

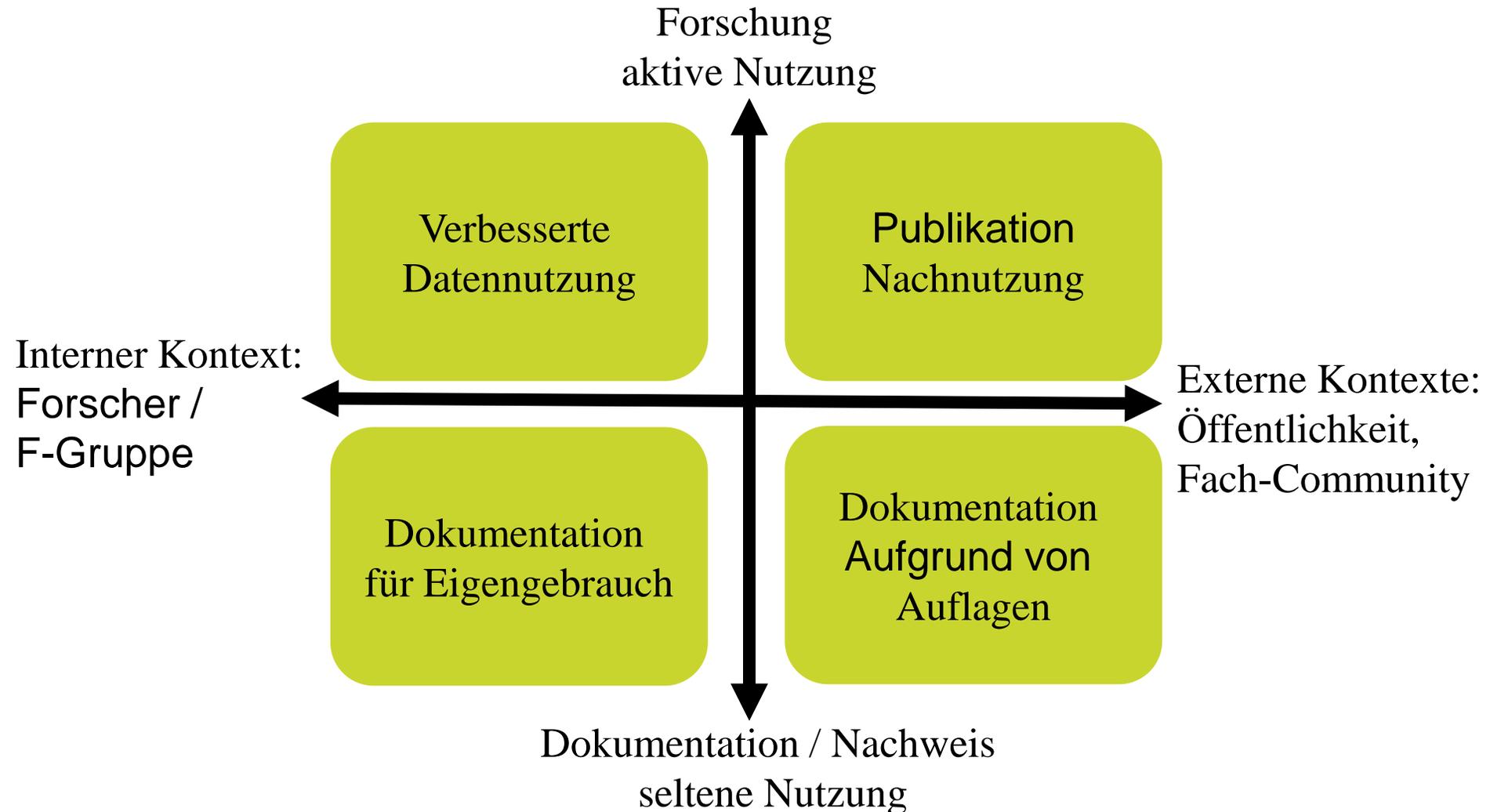
MODUL 3-2: Daten Nutzen – welche Daten wie für die Nachnutzung aufbereiten

