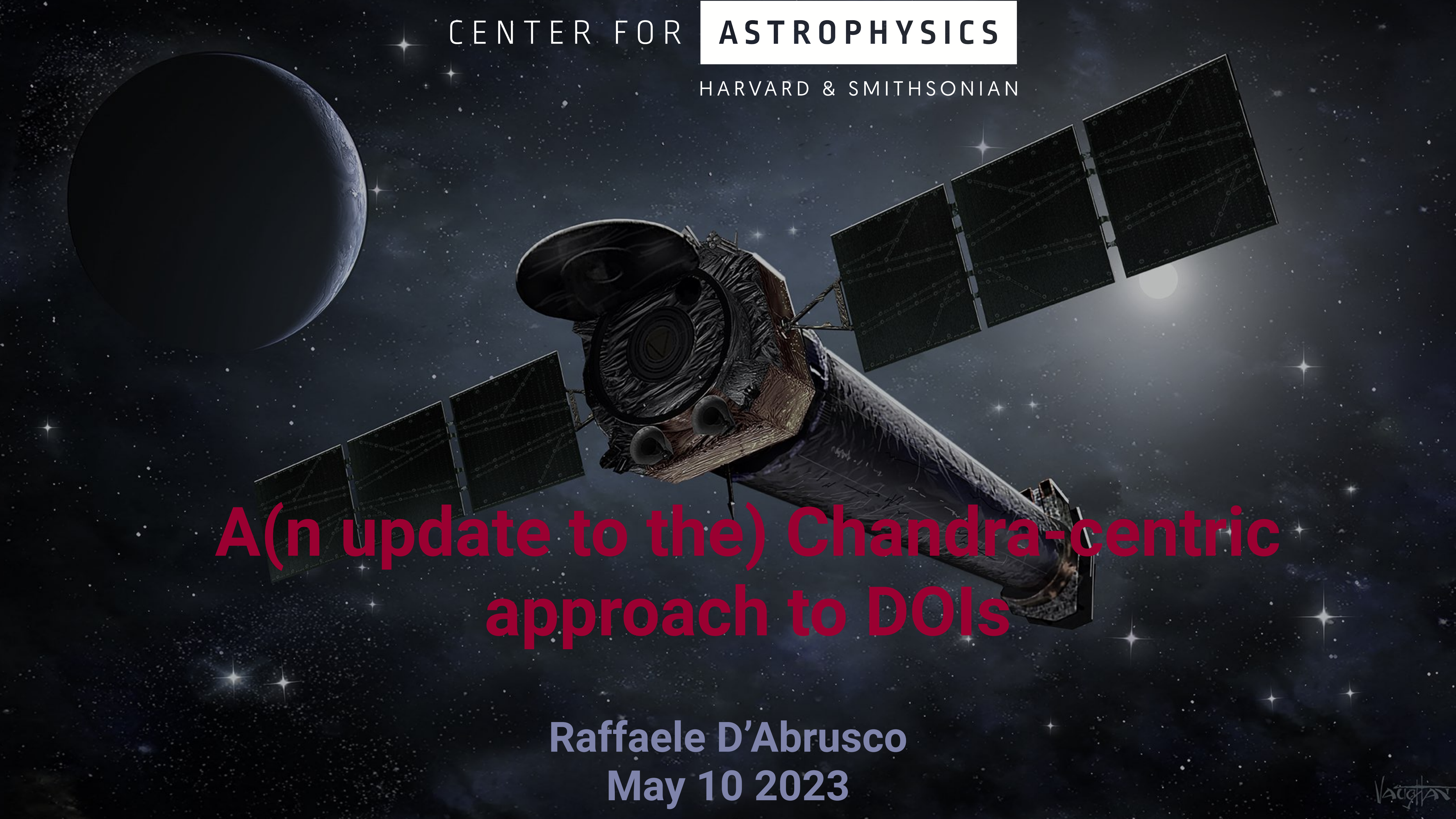


CENTER FOR **ASTROPHYSICS**

HARVARD & SMITHSONIAN



**A(n update to the) Chandra-centric
approach to DOIs**

Raffaele D'Abrusco
May 10 2023

The CDA DOI strategy

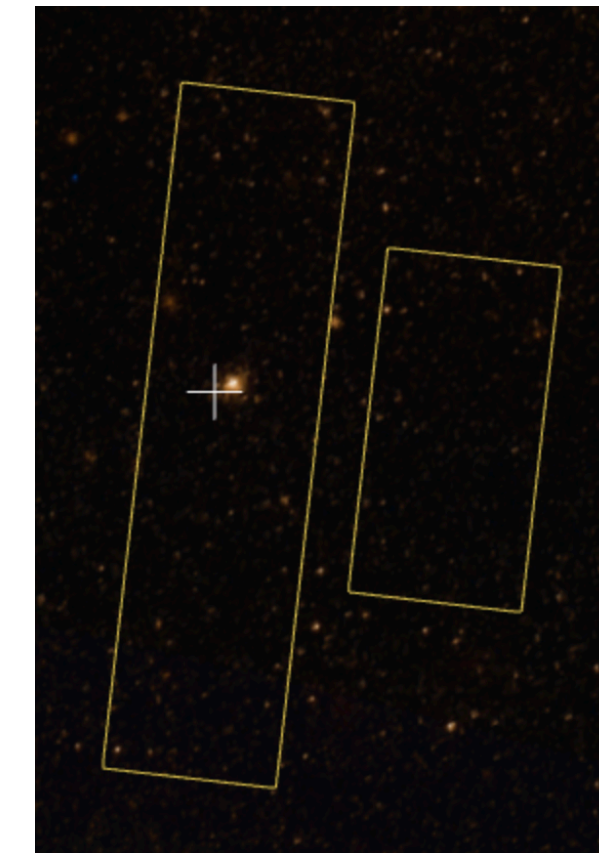
The Chandra strategy for the adoption of DOIs has been determined by

- needs of a well established archive (22 years in the business) that had adopted IVO style Persistent Identifiers since its inception
- new opportunities provided by DOI technology
- emergence of new types of data aggregation and “vertical” development of the Chandra data

