



# DATA PROVENANCE IN THE VO

JUAN DE DIOS SANTANDER VELA

INSTITUTO DE ASTROFÍSICA DE ANDALUCÍA-CSIC

IN COLLABORATION WITH THE SVO



*and who where they?*

# WHAT DID THEY DO TO YOU, PHOTON?

JUAN DE DIOS SANTANDER VELA

INSTITUTO DE ASTROFÍSICA DE ANDALUCÍA-CSIC

IN COLLABORATION WITH THE SVO

# TALK OVERVIEW

---

- Who am I, and how did I get here?
- Data Provenance
- The Observation Data Model
- RADAMS, and Data Provenance Proposals
- Review & Conclusions

# WHO AM I?

---

- Juan de Dios Santander Vela, with the AMIGA international collaboration
- Electronics Engineer, Software Engineer, come... what?
- Should get my Ph.D. in Radio Astronomy & VO by year's end!
- Collaborating with the SVO in archive development
  - Robledo de Chavela's DSS-63 Scientific Archive
  - IRAM 30m Archive (TAPAS)

# WHO AM I?

---

- Juan de Dios Santander Vela, with the AMIGA international collaboration
- Electronics Engineer, Software Engineer, come... what?
- Should get my Ph.D. in Radio Astronomy & VO by year's end!
- Collaborating with the SVO in archive development
  - Robledo de Chavela's DSS-63 Scientific Archive
  - IRAM 30m Archive (TAPAS)

# WHO AM I?

---

- Juan de Dios Santander Vela, with the AMIGA international collaboration
- Electronics Engineer, Software Engineer, come... what?
- Should get my Ph.D. in Radio Astronomy & VO by year's end! *shameless plug + disclaimer*
- Collaborating with the SVO in archive development
  - Robledo de Chavela's DSS-63 Scientific Archive
  - IRAM 30m Archive (TAPAS)

# WHAT'S AMIGA?

---

- Analysis of the interstellar Medium of Isolated GALaxies
- Based on Karachentseva's CIG/KIG Catalogue of Isolated Galaxies
- $N \approx 1000$ , complete sample
- Multi-wavelength, with emphasis in mm/submm (molecular gas)
- Stringent data analysis/physical parameter derivation
  - VERY database dependent!
  - VO Interface

# AMIGA WEB: <http://amiga.iaa.es/>

Instituto de Astrofísica de Andalucía

## AMIGA

Analysis of the Interstellar Medium of Isolated GALAXIES



Search

Home

Science

Technical development

Team and collaborators

Results & Ongoing Work

Publications

Conferences

Public Data

VO Interface

ASCII Files

Links & Tools

### VO Interface

Search by name | [Search by parameters](#)

Object Name

(Ex: CIG 4, UGC 00297,CIG 4%, etc..)

Or/and Input a File

ningún archivo seleccionado (Text file with a name per line)

### Output Options

Fields

Coords  Optical  Velocity  Morphology  FIR  Isolation

Equinox

Image preview?  Yes  No



# AMIGA WEB: <http://amiga.iaa.es/>

Instituto de Astrofísica de Andalucía

## AMIGA

Analysis of the Interstellar Medium of Isolated GALAXIES



Search

Home

Science

Technical development

Team and collaborators

Results & Ongoing Work

Publications

Conferences

Public Data

VO Interface

ASCII Files

Links & Tools

### VO Interface

[Search by name](#) | **Search by parameters**

### Coordinates

Choose Epoch

Choose a circular region

RA  RA:20h43m23s, RA:20 43 45, RA:32.3456

DEC  DEC:20d43m23s, DEC:20 43 45, DEC:32.3456

Search radius

Or choose a range

< RA <

# AMIGA WEB: <http://amiga.iaa.es/>

## AMIGA PUBLIC DATABASE SEARCH RESULTS

Objects Found: 14

### Download Data

Data File: [VOTable](#) [ASCII](#) [Load Data with Javaws Tool](#) [Topcat](#) [Aladin](#) [Aladin applet](#)

### Data

(\*) Please, read note about Morphology

Name	Coords		Optical			Velocity			Morphology (*)				FIR										Isolation																		
CIG Number	RA J2000	DEC J2000	MB	mB-corr	LB	Vr	V3K	Distance	Morph	Morph rc3	conf morph	Bar	Int	F12	UL12	rms12	M12	E12	F25	UL25	rms25	M25	E25	F60	UL60	rms60	M60	E60	F100	UL100	rms100	M100	E100	LFIR	ULFIR	nk	k	Q	QKar	Q05	Q05Kar
<a href="#">CIG 118</a>	41.93608	9.3566	20.97	14.126	10.34	8055	7834	104.4	E/S0	-3				0.084	1	0.028	5	2	0.105	1	0.035	5	2	0.147	1	0.049	5	2	1.200	1	0.400	5	2	10.053	1	0.543	5	3.843	3.843	3.878	-3.878
<a href="#">CIG 128</a>	52.86433	4.3809	20.68	13.990	10.22	6596	6437	85.8	E/S0	-3				0.078	1	0.026	5	2	0.108	1	0.036	5	2	0.190	0	0.062	1	0	1.038	1	0.346	5	2	9.866	1	0.587	5	3.609	4.119	3.642	-4.193
<a href="#">CIG 264</a>	129.00616	30.2662	20.31	14.814	10.07	7715	7939	105.9	E/S0	-3				0.078	1	0.026	5	2	0.138	1	0.046	5	2	0.147	1	0.049	5	2	0.306	1	0.102	5	2	9.703	1	1.505	5	3.332	3.332	3.420	-3.420
<a href="#">CIG 358</a>	145.10037	14.9225	20.42	13.286	10.12	3821	4137	55.2	E/S0	-3				0.099	1	0.033	5	2	0.153	1	0.051	5	2	0.114	1	0.038	5	2	0.318	1	0.106	5	2	9.088	1	1.279	5	2.126	2.814	2.147	-2.862
<a href="#">CIG 415</a>	156.38000	7.4627	20.82	15.216	10.28	11728	12080	161.1	E/S0	-3				0.108	1	0.036	5	2	0.165	1	0.055	5	2	0.126	1	0.042	5	2	0.276	1	0.092	5	2	10.011	1	0.669	5	3.748	3.748	4.056	-4.056
<a href="#">CIG 517</a>	180.71729	26.2529	20.35	15.259	10.09	9633	9935	132.5	E/S0	-3				0.063	1	0.021	5	2	0.096	1	0.032	5	2	0.120	0	0.036	1	0	0.294	1	0.098	5	2	9.820	1	1.346	5	2.971	2.971	3.002	-3.002
<a href="#">CIG 529</a>	186.24083	24.3828	20.39	15.352	10.11	10245	10545	140.6	E/S0	-3				0.099	1	0.033	5	2	0.147	1	0.049	5	2	0.153	1	0.051	5	2	0.342	1	0.114	5	2	9.981	1	1.235	5	1.949	1.949	1.953	-1.953
<a href="#">CIG 599</a>	207.14400	37.1131	21.03	14.691	10.36	10248	10449	139.3	E/S0	-3				0.096	1	0.032	5	2	0.138	1	0.046	5	2	0.099	1	0.033	5	2	0.255	1	0.085	5	2	9.813	1	1.747	5	2.951	3.160	3.198	-3.407
<a href="#">CIG 670</a>	229.87566	67.5046	21.30	14.809	10.47	12477	12502	166.7	E/S0	-3				0.051	1	0.017	5	2	0.066	1	0.022	5	2	0.105	1	0.035	5	2	0.282	1	0.094	5	2	10.004	1	1.021	5	2.704	2.704	2.761	-2.761
<a href="#">CIG 730</a>	244.29566	4.5541		15.382					E/S0	-3				0.057	1	0.019	5	2	0.087	1	0.029	5	2	0.132	1	0.044	5	2	0.351	1	0.117	5	2		9	0.340	5	3.728	3.728		
<a href="#">CIG 735</a>	244.69029	39.3347	20.09	15.427	9.99	9474	9511	126.8	E/S0	-3				0.060	1	0.020	5	2	0.054	1	0.018	5	2	0.057	1	0.019	5	2	0.237	1	0.079	5	2	9.608	1	1.345	5	3.142	3.142	3.312	-3.312

# AMIGA WEB: <http://amiga.iaa.es/>

## AMIGA PUBLIC DATABASE SEARCH RESULTS

Objects Found: 14

### Download Data

Data File: [VOTable](#) [ASCII](#) [Load Data with Javaws Tool](#) [Topcat](#) [Aladin](#) [Aladin applet](#)

### Data

(\*) Please, read note about Morphology

Name	Coords		Optical			Velocity			Morphology (*)				FIR										Isolation																		
CIG Number	RA J2000	DEC J2000	MB	mB-corr	LB	Vr	V3K	Distance	Morph	Morph rc3	conf morph	Bar	Int	F12	UL12	rms12	M12	E12	F25	UL25	rms25	M25	E25	F60	UL60	rms60	M60	E60	F100	UL100	rms100	M100	E100	LFIR	ULFIR	nk	k	Q	QKar	Q05	Q05Kar
<a href="#">CIG 118</a>	41.93608	9.3566	20.97	14.126	10.34	8055	7834	104.4	E/S0	-3				0.084	1	0.028	5	2	0.105	1	0.035	5	2	0.147	1	0.049	5	2	1.200	1	0.400	5	2	10.053	1	0.543	5	3.843	3.843	3.878	-3.878
<a href="#">CIG 128</a>	52.86433	4.3809	20.68	13.990	10.22	6596	6437	85.8	E/S0	-3				0.078	1	0.026	5	2	0.108	1	0.036	5	2	0.190	0	0.062	1	0	1.038	1	0.346	5	2	9.866	1	0.587	5	3.609	4.119	3.642	-4.193
<a href="#">CIG 264</a>	129.00616	30.2662	20.31	14.814	10.07	7715	7939	105.9	E/S0	-3				0.078	1	0.026	5	2	0.138	1	0.046	5	2	0.147	1	0.049	5	2	0.306	1	0.102	5	2	9.703	1	1.505	5	3.332	3.332	3.420	-3.420
<a href="#">CIG 358</a>	145.10037	14.9225	20.42	13.286	10.12	3821	4137	55.2	E/S0	-3				0.099	1	0.033	5	2	0.153	1	0.051	5	2	0.114	1	0.038	5	2	0.318	1	0.106	5	2	9.088	1	1.279	5	2.126	2.814	2.147	-2.862
<a href="#">CIG 415</a>	156.38000	7.4627	20.82	15.216	10.28	11728	12080	161.1	E/S0	-3				0.108	1	0.036	5	2	0.165	1	0.055	5	2	0.126	1	0.042	5	2	0.276	1	0.092	5	2	10.011	1	0.669	5	3.748	3.748	4.056	-4.056
<a href="#">CIG 517</a>	180.71729	26.2529	20.35	15.259	10.09	9633	9935	132.5	E/S0	-3				0.063	1	0.021	5	2	0.096	1	0.032	5	2	0.120	0	0.036	1	0	0.294	1	0.098	5	2	9.820	1	1.346	5	2.971	2.971	3.002	-3.002
<a href="#">CIG 529</a>	186.24083	24.3828	20.39	15.352	10.11	10245	10545	140.6	E/S0	-3				0.099	1	0.033	5	2	0.147	1	0.049	5	2	0.153	1	0.051	5	2	0.342	1	0.114	5	2	9.981	1	1.235	5	1.949	1.949	1.953	-1.953
<a href="#">CIG 599</a>	207.14400	37.1131	21.03	14.691	10.36	10248	10449	139.3	E/S0	-3				0.096	1	0.032	5	2	0.138	1	0.046	5	2	0.099	1	0.033	5	2	0.255	1	0.085	5	2	9.813	1	1.747	5	2.951	3.160	3.198	-3.407
<a href="#">CIG 670</a>	229.87566	67.5046	21.30	14.809	10.47	12477	12502	166.7	E/S0	-3				0.051	1	0.017	5	2	0.066	1	0.022	5	2	0.105	1	0.035	5	2	0.282	1	0.094	5	2	10.004	1	1.021	5	2.704	2.704	2.761	-2.761
<a href="#">CIG 730</a>	244.29566	4.5541		15.382					E/S0	-3				0.057	1	0.019	5	2	0.087	1	0.029	5	2	0.132	1	0.044	5	2	0.351	1	0.117	5	2		9	0.340	5	3.728	3.728		
<a href="#">CIG 735</a>	244.69029	39.3347	20.09	15.427	9.99	9474	9511	126.8	E/S0	-3				0.060	1	0.020	5	2	0.054	1	0.018	5	2	0.057	1	0.019	5	2	0.237	1	0.079	5	2	9.608	1	1.345	5	3.142	3.142	3.312	-3.312

# DATA PROVENANCE IN E-SCIENCE

# WHAT IS E-SCIENCE

---

- Not “electronic” science
- E stands for Enhanced!
- Via Network Effects

# WHAT IS E-SCIENCE

---

- Not “electronic” science
- E stands for Enhancement
- Via Network Effect

Money is nothing,  
but a lot of money,  
that's quite another  
matter.