Niklaus Stettler
Schweizerisches Institut für Informationswissenschaft (SII)
Lizenz CC BY 2.5

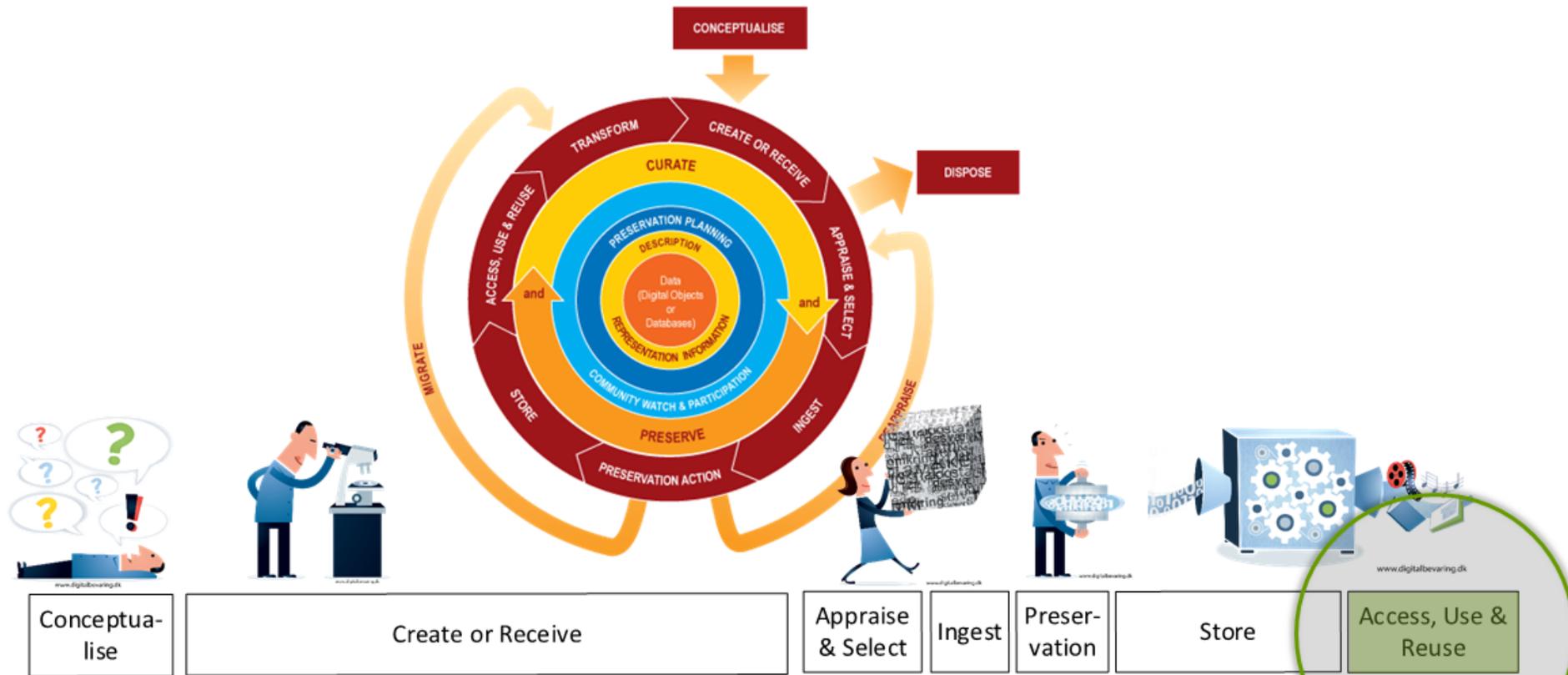
All cartoons courtesy of Jørgen Stamp,
digitalbevaring.dk
CC BY 2.5