The CDA DOI strategy

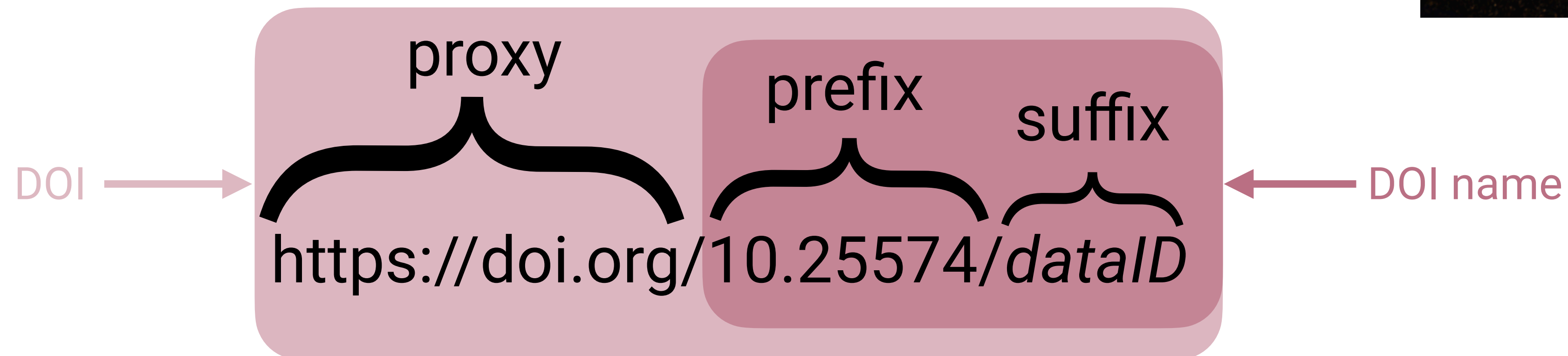
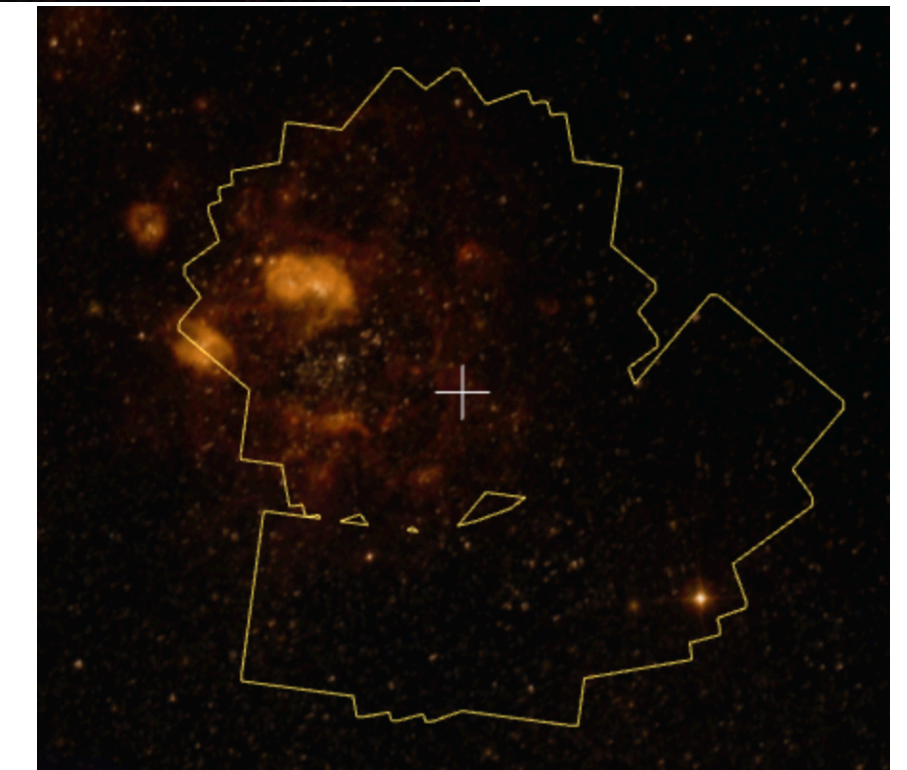
→ Archival observations DOIs

- Distinct DOIs for each distinct observation
 - DOI based on *ObsID* (<https://doi.org/10.25574/23336>)
 - landing page for DOIs data objects generated through ChaSeR



→ Chandra Source Catalog DOIs

- one single DOI for each major version of the CSC,
- Distinct DOIs for full-field data products for CSC stacks and observations
 - CSC Stacks: [10.25574/csc2.stk.acisfJ1509217p073311_001](https://doi.org/10.25574/csc2.stk.acisfJ1509217p073311_001)
 - CSC ObsIDs: [10.25574/csc2.obs.18225](https://doi.org/10.25574/csc2.obs.18225)



- The Smithsonian Institution is a DataCite member and CDA (part of SAO) can mint DOIs with the prefix 10.25574
 - Archival observations: **~24k** DOIs so far
 - average number of new DOIs: **~1,000/year**
 - CSC2.1 (finalizing processing): **~10k** CSC stack DOIs, **~15k** CSC ObsIDs DOIs

DOIs metadata (~70 defined for Archival/CSC ObsIDs DOIs, ~50 for CSC stacks DOIs) are matched with properties of data entities

- Requirements on upkeep of metadata
 - **“one-and-done”** metadata
 - observational/data objects metadata that won't change over time...
 - ...or change seldom
 - **continuously updated metadata**
 - literature objects that keep using the same data products
 - new types/level of aggregations of basic data products