George Bernard Shaw

# WHAT IS E-SCIENCE

---

- Not “electronic” science
- E stands for Enhancement
- Via Network Effect

A datum is nothing,  
but lots of data,  
that's quite another  
matter.

Modern (e-)Scientist

# WHAT IS E-SCIENCE

---

- Not “electronic” science
- E stands for Enhancement
- Via Network Effects

A datum is nothing,  
but lots of data,  
that's quite another  
matter.

Modern (e-)Scientist

Corollary: the Virtual Observatory is e-Science.



# WHAT IS DATA PROVENANCE

---

provenance |'prävənəns|

noun

the place of origin or earliest known history of something : *an orange rug of Iranian provenance.*

- the beginning of something's existence; something's origin : *they try to understand the whole universe, its provenance and fate.*

See note at **origin** .

- a record of ownership of a work of art or an antique, used as a guide to authenticity or quality : *the manuscript has a distinguished provenance.*

ORIGIN late 18th cent.: from French, from the verb **provenir** 'come or stem from,' from Latin **provenire**, from **pro-** 'forth' + **venire** 'come.'

# WHAT IS DATA PROVENANCE

---

provenance |'prävənəns|

noun

the place of origin or earliest known history of something : *an orange rug of Iranian provenance.*

- the beginning of something's existence; something's origin : *they try to understand the whole universe, its provenance and fate.*

See note at **origin** .

- a record of ownership of a work of art or an antique, used as a guide to authenticity or quality : *the manuscript has a distinguished provenance.*

ORIGIN late 18th cent.: from French, from the verb **provenir** 'come or stem from,' from Latin **provenire**, from **pro-** 'forth' + **venire** 'come.'

# WHAT IS DATA PROVENANCE

---

provenance |'prävənəns|

noun

the place of origin or earliest known history of something : *an orange rug of Iranian provenance.*

- the beginning of something's existence; something's origin : *they try to understand the whole universe, its provenance and fate.*

See note at **origin** .

- a record of ownership of a work of art or an antique, used as a guide to authenticity or quality : *the manuscript has a distinguished provenance.*

ORIGIN late 18th cent.: from French, from the verb **provenir** 'come or stem from,' from Latin **provenire**, from **pro-** 'forth' + **venire** 'come.'

# WHAT IS DATA PROVENANCE

---

provenance |'prävənəns|

noun

the place of origin or earliest known history of something : *an orange rug of Iranian provenance.*

- the beginning of something's existence; something's origin : *they try to understand the whole universe, its provenance and fate.*

See note at **origin** .

- a record of ownership of a work of art or an antique, used as a guide to authenticity or quality : *the manuscript has a distinguished provenance.*

ORIGIN late 18th cent.: from French, from the verb *provenir* 'come or stem from,' from Latin *provenire*, from *pro-* 'forth' + *venire* 'come.'

# WHAT IS DATA PROVENANCE

---

- Data from heterogeneous sources
- Different things to trace:
  - Attribution, Data Quality, and Replication
  - For legally binding databases, Audit Trail
  - In many disciplines, just informational

# WHAT IS DATA PROVENANCE

## *Consultative Committee for Space Data Systems*

RECOMMENDATION FOR SPACE  
DATA SYSTEM STANDARDS

### Reference Model for an Open Archival Information System (OAIS)

CCSDS 650.0-B-1  
BLUE BOOK

January 2002

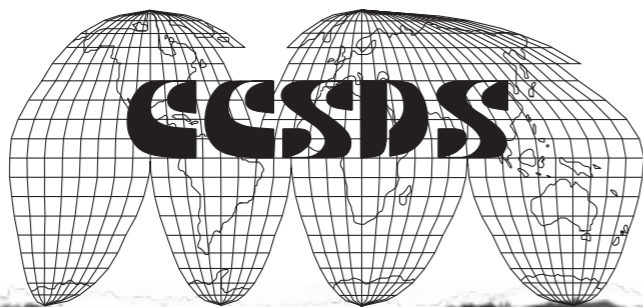
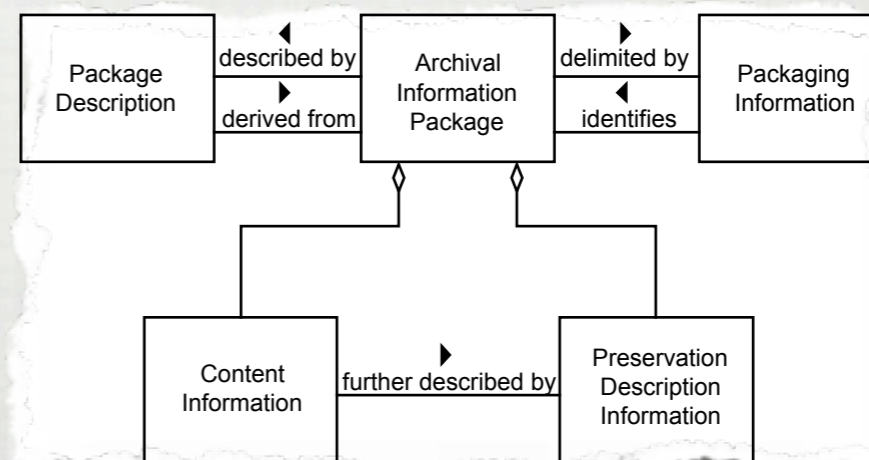


Table 4-1: Examples of PDI Types

Content Information Type	Reference	Provenance	Context	Fixity
Space Science Data	<ul style="list-style-type: none"> <li>Object identifier</li> <li>Journal reference</li> <li>Mission, instrument, title, attribute set</li> </ul>	<ul style="list-style-type: none"> <li>Instrument description</li> <li>Processing history</li> <li>Sensor description</li> <li>Instrument</li> <li>Instrument mode</li> <li>Decommutation map</li> <li>Software interface specification</li> </ul>	<ul style="list-style-type: none"> <li>Calibration history</li> <li>Related data sets</li> <li>Mission</li> <li>Funding history</li> </ul>	<ul style="list-style-type: none"> <li>CRC</li> <li>Checksum</li> <li>Reed-Solomon coding</li> </ul>
	• Bibliographic	• For scanned	• Pointers to related	• Digital



	<b>LIP</b>	<b>Chimera</b>	<b>myGRID</b>	<b>CMCS</b>	<b>PASOA</b>	<b>ESSW</b>	<b>Tioga</b>	<b>Buneman</b>	<b>Trio</b>
<i>Domain</i>	GIS	Physics, Astronomy	Biology	Chemistry	Biology	Earth Sciences	Atmospheric Sciences	Generic	Generic
<i>Processing</i>	Commands	Services	Services	Services	Services	Scripts	RDBMS	RDBMS	RDBMS
<i>Application</i>	IR	IARP	IR	IU	IR	I	IE	IU	IU
<i>Orientation</i>	Data	Process	Process	Data	Process	Data & Process	Data	Data	Data
<i>Granularity</i>	Spatial layers	Abstract datasets (...)	Abstract resources	Files	Workflow Parameters	Files	DB Attributes	DB Attributes & Tuples	DB Tuples
<i>Representation</i>	Ar...							Inverse queries	Inverse queries
<i>Semantics</i>	No	No	Yes	Limited	No	Proposed	No	No	No
<i>Storage</i>	RDBMS	RDBMS	RDBMS	RDBMS	RDBMS + File System	RDBMS	RDBMS	N/A	RDBMS
<i>Overhead</i>	user commands; MD entry	user definition; automatic WF	service semantics; WF calls	service calls; working portal	manual	manual	manual registry of inverse functions	N/A	Automatic generation of inverse queries
<i>Scalability</i>	No	Yes	No	No	Proposed	Proposed	Yes	N/A	No
<i>Dissemination</i>	Queries	Queries	Semantic browser; Lineage graph	Queries; Browsing	Queries	Browsing	Queries; Graph	N/A	Queries

Based on *A Survey of Data Provenance in e-Science*,  
Simmhan et al, SIGMOD Record 34, 3, Sept. 2005

I Informational  
 R Regeneration  
 A Audit  
 E Error Tracking  
 U Information Update  
 P Planning

	<i>LIP</i>	<i>Chimera</i>	<i>myGRID</i>	<i>CMCS</i>	<i>PASOA</i>	<i>ESSW</i>	<i>Tioga</i>	<i>Buneman</i>	<i>Trio</i>
<i>Domain</i>	GIS	Physics, Astronomy	Biology	Chemistry	Biology	Earth Sciences	Atmospheric Sciences	Generic	Generic
<i>Processing</i>	Commands	Services	Services	Services	Services	Scripts	RDBMS	RDBMS	RDBMS
<i>Application</i>	IR	IARP	IR	IU	IR	I	IE	IU	IU
<i>Orientation</i>	Data	Process	Process	Data	Process	Data & Process	Data	Data	Data
<i>Granularity</i>	Spatial layers	Abstract datasets	Abstract resources	Files	Workflow Parameters	Files	DB Attributes	DB Attributes & Tuples	DB Tuples
<i>Representation</i>	Ar							Inverse queries	Inverse queries
<i>Semantics</i>	No	No	Yes	Limited	No	Proposed	No	No	No
<i>Storage</i>	RDBMS	RDBMS	RDBMS	RDBMS	RDBMS + File System	RDBMS	RDBMS	N/A	RDBMS
<i>Overhead</i>	user commands; MD entry	user definition; automatic WF	service semantics; WF calls	service calls; working portal	manual	manual	manual registry of inverse functions	N/A	Automatic generation of inverse queries
<i>Scalability</i>	No	Yes	No	No	Proposed	Proposed	Yes	N/A	No
<i>Dissemination</i>	Queries	Queries	Semantic browser; Lineage graph	Queries; Browsing	Queries	Browsing	Queries; Graph	N/A	Queries

Based on *A Survey of Data Provenance in e-Science*,  
 Simmhan et al, SIGMOD Record 34, 3, Sept. 2005



I Informational  
 R Regeneration  
 A Audit  
 E Error Tracking  
 U Information Update  
 P Planning

	<b>LIP</b>	<b>Chimera</b>	<b>myGRID</b>	<b>CMCS</b>	<b>PASOA</b>	<b>ESSW</b>	<b>Tioga</b>	<b>Buneman</b>	<b>Trio</b>
<i>Domain</i>	GIS	Physics, Astronomy	Biology	Chemistry	Biology	Earth Sciences	Atmospheric Sciences	Generic	Generic
<i>Processing</i>	Commands	Services	Services	Services	Services	Scripts	RDBMS	RDBMS	RDBMS
<i>Application</i>	IR	IARP	IR	IU	IR	I	IE	IU	IU
<i>Orientation</i>	Data	Process	Process	Data	Process	Data & Process	Data	Data	Data
<i>Granularity</i>	Spatial layers	Abstract datasets (files)	Abstract resources	Files	Workflow Parameters	Files	DB Attributes	DB Attributes & Tuples	DB Tuples
<i>Representation</i>	Annotations	Annotations	Annotations	Annotations	Annotations	Annotations	Inverse functions	Inverse queries	Inverse queries
<i>Semantics</i>	No	No	Yes	Limited	No	Proposed	No	No	No
<i>Storage</i>	RDBMS	RDBMS	RDBMS	RDBMS	RDBMS + File System	RDBMS	RDBMS	N/A	RDBMS
<i>Overhead</i>	user commands; MD entry	user definition; automatic WF	service semantics; WF calls	service calls; working portal	manual	manual	manual registry of inverse functions	N/A	Automatic generation of inverse queries
<i>Scalability</i>	No	Yes	No	No	Proposed	Proposed	Yes	N/A	No
<i>Dissemination</i>	Queries	Queries	Semantic browser; Lineage graph	Queries; Browsing	Queries	Browsing	Queries; Graph	N/A	Queries