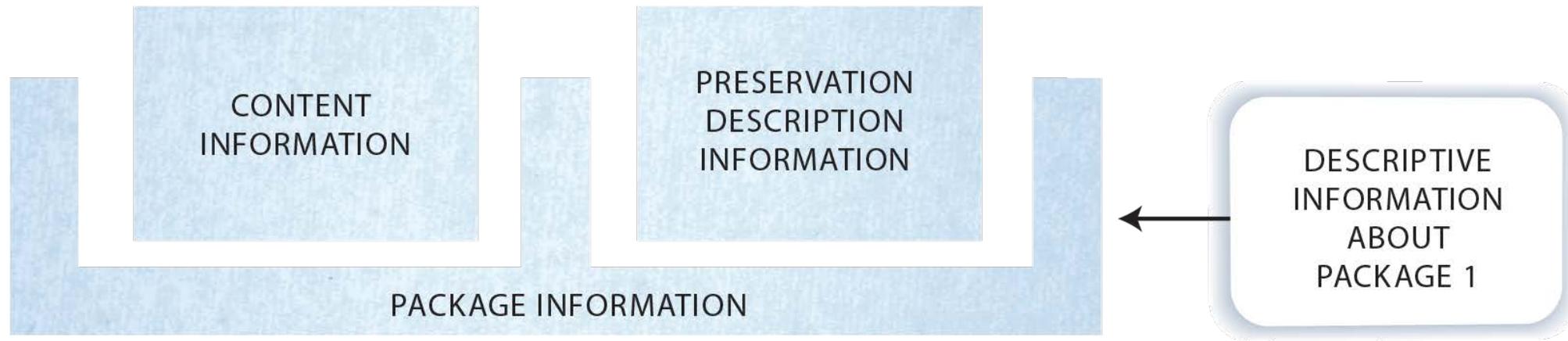
Ziele des Forschungsdatenmanagement



DCC CURATION LIFECYCLE MODEL: ACCESS, USE & REUSE



Im Ingest entstehen Informationspakete



PACKAGE 1

www.digitalbevaring.dk



Daten sollen benutzbar bleiben

In Modul 2-4 wurde definiert: Damit Daten auch in 50 Jahren noch benutzbar sind, ...

- müssen die Daten in einem Katalog gefunden werden
- müssen die Daten in einem Format sein, das noch geöffnet werden kann
- muss das Format in einer format-registry beschrieben sein, damit wir auf allfällig notwendige Migrationen aufmerksam werden

Mit dem richtigen Format ist es noch nicht getan

Die Erhaltung der Daten in einem lesbaren Format garantiert noch nicht deren Verständlichkeit

- Daten müssen richtig dargestellt werden
- Ev. braucht es zusätzlich zu den Daten Erläuterungen, um die Daten zu verstehen

ID	TLOC	NbPers	Anz_Ki_HI	Ki_HH	CibleSex	CibleAge	Alter_10	Alter_20	EK_grob	EK_detaill	D1	D2	D3	D3_g
471	1	1	0	2	2	52	4	3	1	1	1	1	4	2
2124	3	1	0	2	2	35	3	2	1	1	1	1	4	2
1841	2	1	0	2	1	44	3	2	1	1	1	1	4	2
2447	3	2	0	2	1	23	1	1	1	1	1	3	7	3
2010	1	1	0	2	1	39	3	2	1	1	1	1	4	2
586	1	1	0	2	1	69	6	4	1	1	1	4	5	2

Die Zusatzinformation, die der künftige Nutzer für das Verständnis der Daten braucht, nennt OASIS: representation information

Representation Information im OAIS Informationsmodell

Representation Information ist die Anleitung zur **Anzeige** und **Nutzung** des Data Objects. Sie enthält:

- Strukturinformation:
z.B. Info. zum Datenformat
- Semantische Information:
z.B. Erklärung zur Bedeutung der Daten in einer Spalte.

Data Object plus RI bilden zusammen das **Informationsobjekt**, das archiviert werden soll.