The CDA DOI strategy

→ Archival observations DOIs

- Distinct DOIs for each distinct observation
 - DOI based on *ObsID* (<https://doi.org/10.25574/23336>)
 - landing page for DOIs data objects generated through ChaSeR

→ Chandra Source Catalog DOIs

- One single DOI for each major version of the CSC,
- Distinct DOIs for full-field data products for CSC stacks and observations
 - CSC Stacks: 10.25574/csc2.stk.acisfJ1509217p073311_001
 - CSC ObsIDs: 10.25574/csc2.obs.18225

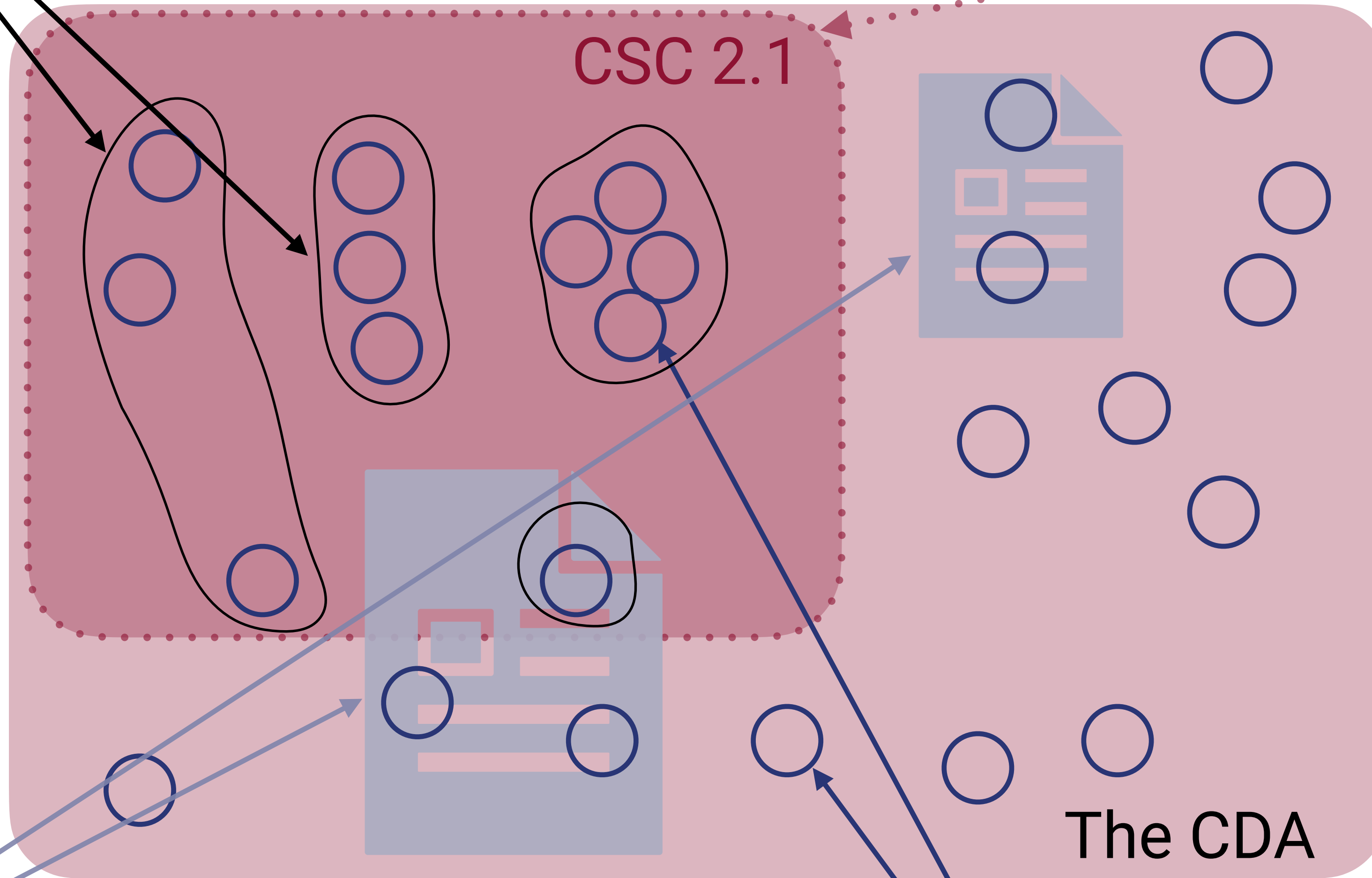
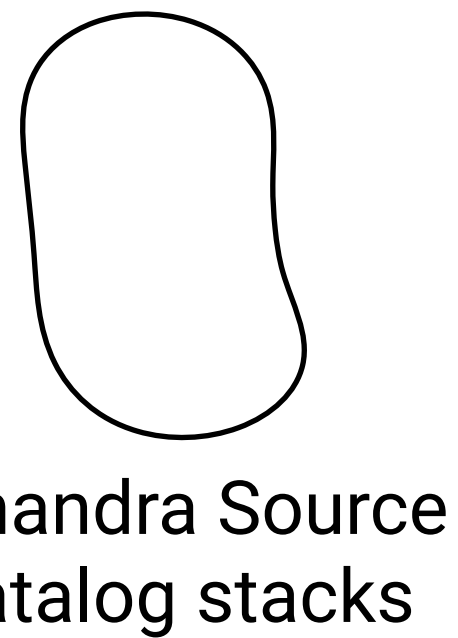
→ Chandra Data Collection (CDC) DOIs