# TRADITIONAL DATA PROVENANCE IN ASTROPHYSICS

```
SIMPLE = T /Standard FITS file
BITPIX = -32 /FITS data type
NAXIS = 4 /Dimensionality of array
NAXIS1 = 1024
NAXIS2 = 512
NAXIS3 = 1
NAXIS4 = 1
EXTEND = T /Extensions may be present
PCOUNT = 0 /Parameter count
GCTYPE = 1 /Group count
CTYPE1 = 'RA---SIN' /Axis name
CRPIX1 = 512.000000000000 /Reference pixel
CRVAL1 = 68.296229138839 /Reference value
CDELT1 = -8.3333331542464E-09 /Pixel increment
CROTA1 = 0.00000000000000 /Axis rotation
CTYPE2 = 'DEC--SIN' /Axis name
CRPIX2 = 257.000000000000 /Reference pixel
CRVAL2 = 5.3543381055544 /Reference value
CDELT2 = 8.3333331542464E-09 /Pixel increment
CROTA2 = 0.00000000000000 /Axis rotation
CTYPE3 = 'FREQ' /Axis name
CRPIX3 = 1.00000000000000 /Reference pixel
CRVAL3 = 22233458750.000 /Reference value
CDELT3 = 32000000.000000 /Pixel increment
CROTA3 = 0.00000000000000 /Axis rotation
CTYPE4 = 'STOKES' /Axis name
CRPIX4 = 1.00000000000000 /Reference pixel
CRVAL4 = 1.00000000000000 /Reference value
CDELT4 = 1.00000000000000 /Pixel increment
CROTA4 = 0.00000000000000 /Axis rotation
ORIGIN = 'AIPSLocalhost MACMAR 31DEC08' /Origin of data
DATE-OBS = '2007-11-07' /Observation date
TELESCOP = 'VLBA' /Telescope used
OBSERVER = 'BG182' /Observer name
OBJECT = '3C120' /Name of observed source
EPOCH = 2000.000000000000 /Equinox of coordinates
BSCALE = 1.00000000000000 /Scale factor of array
BZERO = 0.00000000000000 /Zero offset of array
BUNIT = 'JY/BEAM' /Unit of measurement
DATAMIN = -0.0019036284647882 /Min data value
DATAMAX = 1.1940656900406 /Max data value
BMAJ = 3.0594406858596E-07 /Clean beam major axis diameter (degrees).
BMIN = 9.1569899501081E-08 /Clean beam minor axis diameter (degrees).
BPA = -19.967098396954 /Clean beam position angle (degrees).
NITER = 1620 /Number of model components.
OBSRA = 68.296231370800 /Antenna pointing RA
OBSDEC = 5.3543386611100 /Antenna pointing Dec
NOISE = 0.00035333831328899 /Theoretical RMS noise estimate
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY EXTEND = T /Extensions may be present
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY EXTEND = T /This is the antenna file
HISTORY BLOCKED = T /Tape may be blocked
HISTORY / Where baseline = 256*ant1 + ant2 + (array#-1)/100
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY EXTEND = T / All data in tables
HISTORY BLOCKED = T / Tape may be blocked
```

```
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY FITLD EXTEND = T /
HISTORY FITLD BLOCKED = T /
HISTORY FITLD FXCORVER = '4.22' /
HISTORY FITLD OPENED FITS FILE : 2007NOV15 6h46m41.75s
HISTORY FITLD SOFTWARE DIR : /home/fxcorr/code
HISTORY FITLD CORREL SOFTWARE VERSION : 4.22
HISTORY FITLD CORREL ON LINE : Motorola MVME167
HISTORY FITLD VXWORKS VERSION : 5.4
HISTORY FITLD LOG FILE : /home/fxcorr/logs/v47c071115:0019.log
HISTORY FITLD OBSCODE : BG182
HISTORY FITLD JOBNUM : 6720
HISTORY FITLD JOBSTART : 2007NOV07 2h38m46.05s
HISTORY FITLD JOBSTOP : 2007NOV07 3h31m19.00s
HISTORY FITLD FILESTART : 2007NOV07 2h38m46.05s
HISTORY FITLD FILESTOP : 2007NOV07 3h31m19.00s
HISTORY /End FITS tape header "HISTORY" information
HISTORY FITLD RELEASE = '31DEC07' /-----
HISTORY FITLD OUTNAME = ' ' OUTCLASS = ' '
HISTORY FITLD OUTSEQ = 0 OUTDISK = 0
HISTORY FITLD RELEASE = '31DEC07' /-----
HISTORY FITLD SCLVIS = 0.54496 / Correlator scaling value
HISTORY FITLD DIGICOR = 1 / VLBA correlator digital corrs
HISTORY FITLD / FFT artifact corrections no longer necessary
HISTORY FITLD RELEASE = '31DEC07' /***** Start 16-NOV-2007 10:27:09
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY FITLD EXTEND = T /
HISTORY FITLD BLOCKED = T /
HISTORY FITLD FXCORVER = '4.22' /
HISTORY FITLD OPENED FITS FILE : 2007NOV15 8h15m22.70s
HISTORY FITLD SOFTWARE DIR : /home/fxcorr/code
HISTORY FITLD CORREL SOFTWARE VERSION : 4.22
HISTORY FITLD CORREL ON LINE : Motorola MVME167
HISTORY FITLD VXWORKS VERSION : 5.4
HISTORY FITLD LOG FILE : /home/fxcorr/logs/v47c071115:0809.log
HISTORY FITLD OBSCODE : BG182
HISTORY FITLD JOBNUM : 6721
HISTORY FITLD JOBSTART : 2007NOV07 3h31m21.08s
HISTORY FITLD JOBSTOP : 2007NOV07 4h36m16.00s
HISTORY FITLD FILESTART : 2007NOV07 3h31m21.08s
HISTORY FITLD FILESTOP : 2007NOV07 4h36m16.00s
HISTORY /End FITS tape header "HISTORY" information
HISTORY FITLD RELEASE = '31DEC07' /-----
HISTORY FITLD OUTNAME = 'BG182B' OUTCLASS = 'UVDATA'
HISTORY FITLD OUTSEQ = 1 OUTDISK = 0
HISTORY FITLD RELEASE = '31DEC07' /-----
HISTORY FITLD SCLVIS = 0.54496 / Correlator scaling value
HISTORY FITLD DIGICOR = 1 / VLBA correlator digital corrs
HISTORY FITLD / FFT artifact corrections no longer necessary
HISTORY FITLD RELEASE = '31DEC07' /***** Start 16-NOV-2007 10:27:22
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY FITLD EXTEND = T /
HISTORY FITLD BLOCKED = T /
HISTORY FITLD FXCORVER = '4.22' /
HISTORY FITLD OPENED FITS FILE : 2007NOV15 10h33m52.17s
HISTORY FITLD SOFTWARE DIR : /home/fxcorr/code
HISTORY FITLD CORREL SOFTWARE VERSION : 4.22
```

# TRADITIONAL DATA PROVENANCE IN ASTROPHYSICS

```
SIMPLE = T /Standard FITS file
BITPIX = -32 /FITS data type
NAXIS = 4 /Dimensionality of array
NAXIS1 = 1024
NAXIS2 = 512
NAXIS3 = 1
NAXIS4 = 1
EXTEND = T /Extensions may be present
PCOUNT = 0 /Parameter count
GCOUNT = 1 /Group count
CTYPE1 = 'RA---SIN' /Axis name
CRPIX1 = 512.000000000000 /Reference pixel
CRVAL1 = 68.296229138839 /Reference value
CDELT1 = -8.3333331542464E-09 /Pixel increment
CROTA1 = 0.00000000000000 /Axis rotation
CTYPE2 = 'DEC--SIN' /Axis name
CRPIX2 = 257.000000000000 /Reference pixel
CRVAL2 = 5.3543381055544 /Reference value
CDELT2 = 8.3333331542464E-09 /Pixel increment
CROTA2 = 0.00000000000000 /Axis rotation
CTYPE3 = 'FREQ' /Axis name
CRPIX3 = 1.00000000000000 /Reference pixel
CRVAL3 = 22233458750.000 /Reference value
CDELT3 = 32000000.000000 /Pixel increment
CROTA3 = 0.00000000000000 /Axis rotation
CTYPE4 = 'STOKES' /Axis name
CRPIX4 = 1.00000000000000 /Reference pixel
CRVAL4 = 1.00000000000000 /Reference value
CDELT4 = 1.00000000000000 /Pixel increment
CROTA4 = 0.00000000000000 /Axis rotation
ORIGIN = 'AIPSLocalhost MACMAR 31DEC08' /Origin of
DATE-OBS= '2007-11-07' /Observation date
TELESCOP= 'VLBA' /Telescope used
OBSERVER= 'BG182' /Observers name
OBJECT = '3C120' /Name of observed source
EPOCH = 2000.000000000000 /Equinox of observations
BSCALE = 1.00000000000000 /Scale factor for axes
BZERO = 0.00000000000000 /Zero offset for array
BUNIT = 'JY/BEAM' /Units measurement
DATAMIN = -0.0019036284647882 /Min data value
DATAMAX = 1.1940656900406 /Max data value
BMAJ = 3.0594406858596E-07 /Clear beam major axis diameter (degrees).
BMIN = 9.1569899501081E-08 /Clear beam minor axis diameter (degrees).
BPA = -19.967098396954 /Clear beam position angle (degrees).
NITER = 1 /Number of model components.
OBSRA = 68.296231371000 /Antenna pointing RA
OBSDEC = 5.354338661000 /Antenna pointing Dec
NOISE = 0.0003533132009 /Theoretical RMS noise estimate
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY EXTEND = T /Extensions may be present
HISTORY /-----
HISTORY /Begin "STOKES" information found in FITS tape header by FITLD
HISTORY EXTEND = T /This is the antenna file
HISTORY BLOCKED = T /Tape may be blocked
HISTORY /-----
HISTORY /Begin "STOKES" information found in FITS tape header by FITLD
HISTORY EXTEND = T /This is the antenna file
HISTORY BLOCKED = T /Tape may be blocked
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY EXTEND = T /All data in tables
HISTORY BLOCKED = T /Tape may be blocked
```

```
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY FITLD EXTEND = T /
HISTORY FITLD BLOCKED = T /
HISTORY FITLD FXCORVER= '4.22' /
HISTORY FITLD OPENED FITS FILE : 2007NOV15 6h31m19.00s
HISTORY FITLD SOFTWARE DIR : /home/fxcorr/code
HISTORY FITLD CORREL SOFTWARE VERSION : 4.22
HISTORY FITLD CORREL ON LINE : Motorola MVME167
HISTORY FITLD VXWORKS VERSION : 5.4
HISTORY FITLD LOG FILE : /home/fxcorr/logs/v47c071115:0019.log
HISTORY FITLD OBSCODE : BG182
HISTORY FITLD JOBNUM : 6721
HISTORY FITLD JOBSTART : 2007NOV07 3h38m46.05s
HISTORY FITLD JOBSTOP : 2007NOV07 4h31m19.00s
HISTORY FITLD FILESTART : 2007NOV07 2h38m46.05s
HISTORY FITLD FILESTOP : 2007NOV07 3h31m19.00s
HISTORY /End FITS tape header "HISTORY" information
HISTORY FITLD RELEASE = '31DEC07' /-----
HISTORY FITLD OUTNAME = 'BG182B' OUTCLASS= 'UVDATA'
HISTORY FITLD OUTSEQ= 1 OUTDISK= 0
HISTORY FITLD RELEASE= '31DEC07' /-----
HISTORY FITLD SCLVIS = 0.54496 / Correlator scaling value
HISTORY FITLD DIGICOR = 1 / VLBA correlator digital corrs
HISTORY FITLD / FFT artifact corrections no longer necessary
HISTORY FITLD RELEASE = '31DEC07' /***** Start 16-NOV-2007 10:27:09
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY FITLD EXTEND = T /
HISTORY FITLD BLOCKED = T /
HISTORY FITLD FXCORVER= '4.22' /
HISTORY FITLD OPENED FITS FILE : 2007NOV15 8h15m22.70s
HISTORY FITLD SOFTWARE DIR : /home/fxcorr/code
HISTORY FITLD CORREL SOFTWARE VERSION : 4.22
HISTORY FITLD CORREL ON LINE : Motorola MVME167
HISTORY FITLD VXWORKS VERSION : 5.4
HISTORY FITLD LOG FILE : /home/fxcorr/logs/v47c071115:0809.log
HISTORY FITLD OBSCODE : BG182
HISTORY FITLD JOBNUM : 6721
HISTORY FITLD JOBSTART : 2007NOV07 3h31m21.08s
HISTORY FITLD JOBSTOP : 2007NOV07 4h36m16.00s
HISTORY FITLD FILESTART : 2007NOV07 3h31m21.08s
HISTORY FITLD FILESTOP : 2007NOV07 4h36m16.00s
HISTORY /End FITS tape header "HISTORY" information
HISTORY FITLD RELEASE = '31DEC07' /-----
HISTORY FITLD OUTNAME = 'BG182B' OUTCLASS= 'UVDATA'
HISTORY FITLD OUTSEQ= 1 OUTDISK= 0
HISTORY FITLD RELEASE= '31DEC07' /-----
HISTORY FITLD SCLVIS = 0.54496 / Correlator scaling value
HISTORY FITLD DIGICOR = 1 / VLBA correlator digital corrs
HISTORY FITLD / FFT artifact corrections no longer necessary
HISTORY FITLD RELEASE = '31DEC07' /***** Start 16-NOV-2007 10:27:22
HISTORY /-----
HISTORY /Begin "HISTORY" information found in FITS tape header by FITLD
HISTORY FITLD EXTEND = T /
HISTORY FITLD BLOCKED = T /
HISTORY FITLD FXCORVER= '4.22' /
HISTORY FITLD OPENED FITS FILE : 2007NOV15 10h33m52.17s
HISTORY FITLD SOFTWARE DIR : /home/fxcorr/code
HISTORY FITLD CORREL SOFTWARE VERSION : 4.22
```

APPLICATION DEPENDENT!