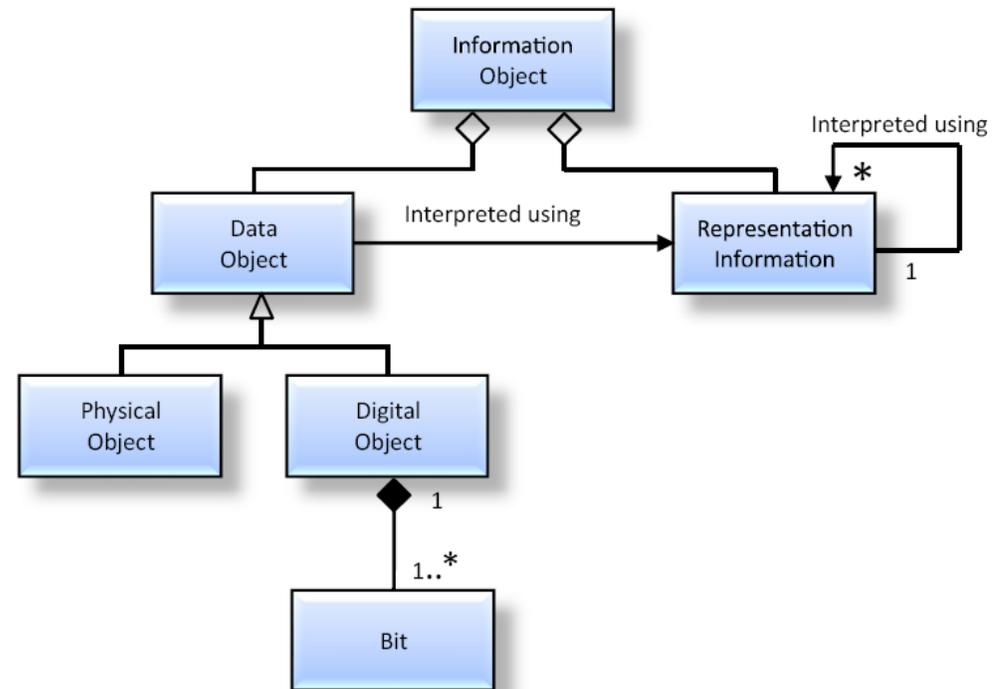


Figure 4-10: Information Object

Representation Information bei Forschungsdaten

Um Forschungsdaten nutzbar zu halten, braucht es z.b:

- Information zum Format
- Information zur Software, um das Objekt darzustellen
- Information zur Methode
- Information zur Erstnutzung
- Erklärungen zu den einzelnen Daten
- Etc.

Viele dieser Daten sind in Form von zusätzlichen Dokumenten auf der Landing Page zugänglich