- DOIs expressing an arbitrary collection of unitary Chandra data entities
 - **Usually** associated with a publication
 - **Usually** created based on users' requests

The CDA DOI strategy

10.25574/csc<version_number>.<stack_id>

10.25574/csc<version_number>

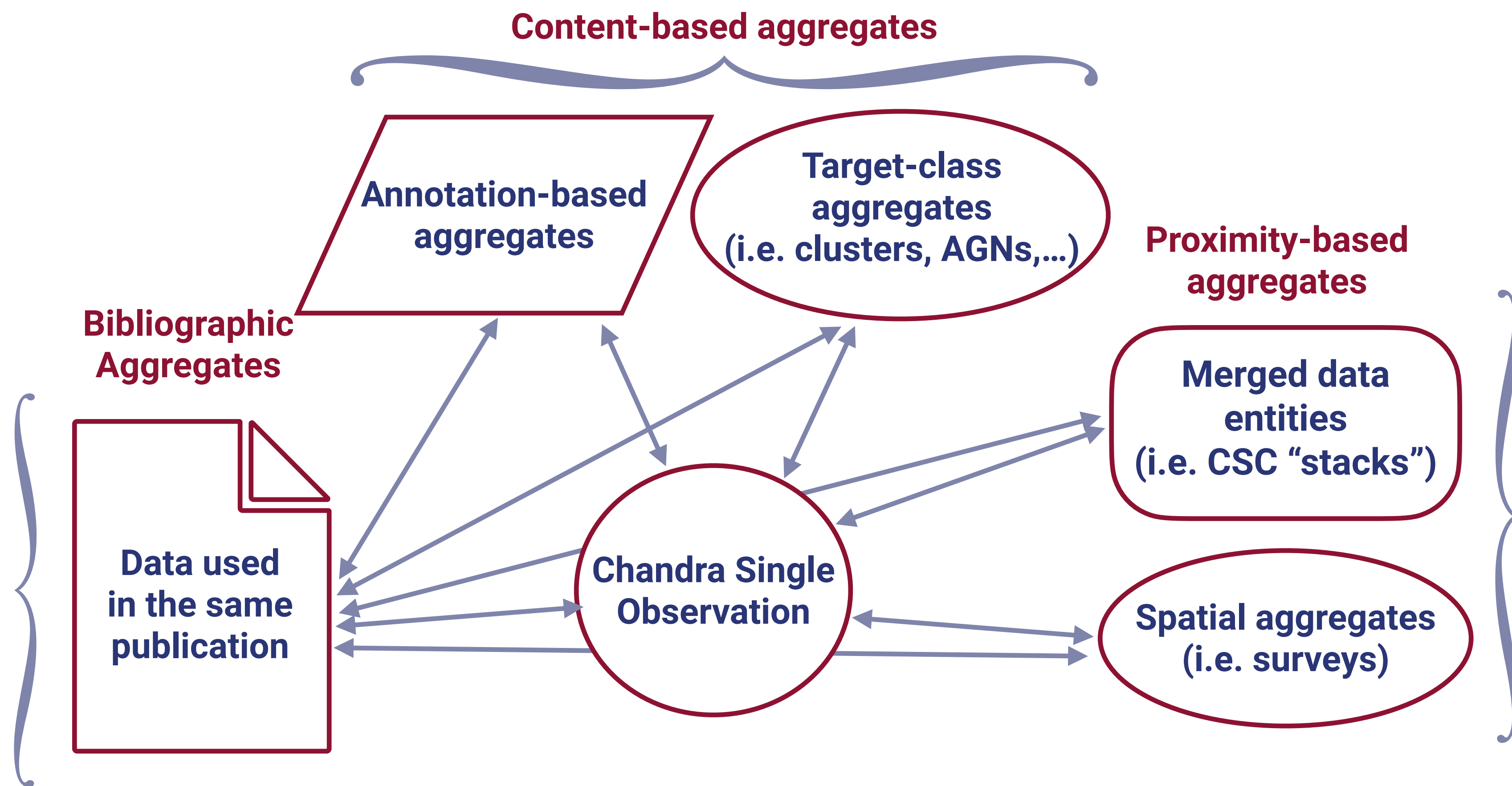


10.25574/cdc.<collection_id>

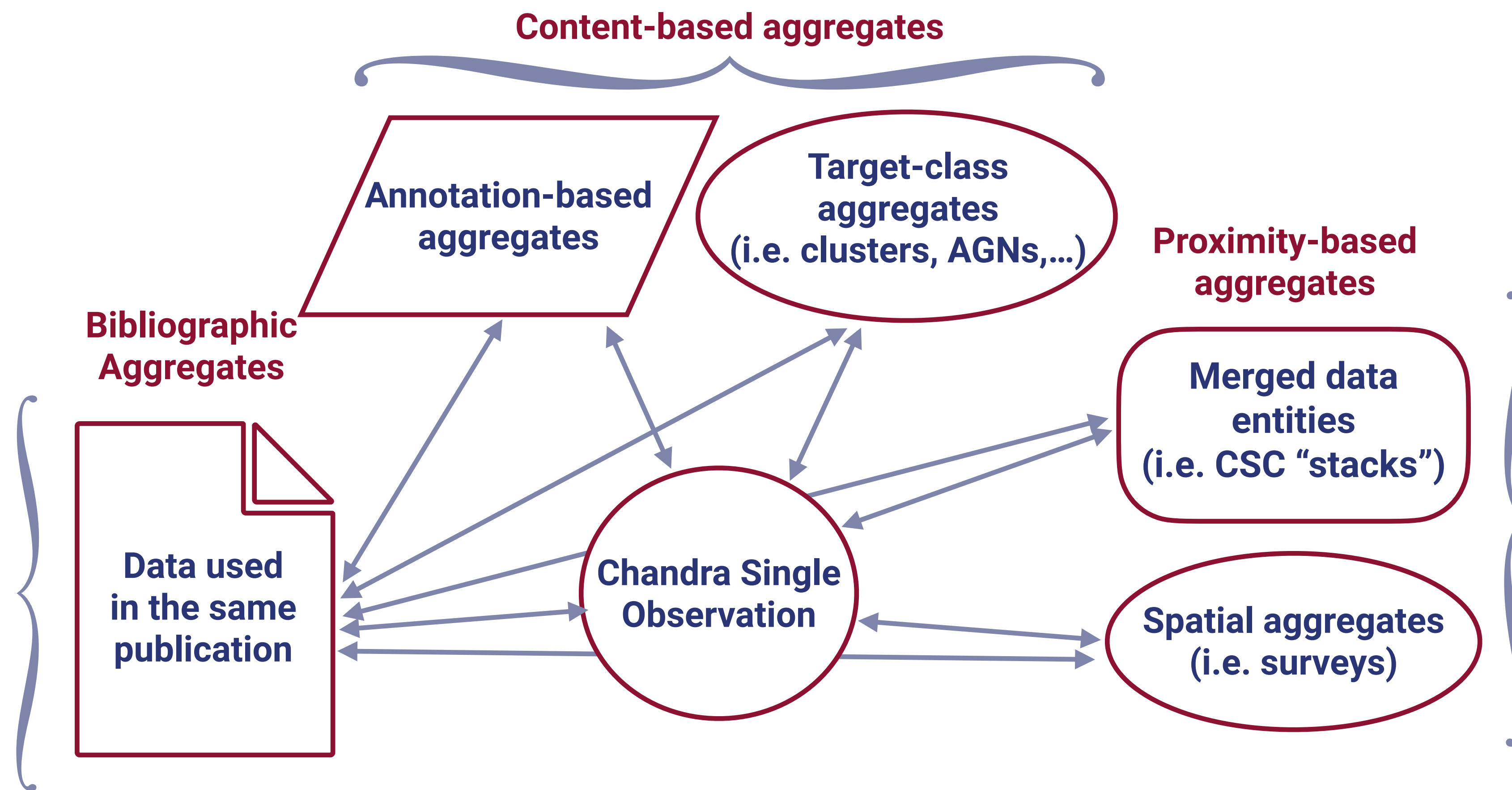
10.25574/<ObsID>

The CDA

Aggregations of Chandra data

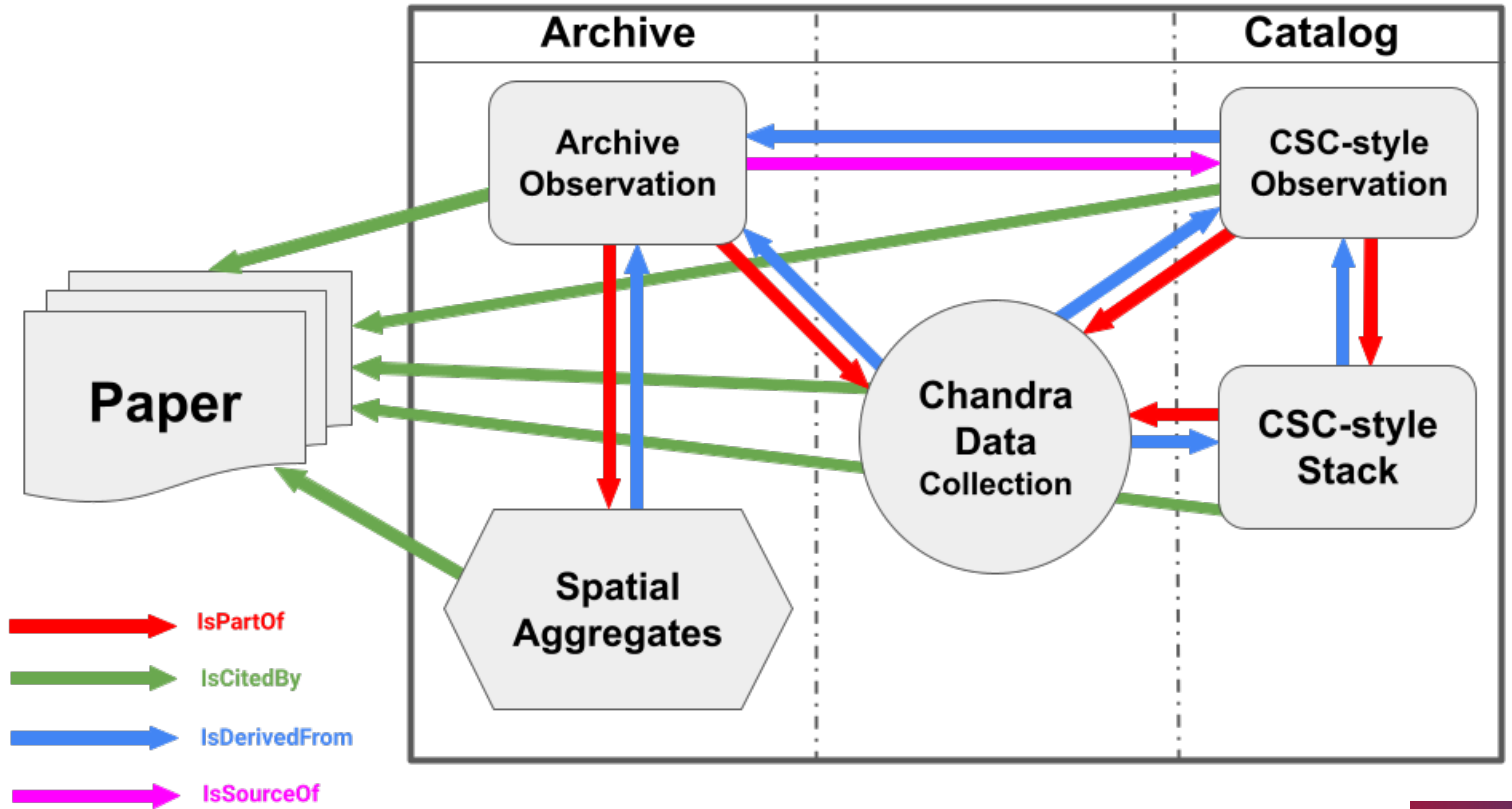


Aggregations of Chandra data