# TRADITIONAL DATA PROVENANCE IN ASTROPHYSICS

---

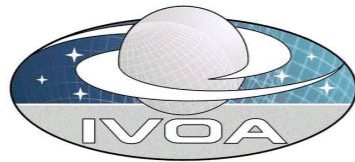
- Application/task based
- No standardisation** → Informational provenance (sometimes, Audit)
- Valid for individual astronomers/small groups, **not for archival data**
- Who queries (cares) about Provenance?

	<b>VO Provenance</b>
<i>Domain</i>	Astronomy & Astrophysics (IVOA)
<i>Processing</i>	Commands; Message Interchange; Workflow Steps; Scripting
<i>Application</i>	Informational; Audit; Error Tracking; Data Regeneration?; Planning
<i>Orientation</i>	Data & Process
<i>Granularity</i>	Scan / Observing Run
<i>Representation</i>	Metadata Annotations; WF steps and parameters
<i>Semantics</i>	Yes
<i>Storage</i>	XML + Web + VOSpace...
<i>Overhead</i>	API calls; Generation of files, links, extra metadata
<i>Scalability</i>	Yes (Mandatory!)
<i>Dissemination</i>	Protocol Metadata

# PROVENANCE'S PROVENANCE: THE OBSERVATION DATA MODEL

# THE OBSERVATION DATA MODEL

---



## Data Model for Observation

Version 0.23

IVOA DM WG Internal Draft

2004-05-16

**Working Group:** <http://www.ivoa.net/twiki/bin/view/IVOA/IvoaDataModel>

**This version:** Added Alberto's introduction.

**Editors:**

Jonathan McDowell, Francois Bonnarel, David Giaretta, Gerard Lemson,  
Mireille Louys, Alberto Micol

**Authors:**

IVOA Data Model Working Group

## Abstract

This document defines the high level Observation data model giving the structure of metadata used to describe the content and context of datasets.

# OVERVIEW OF THE OBSERVATION DM

---

- Scope**

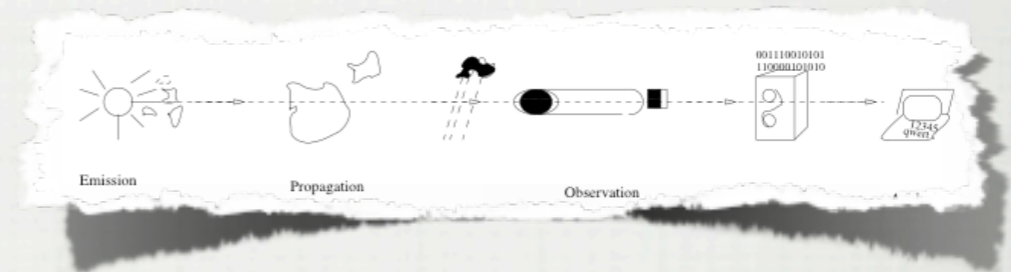
- Describe the complete observation

- Derived Data Models

- Characterisation** (where in parameter space?)

- Provenance

- Target, Curation, Packaging...





# OVERVIEW OF THE OBSERVATION DM

---

- Scope**

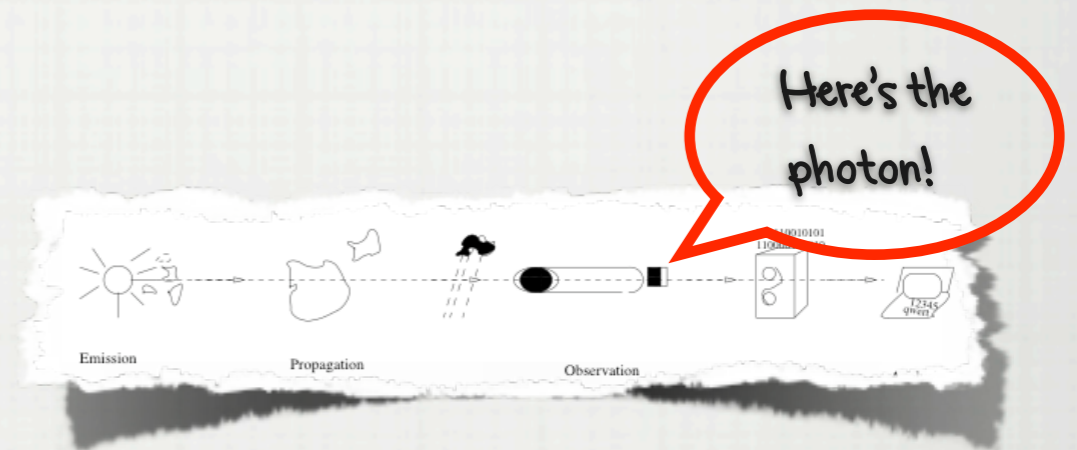
- Describe the complete observation

- Derived Data Models

- Characterisation** (where in parameter space?)

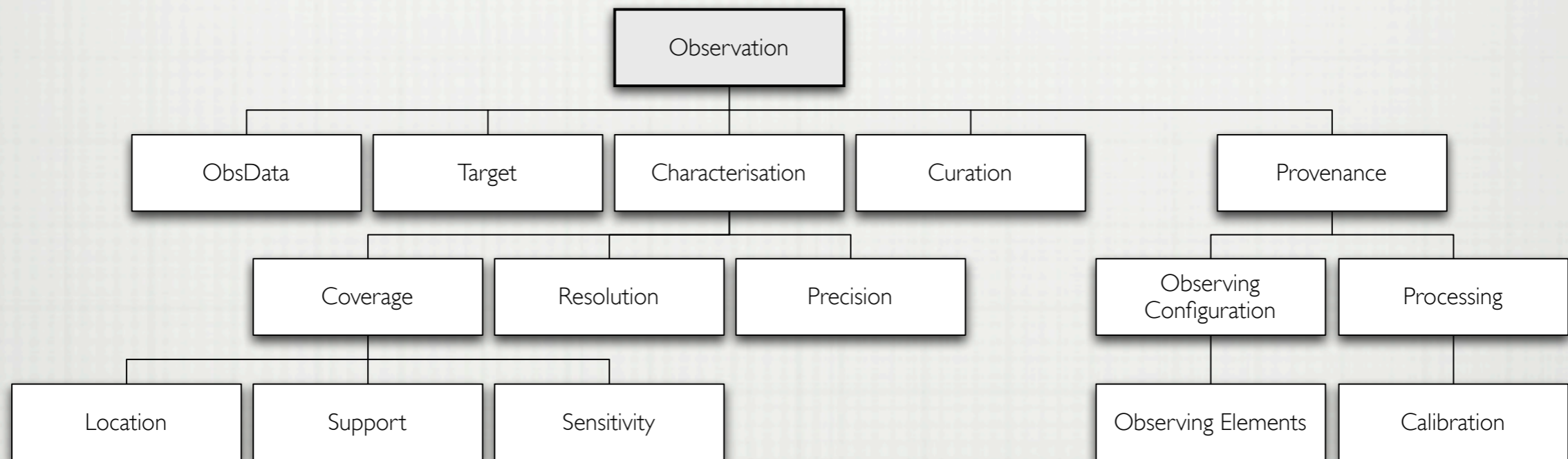
- Provenance

- Target, Curation, Packaging...