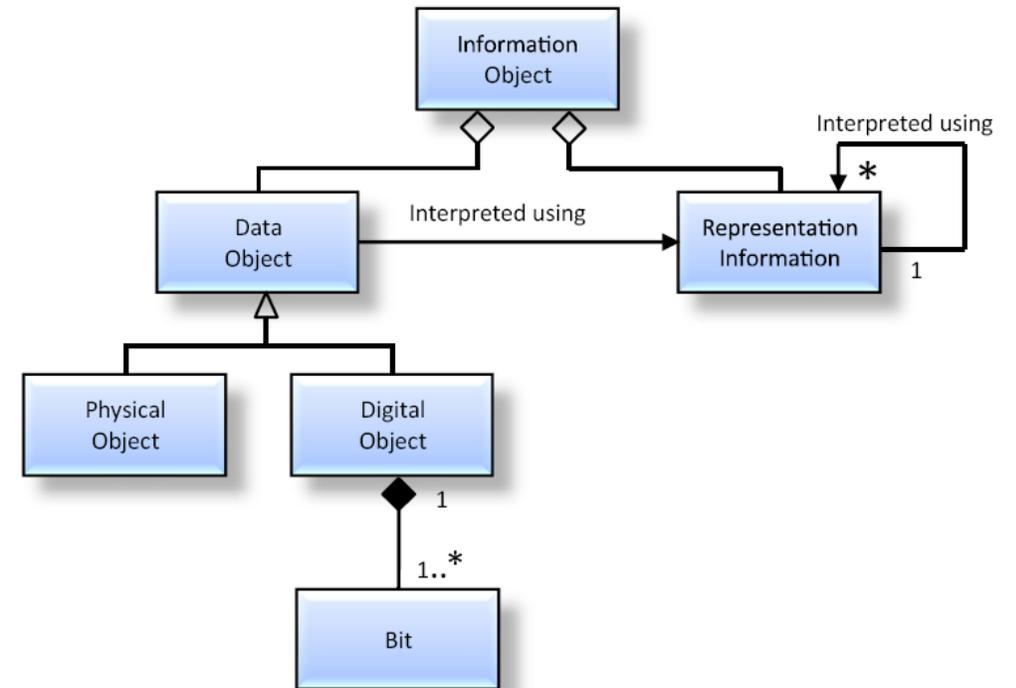


Figure 4-10: Information Object



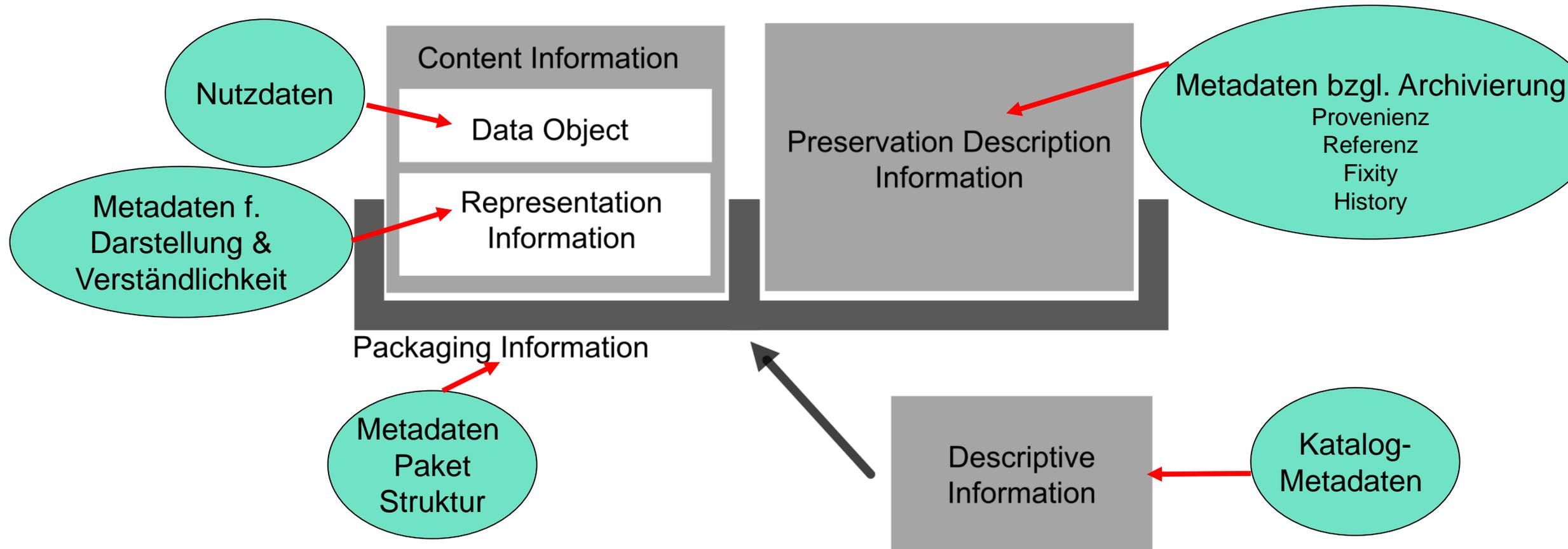
Das Informationspaket ist ‚autonom‘

Zur langfristigen Nutzbarhaltung, konzipiert OAIS das Informationspaket als autonom