The CDC DOIs provide the most flexible tool to formally represent all classes of aggregation of all Chandra data entities, and this provides a path to a **formal, public, abstract description of the internal structure** of the *Chandra* data holdings and their impact on bibliography

D0Is and Aggregation



DOLs and Aggregation

The rich set of **RelationTypes** for RelatedIdentifiers available under the Datacite DOLs metadata schema (4.*) provides a great opportunity to mirror the growing complexity of Chandra and other data, and express the various relationships between data objects.

12.b	relationType	1	Description of the relationship of the resource being registered (A) and the related resource (B)	If RelatedIdentifier is used, relationType is mandatory. <i>Controlled List Values:</i> IsCitedBy Cites IsSupplementTo IsSupplementedBy IsContinuedBy Continues IsDescribedBy Describes HasMetadata IsMetadataFor HasVersion IsVersionOf IsNewVersionOf IsPreviousVersionOf IsPartOf HasPart IsPublishedIn IsReferencedBy References IsDocumentedBy Documents IsCompiledBy Compiles IsVariantFormOf IsOriginalFormOf IsIdenticalTo IsReviewedBy Reviews IsDerivedFrom IsSourceOf IsRequiredBy Requires IsObsoletedBy
------	--------------	---	---	--

The question(s)

Are we using the ***correct relational identifiers*** for astronomical data? DataCite definitions are very broad but the semantics of relationships between data should aim to be generally accepted and unambiguous.

How to allow the citability of subsets of the Chandra Source Catalogs? We are evaluating how much effort it would take to start assigning DOIs to queries.



Thank you!