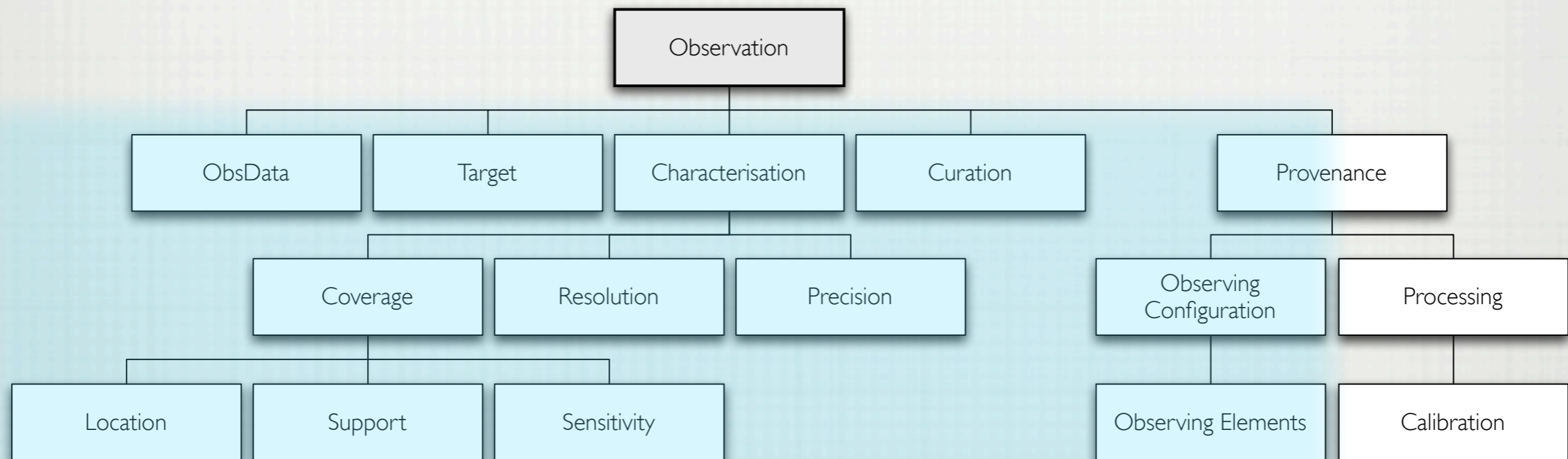
# OBSERVATION DM 2003-2006

---



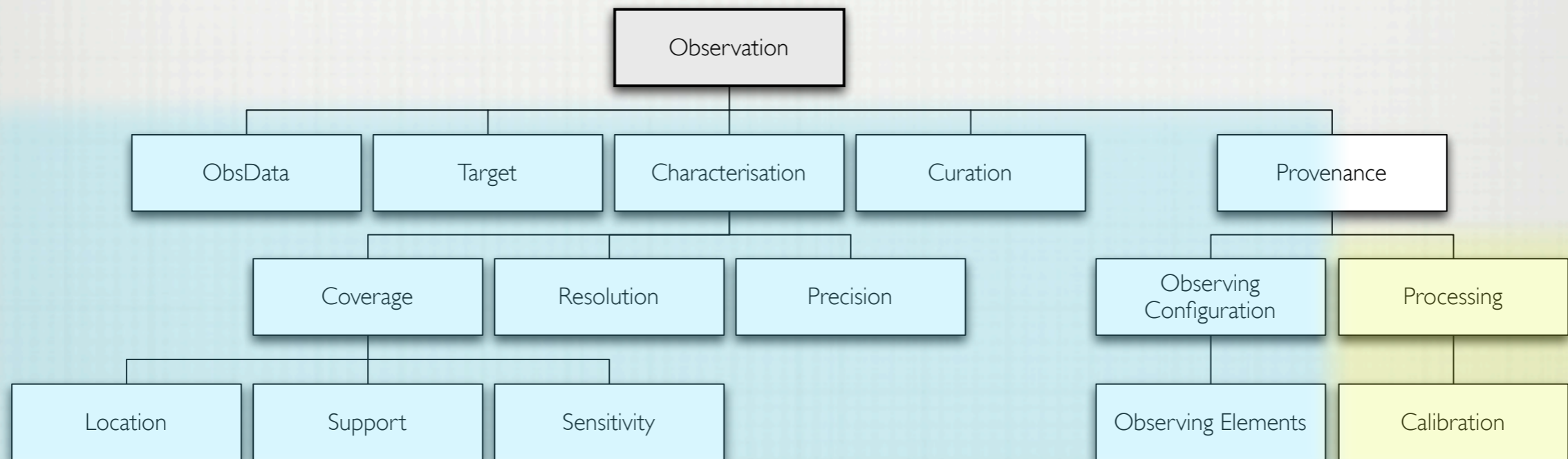
# OBSERVATION DM 2003-2006

---



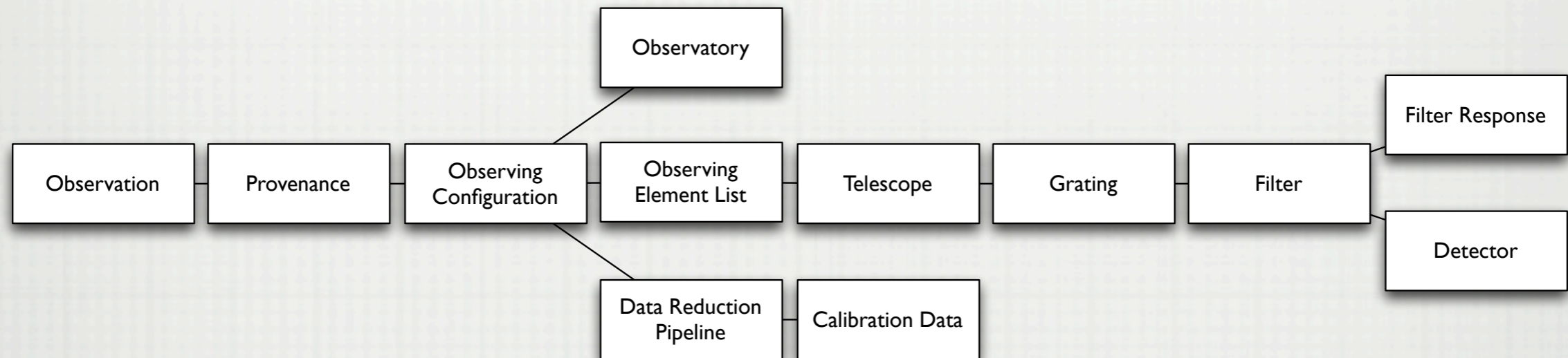
# OBSERVATION DM 2003-2006

---



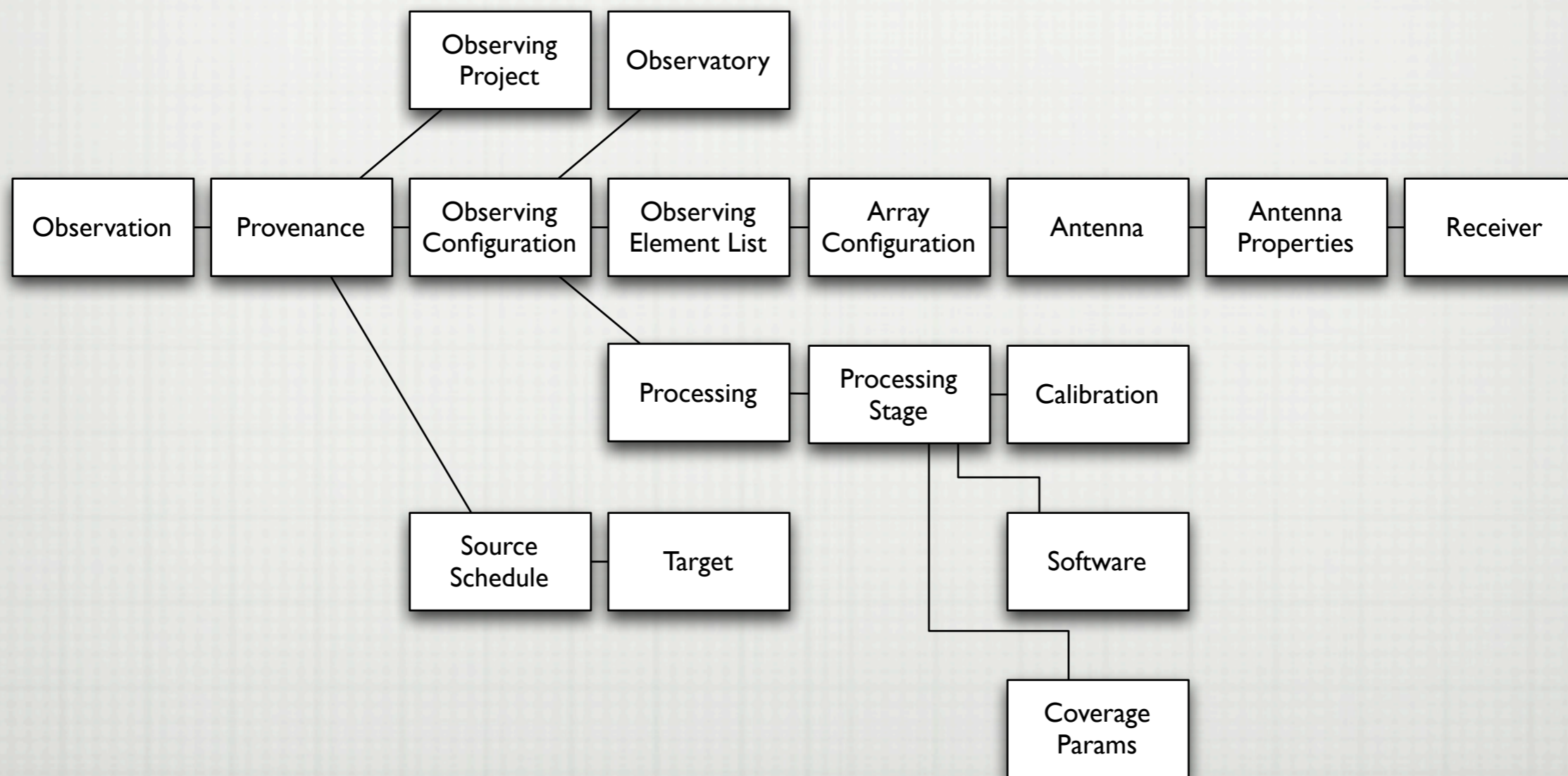
# EARLY PROVENANCE PROPOSALS

---



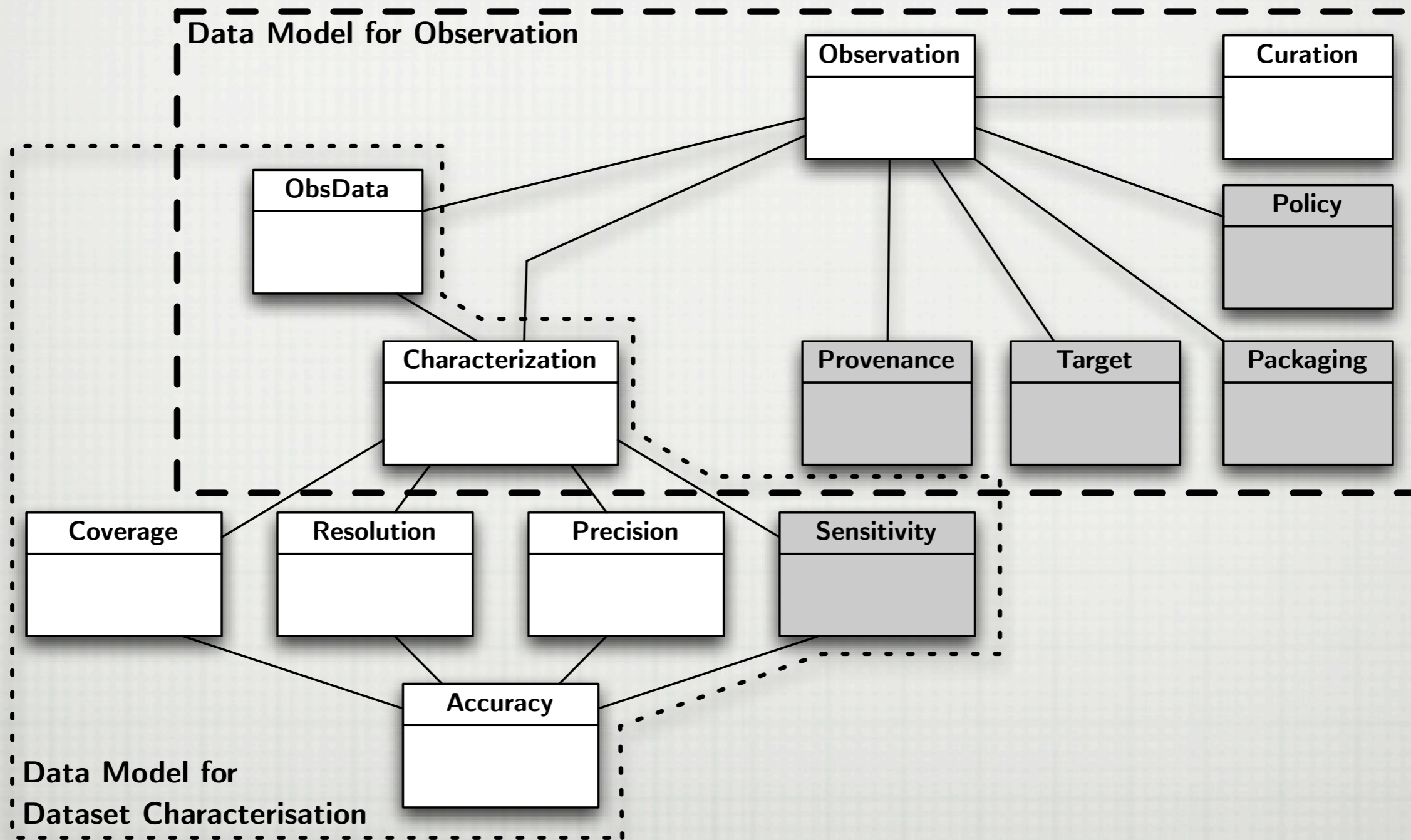
# EARLY PROVENANCE PROPOSALS

---



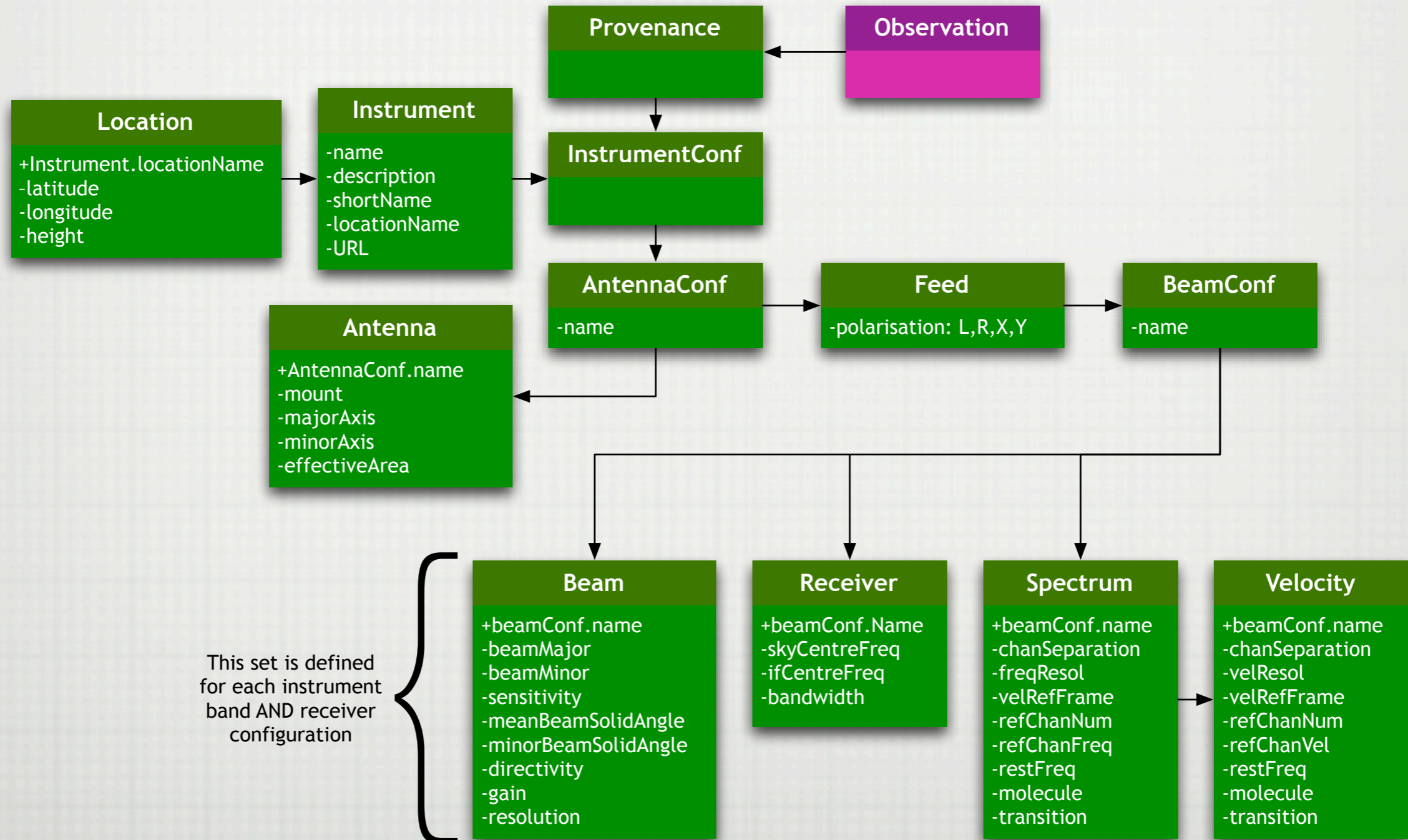
**RADAMS:  
RADIO ASTRONOMY DATA  
MODEL FOR SINGLE-DISH**