D.h., das Paket muss alle Informationen enthalten

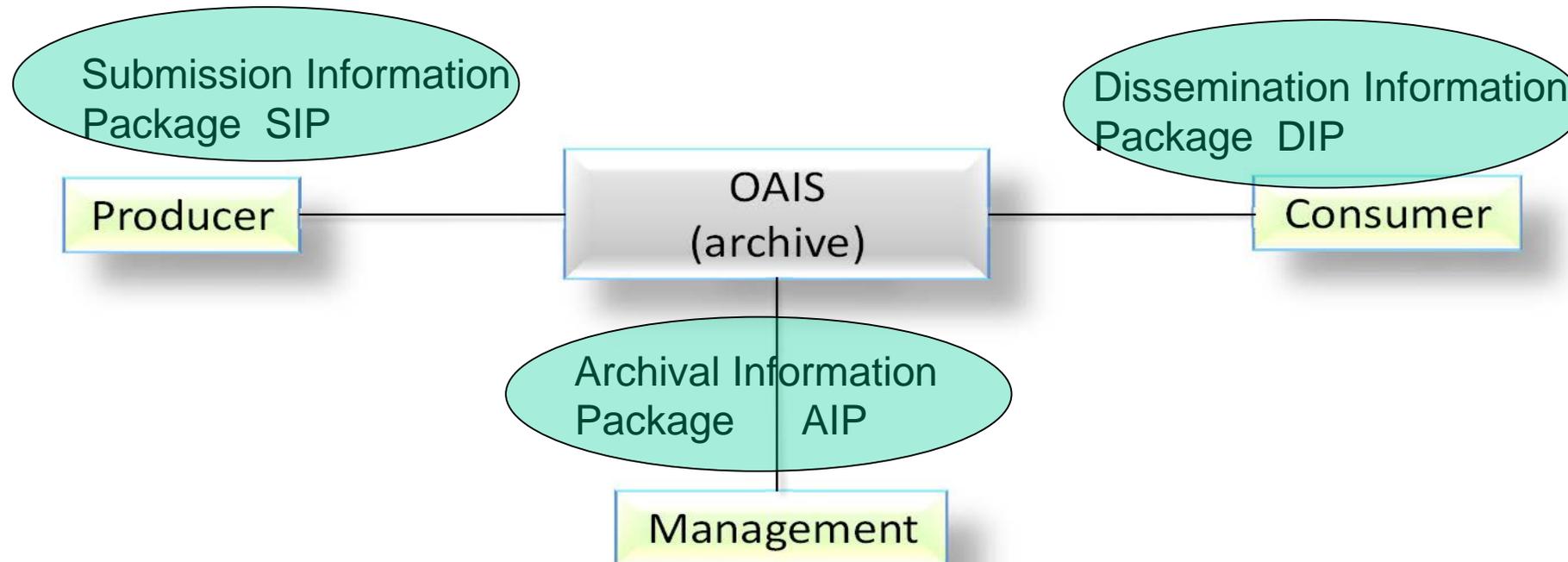
- die das Paket beschreiben und es wiederauffindbar machen
- die es ermöglichen, die Nutzdaten darstellen zu können
- die der designierten Nutzer braucht, um die Nutzdaten verstehen zu können.

Das InformationsPaket



Unterschiedliche Informationspakete

Nicht in jeder Phase des Archivierungsprozesses werden dieselben Informationen benötigt, daher sind die Informationspakete leicht unterschiedlich ausgebildet



Zusammenfassend: Anforderungen an die Daten

- Daten einige Jahre aufbewahren: Im Originalformat erhalten
 Beschreiben für die Suche (DCAT)
 Beschreibung der Methode für die Qualitätskontrolle
- Daten für Nachnutzung aktuell: Daten bereinigen
 Ev. in RDF umwandeln
 Semantik der Daten dokumentieren
- Daten für spätere Nachnutzung: Signifikante Eigenschaften definieren
 In archivtaugliches Format umwandeln
 Ingest-Prozess vollziehen
 Datenpakete schnüren
 Dokumentieren für künftige NutzerInnen, die über eine neue
 Wissensbasis verfügen.

