutzungsstatistiken aus PUB

Aggregated Usage Statistics from recent years



Quellen und weiterführende Links

COAR - Next Generation Repositories	http://ngr.coar-repositories.org/
Signposting	http://signposting.org/
ResourceSync	http://www.openarchives.org/rs/toc
SUSHI-Lite	https://www.niso.org/standards-committees/sushi/sushi-lite
Hypothes.is	https://web.hypothes.is/
Web Annotation Model and Protocol	https://www.w3.org/TR/annotation-model/ https://www.w3.org/TR/annotation-protocol/

Klein, M., Sanderson, R., Sompel, H.v.d (2013). A Technical Framework for Resource Synchronization. doi:10.1045/january2013-klein

Dankeschön!