# PROVENANCE IN RADAMS



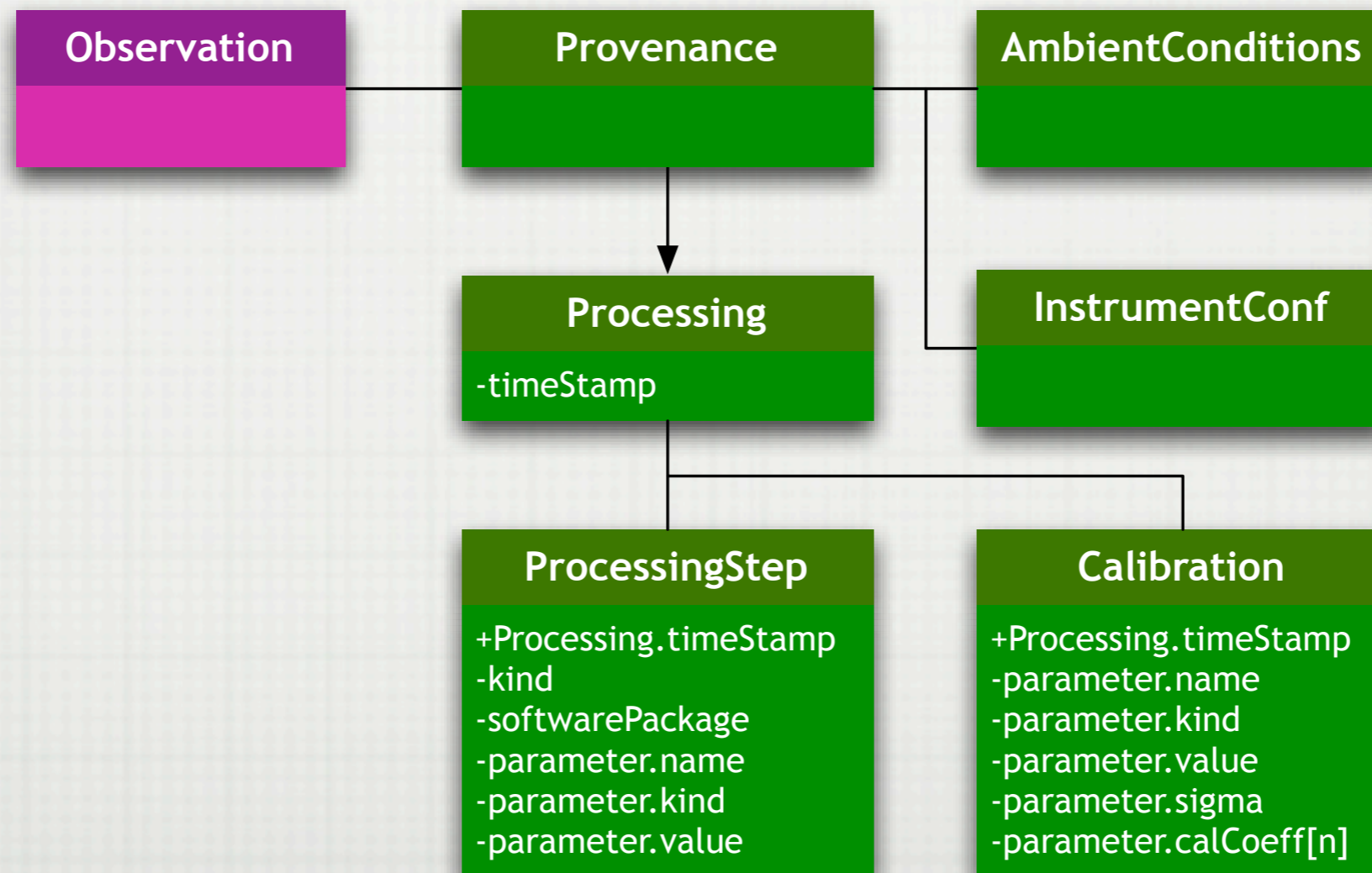


# PROVENANCE IN RADAMS: INSTRCONF



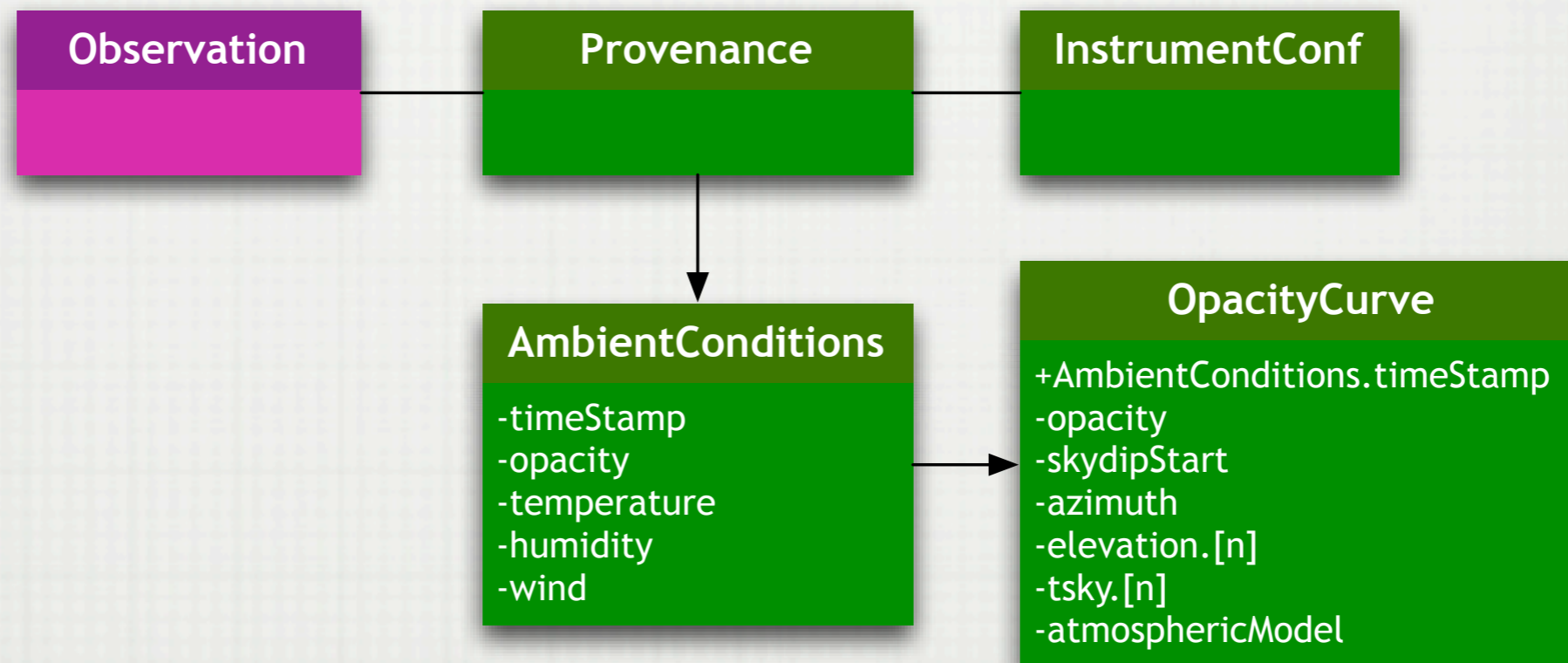
# PROVENANCE IN RADAMS

---

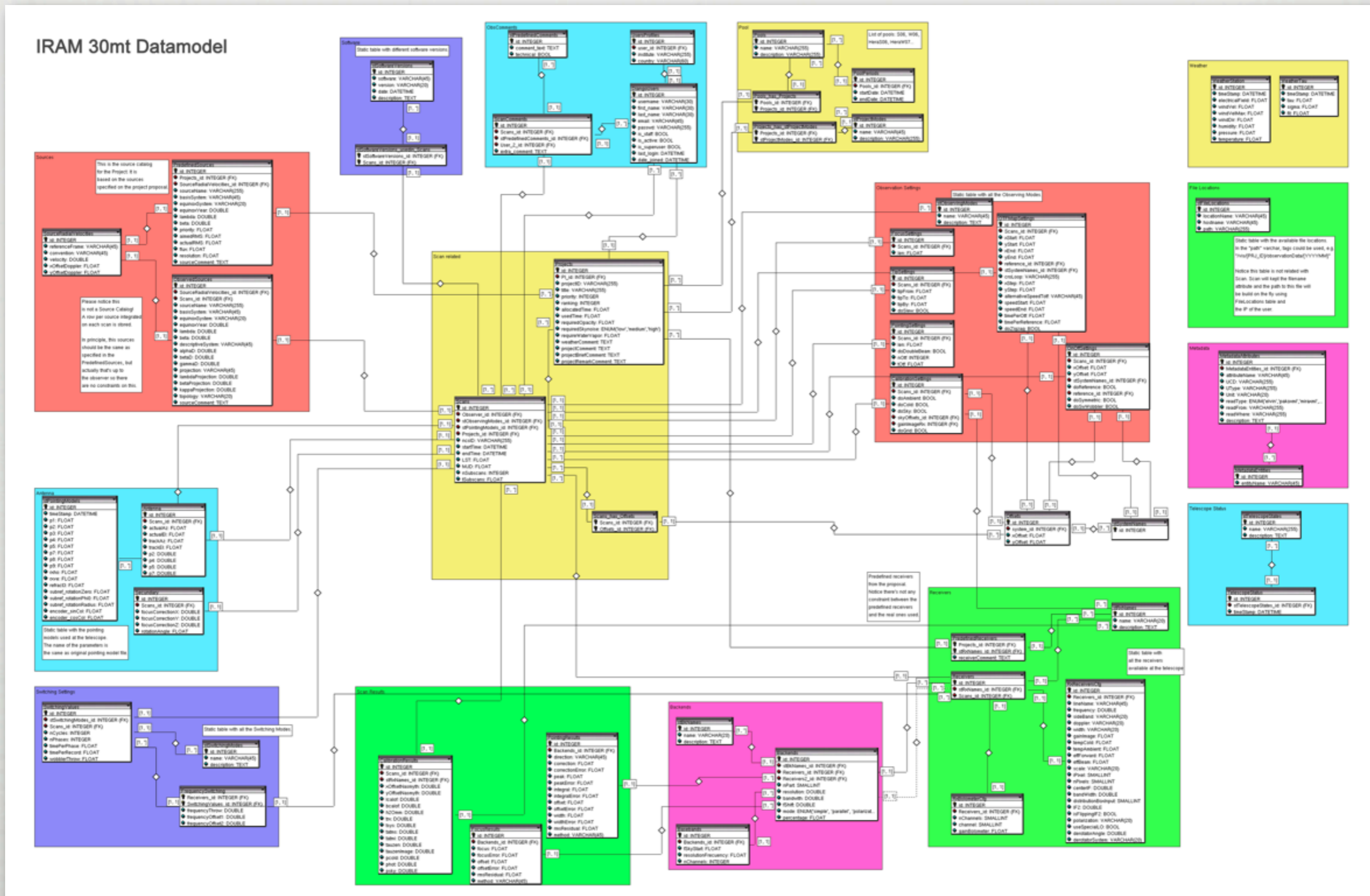


# PROVENANCE IN RADAMS: AMBCONDS

---



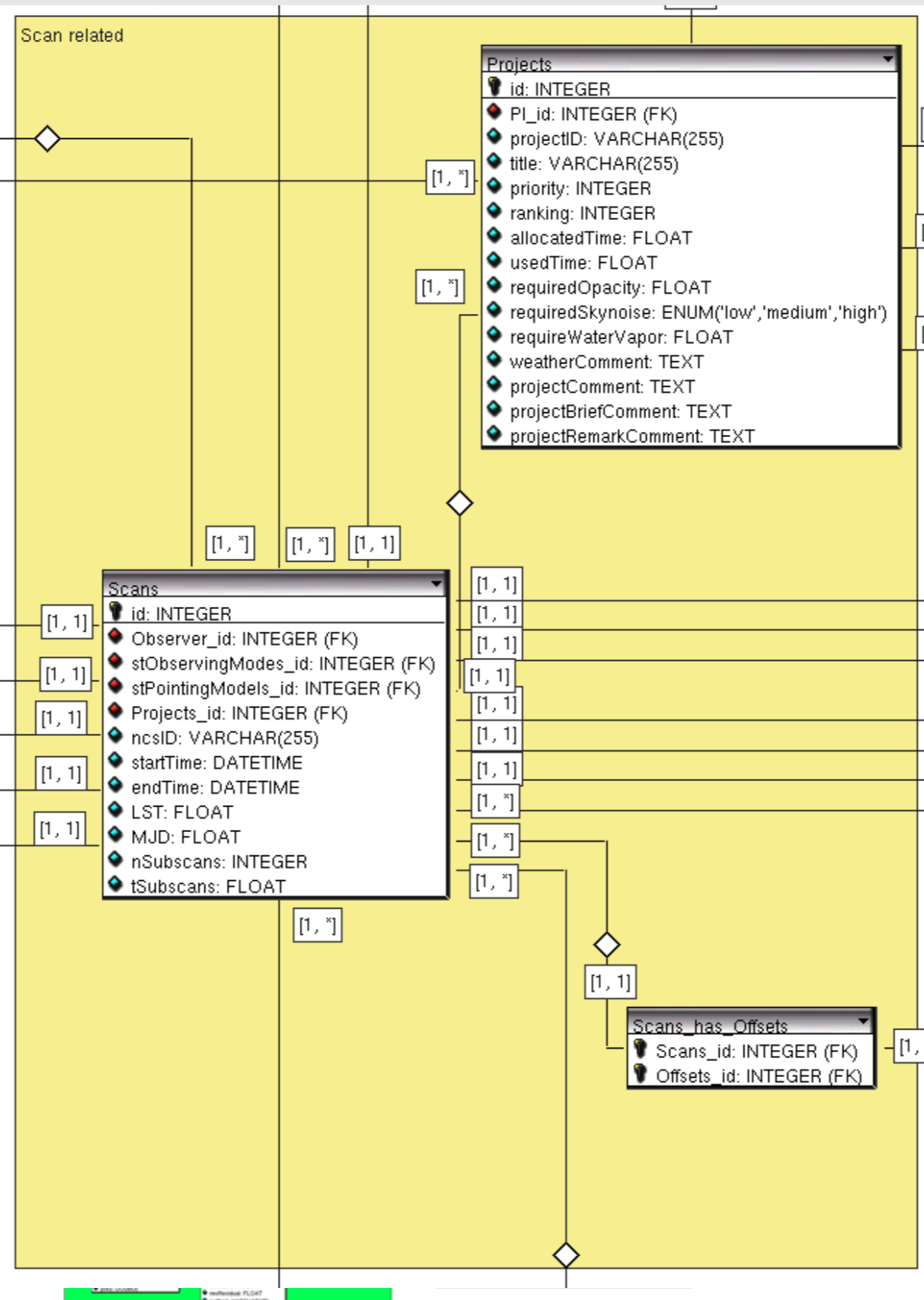
# RADAMS IN PRACTICE: IRAM 30M



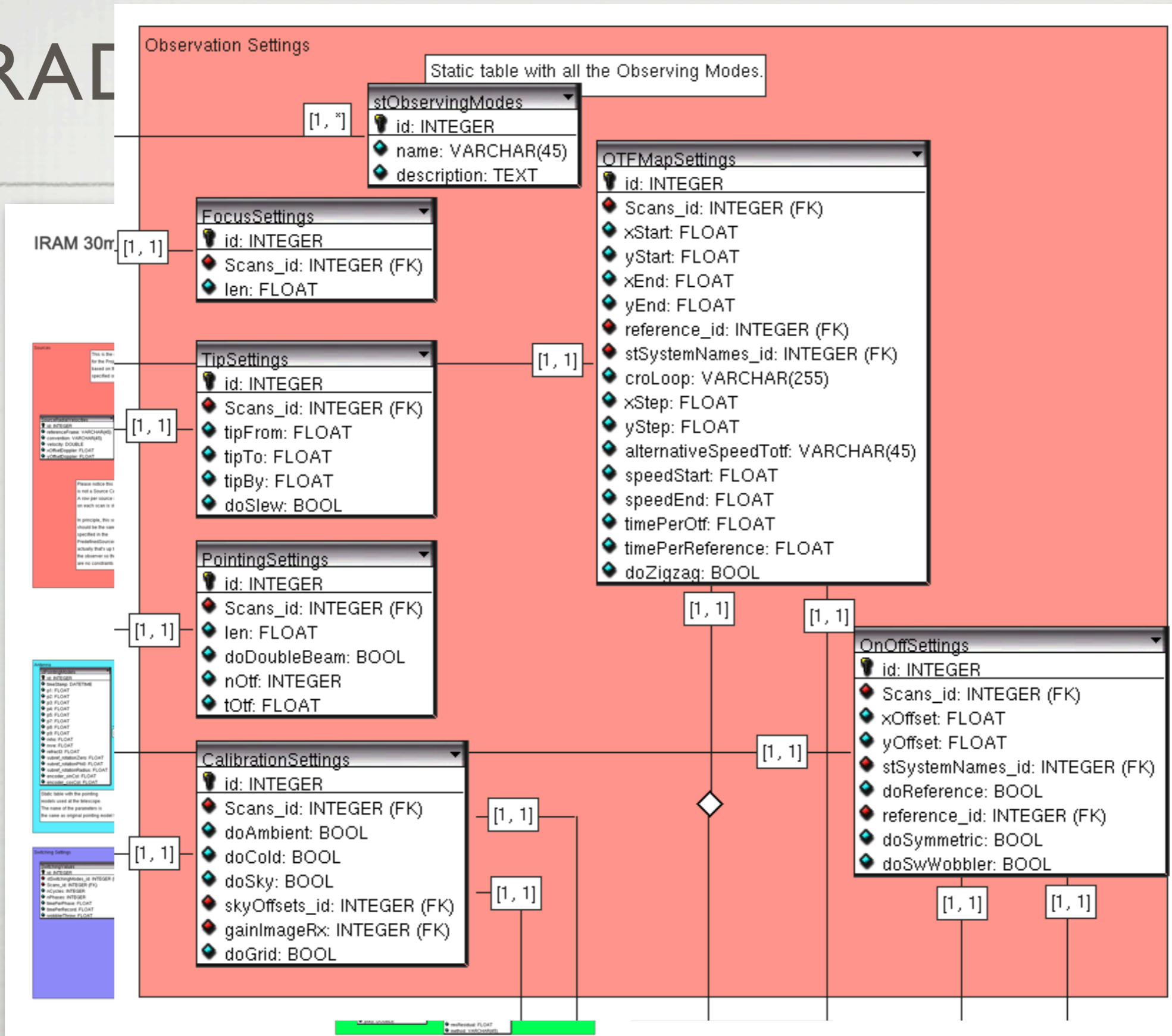
# RADAMS

30M