Zusammenfassend: Informationsverlust minimieren resp. planen

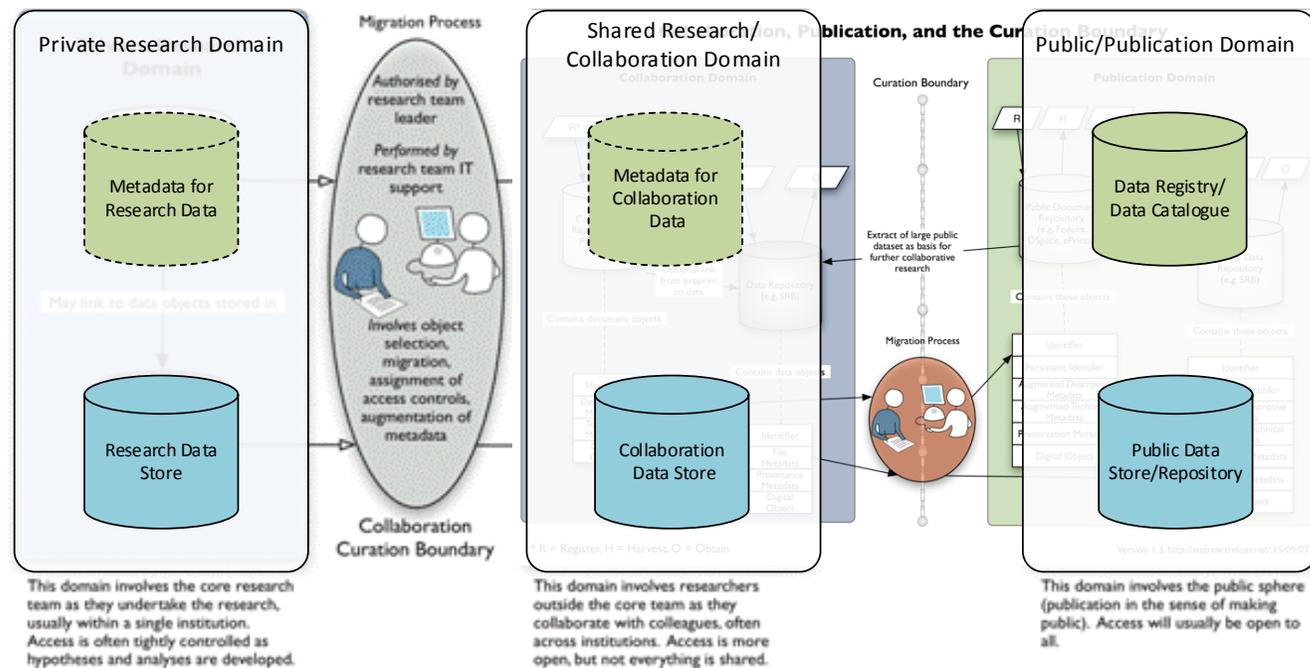
Die Aufbereitung der Daten für die spätere Nachnutzung kann mit Informationsverlust einher gehen.

Um dies zu verhindern:

- Planung der Aufbewahrung bereits bei der Planung der Datenerhebung
- ‚einfache‘ Applikationen
- Frühe präzise Beschreibung der Daten
- Definition der Eigenschaften, die erhalten werden müssen
- Formate einsetzen, die die signifikanten Eigenschaften zu erhalten gewährleisten und eine lange Lebensdauer haben

