GERMAN ASTROPHYSICAL
GAVO
VIRTUAL OBSERVATORY

Fig. 1

Federal Ministry
of Education
and Research

Fig. 2

1. DOIs and the Registry

(cf. Fig. 1)

Markus Demleitner
msdemlei@ari.uni-heidelberg.de

(cf. Fig. 2)

This is my personal view on how we should proceed with DOIs in the VO.

- What's the plan?
- What's not?
- The story so far
- A program

2. What's the plan?

If you're in the Registry, you have a DOI
(so proper data publication becomes a desirable thing)

3. Reasoning

- To provide an incentive for publishing data, it must be citeable
- IVOIDs might have worked, but we didn't push that. DOIs have street cred already, so there's little pushing required
- Not everyone publishing in the VO should need to make a contract with Datacite
- A well-maintained mapping IVOID ↔ DOI would immediately bring the entire VO into the DOIverse.

4. More detailed plan

- Map VOResource to Datacite Metadata (DCMeta); map IVORNs to DOIs (plus collision detection)
- Have something monitor two or three global publishing OAI-PMH endpoints (or harvest separately?)
- Convert and upload new records, do something smart with deletions
- Invent some magic to get the DOI back to the publishing registry.

5. What this is not

This is *not* about citing individual datasets.

I claim there is nothing we, as a VO, can sensibly do about those, as publishers can pull them at will, and we can't even give weak guarantees for persistence there.

I *think* there's nothing wrong with individual publishers minting DOIs for datasets contained in VO services (but we'll have to work this out with DataCite).

6. Mapping Problems: general

I have tried to map VOResource to Datacite Metadata.

https://volute.g-vo.org/svn/trunk/projects/registry/dois

Problems and suggested solutions:

- DCMeta: `creator` 1+ vs. VOResource `creator` 0+ – Solution: Use `{:unap}`, `{:unav}`, `{:tba}` machine readable NULLs
- DCMeta: `(Title,Subject,Description)@xml:lang` vs. VOResource implicit English – Solution: make English required in VOResource and insert the language declaration? Have `@lang` in VOResource, too? Ignore the issue?

7. Mapping Problems: creator

DCMeta wants parsed authors in separate `creatorName` elements. VOResource in principle concurs, but look at the mess people are dumping into `creator/name`:

```
Kornilov, V.G., Volkov, I.M., Zakharov, A.I., Kozyreva, V.S., Kornil...
NRAO, generated by J.J. Condon, J.J. Broderick and G.A. Seielstad, D...
CompTel Instrument Team.  Maps generated \n\n by Andrew Strong, Max-...
Helio
The Two Micron All Sky Survey is a joint project of the University o...
Data taken by ROE, AAO, and CalTech, Compression\nand distribution b...
All data is downloaded from the <a href=http://galex.stsci.edu>\nMAS...
Max Planck Institute for Exterrestrial Physics (Garching FRG)
PI: Ata Sarajedini
ASB branch STScI
Akylas et al.
```

Solution: No idea. It'd be nice to get this fixed for VO data discovery, too.

8. Mapping Problems: date terms

- DCMeta `Date/@dateType` with about 10 terms vs. VOResource `date/@role` with "creation" and "update" as part of an open vocabulary. – Solution: Copy the controlled vocabulary from DCMeta?

- DCMeta: `PublicationYear` vs. VOResource `@created` – Solution: Define that the metadata is what the DOI really references? Will Datacite let us do that? Use `curation/date` with a special `@role`?

9. Mapping Problems: more terms

- DCMeta: "fine-grained" `ResourceType` vs. VOResource `capability` (?) – Solution: have AstroObjects/AstroImage/Spectrum/Lines in ResourceType based on `capability/@standardId`

- DCMeta: `@contributorType` vs. VOResource `ResourceName`-typed `contributor` – Solution: add `contributorType` with vocabulary to VOResource

- DCMeta: `resourceTypeGeneral` vs. VOResource `Resource/@xsi:type` – Solution: ad-hoc mapping of types to DCMeta Collection, Dataset, Service, or Other

10. Mapping relationship terms

VOR relationshipType

1. mirror-of
2. service-for
3. served-by
4. derived-from
5. related-to
6. ?

DCMeta relatedType

1. IsIdenticalTo
2. ?
3. ?
4. IsDerivedFrom
5. appropriate null value?
6. Almost 30 terms Specifically, DCMeta 3.1 lists IsCitedBy Cites IsSupplementTo IsSupplementedBy IsContinuedBy Continues HasMetadata IsMetadataFor IsNewVersionOf IsPreviousVersionOf IsPartOf HasPart IsReferencedBy References IsDocumentedBy Documents IsCompiledBy Compiles IsVariantFormOf IsOriginalFormOf IsReviewedBy Reviews IsSourceOf– since we don't intend to turn DCMeta into VOResource, there's nothing wrong with having no matches for those. Maybe we should investigate if some are relevant for us, too?

11. Mapping Problems: URIs

- DCMeta `relatedIdentifier` vs. VOResource `relationship`: Keep ivoid? Translate to DOI?

- `relatedIdentifierType`='URL' for IVOIDs? Can we get a term for us?

- DCMeta `AlternateIdentifier` has `@alternateIdentifierType` that we'd have to define. I'd say IVOID or ivoid.

12. Mapping Problems: rights

DCMeta: freetext `rights` plus `rightsURI` vs. VOResource `rights` with values "public", "secure", and "proprietary".

Solution: release VOResource `rights` content limitations and add `@rights-uri`

13. altIdentifier in VOResource

- To keep the DOI in the Registry record, `vr:Resource` should receive a new `altIdentifier` element.

- DCMeta allows identification of authors using ORCID – in VOResource, we could allow `altIdentifier` in `curation/ceator`, too.

14. Revised plan

- Figure out how far Datacite will indulge us
- Uniform landing page: VOResource metadata plus style sheet? referenceURL?
- Prepare VOResource 1.1 (schema change?)
- Write "DOI minter"
- **Figure out how to inject DOIs into the Registry records**
Help!