IRAM 30mt Datamodel



# RAC



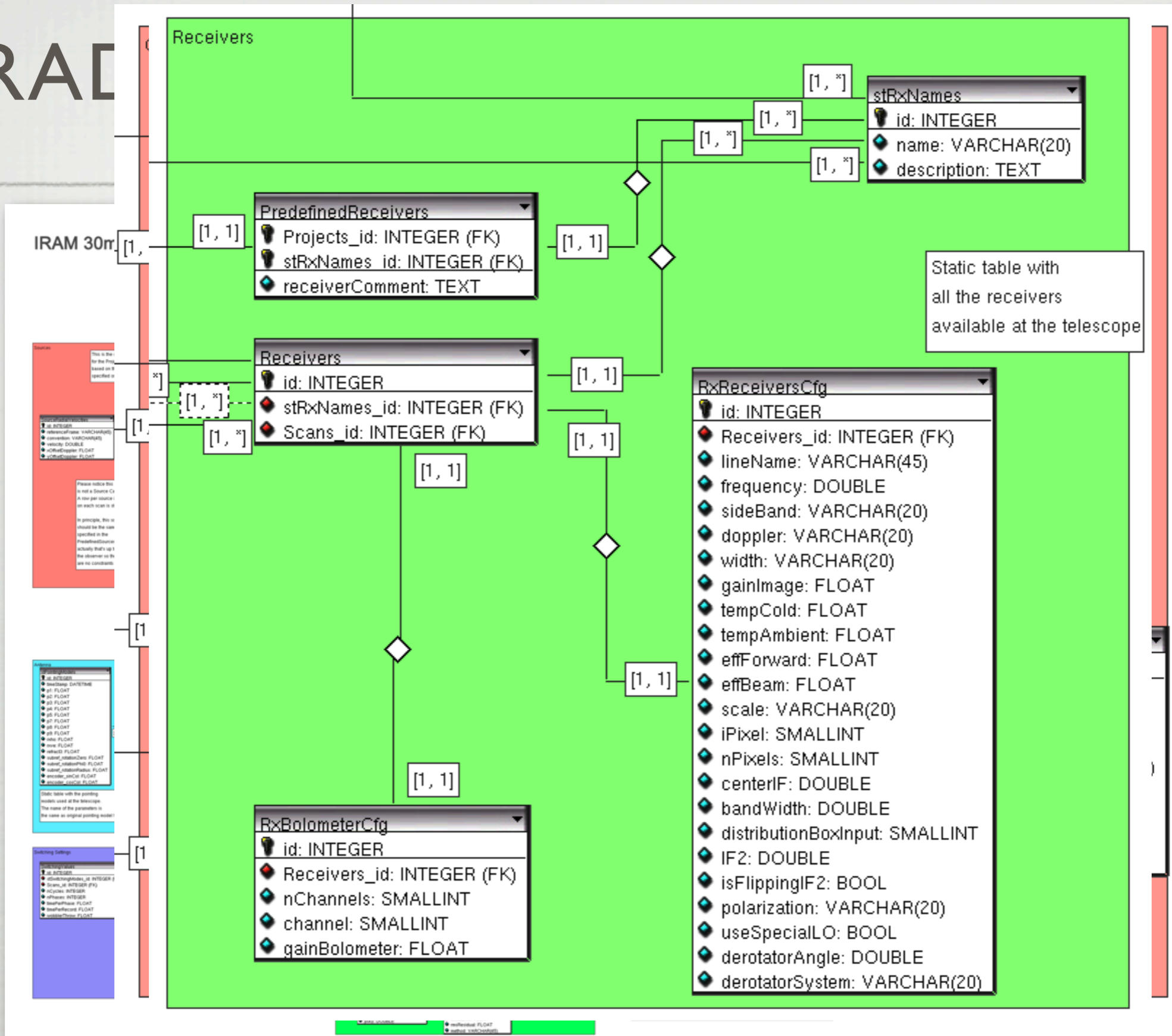
IRAM 30m

This is the for the Pro...  
Please notice this is not a Source C...  
In principle, this is should be the use specified in the...  
actually that's up to the observer as th...  
are no constraints.

Static table with the pointing...  
The name of the parameters in...  
the same as original pointing model!