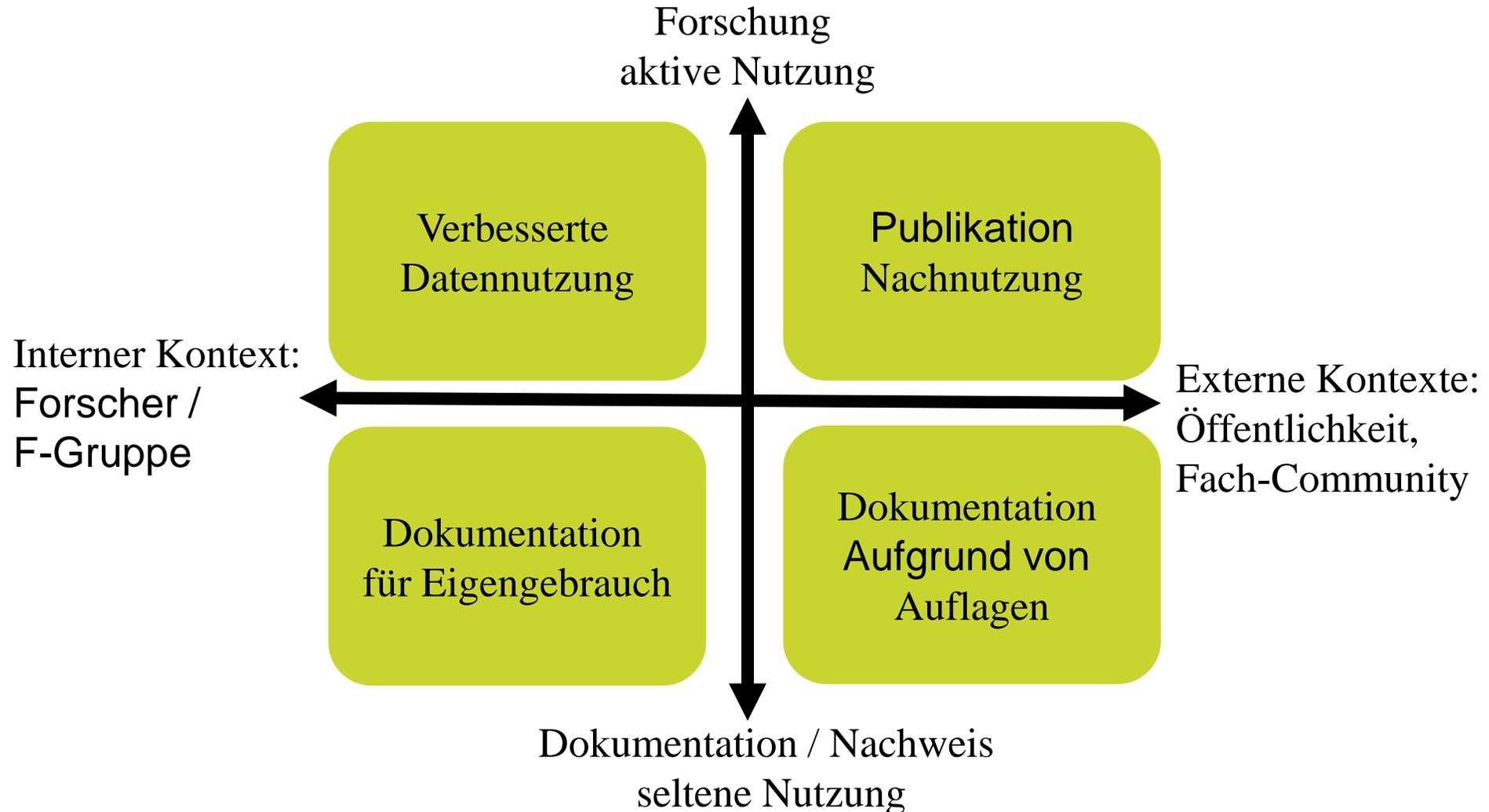
Ein Forschungsdatenportal schafft Zugang zu unterschiedlichen Daten

- Was in der Curation Domain geschieht, kann sehr unterschiedlich sein
- Nicht zwingend müssen die Daten für die Langzeitarchivierung aufbereitet werden
- Welche Daten in welcher Form erhalten werden, entscheidet das Data Center



Welche Daten aufbewahren?

Der Nutzungszweck bestimmt die Form der Daten



Welche Daten aufbewahren?

- Daten müssen im Hinblick auf ihre potentielle Nutzung aufbereitet werden
- Nutzung der Daten für Validierung der Forschungsergebnisse oder für langfristige Nachnutzung?
- Können die Daten ohne Informationsverlust erhalten werden?
- Was bedeutet der Informationsverlust für die Daten, sind sie noch brauchbar?
- Auch für Daten gilt: Bewertung ist entscheidend, insbes. da die Aufbereitung der Daten extrem aufwändig ist.

Vielen Dank