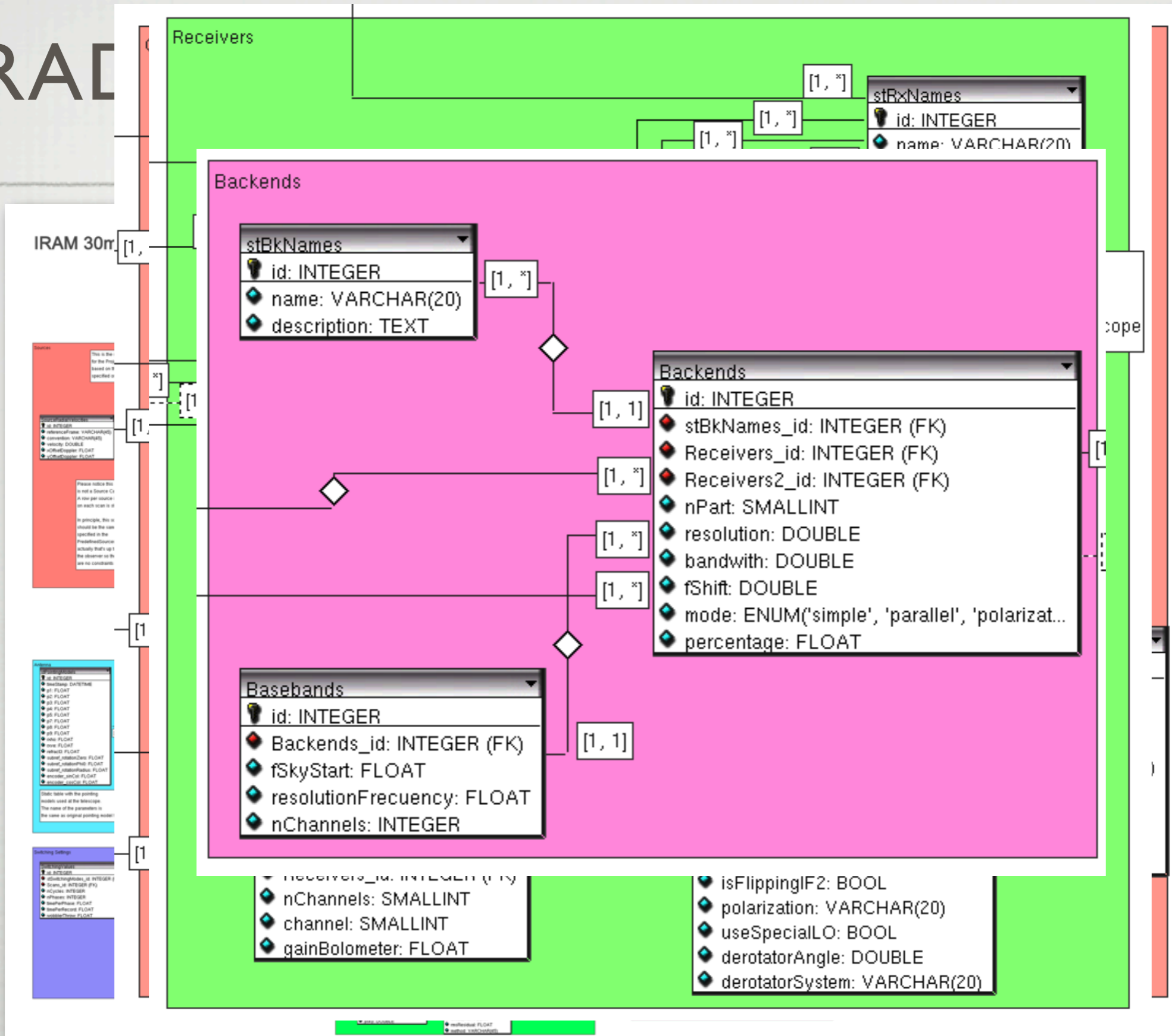
Switching Settings

# RAC



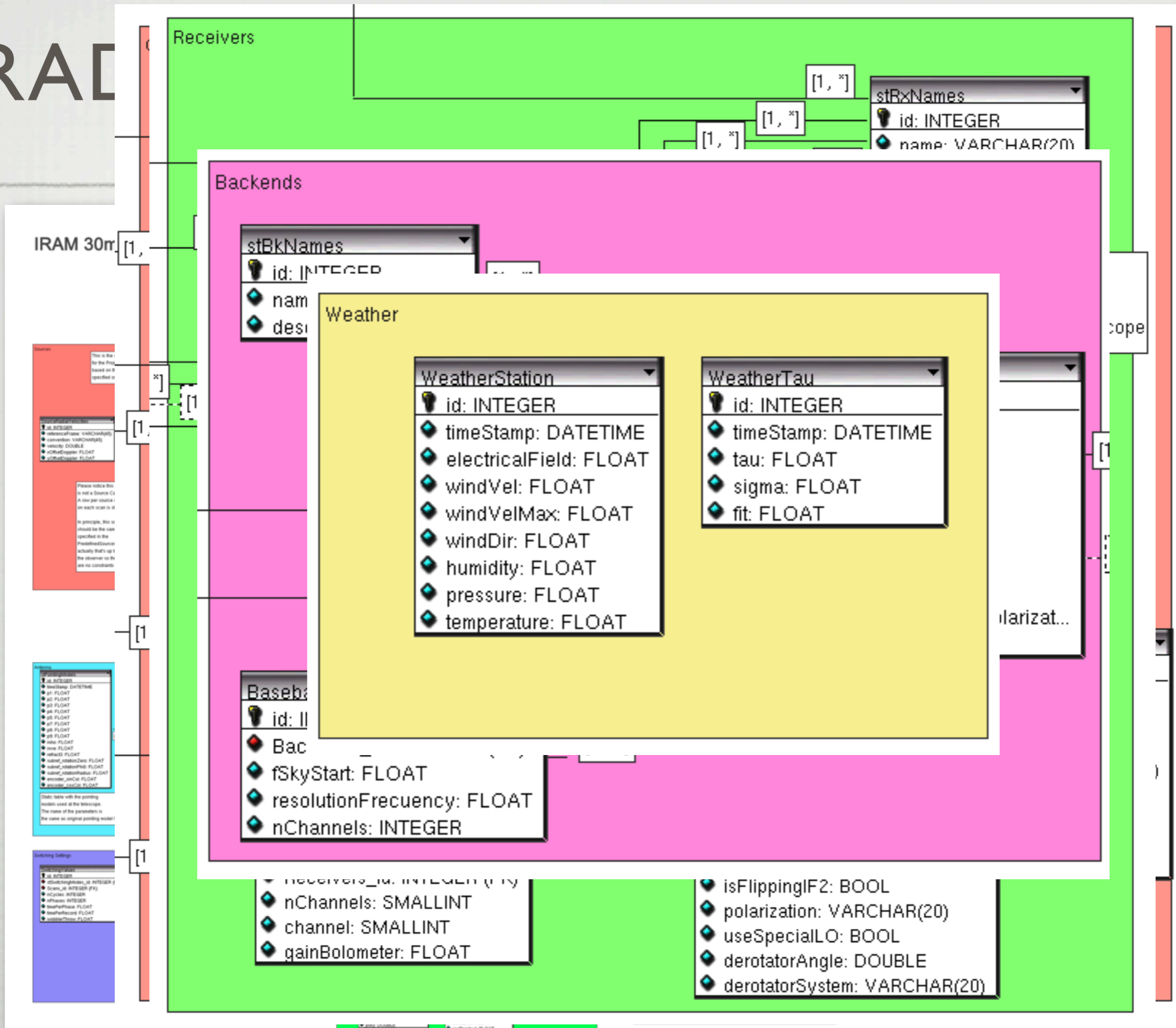
Vertical sidebar containing several small thumbnail images of database-related documents or screenshots.

# RAD

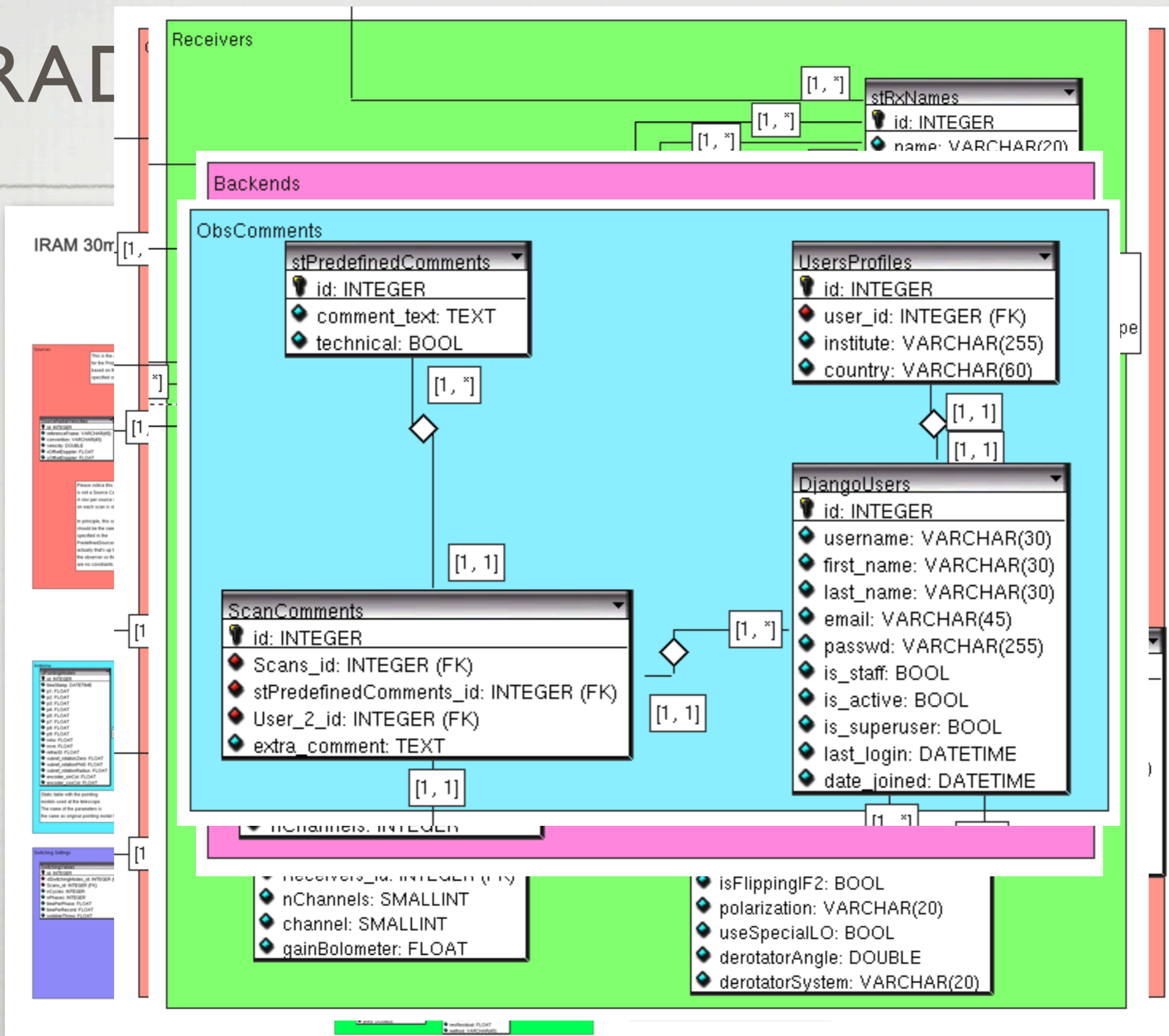




# RAD



# RAC



IRAM 30m [1, \*]

This is the... for the... based on... specified in...  
Please notice this is not a Source CI (A star per source) on each scan is it  
In principle, this is... should be the case specified in the... actually that's up to the observer as there are no constraints

Static table with the pointing models used at the telescope. The name of the parameters in the same as original pointing model!

Switching Settings  
is: INTEGER  
isFlippingIF2: BOOL  
isFlippingIF1: INTEGER (FK)  
isFlippingIF2: INTEGER (FK)  
isFlippingIF3: INTEGER (FK)  
isFlippingIF4: INTEGER (FK)  
isFlippingIF5: INTEGER (FK)  
isFlippingIF6: INTEGER (FK)  
isFlippingIF7: INTEGER (FK)  
isFlippingIF8: INTEGER (FK)  
isFlippingIF9: INTEGER (FK)  
isFlippingIF10: INTEGER (FK)  
isFlippingIF11: INTEGER (FK)  
isFlippingIF12: INTEGER (FK)  
isFlippingIF13: INTEGER (FK)  
isFlippingIF14: INTEGER (FK)  
isFlippingIF15: INTEGER (FK)  
isFlippingIF16: INTEGER (FK)  
isFlippingIF17: INTEGER (FK)  
isFlippingIF18: INTEGER (FK)  
isFlippingIF19: INTEGER (FK)  
isFlippingIF20: INTEGER (FK)  
isFlippingIF21: INTEGER (FK)  
isFlippingIF22: INTEGER (FK)  
isFlippingIF23: INTEGER (FK)  
isFlippingIF24: INTEGER (FK)  
isFlippingIF25: INTEGER (FK)  
isFlippingIF26: INTEGER (FK)  
isFlippingIF27: INTEGER (FK)  
isFlippingIF28: INTEGER (FK)  
isFlippingIF29: INTEGER (FK)  
isFlippingIF30: INTEGER (FK)  
isFlippingIF31: INTEGER (FK)  
isFlippingIF32: INTEGER (FK)  
isFlippingIF33: INTEGER (FK)  
isFlippingIF34: INTEGER (FK)  
isFlippingIF35: INTEGER (FK)  
isFlippingIF36: INTEGER (FK)  
isFlippingIF37: INTEGER (FK)  
isFlippingIF38: INTEGER (FK)  
isFlippingIF39: INTEGER (FK)  
isFlippingIF40: INTEGER (FK)  
isFlippingIF41: INTEGER (FK)  
isFlippingIF42: INTEGER (FK)  
isFlippingIF43: INTEGER (FK)  
isFlippingIF44: INTEGER (FK)  
isFlippingIF45: INTEGER (FK)  
isFlippingIF46: INTEGER (FK)  
isFlippingIF47: INTEGER (FK)  
isFlippingIF48: INTEGER (FK)  
isFlippingIF49: INTEGER (FK)  
isFlippingIF50: INTEGER (FK)  
isFlippingIF51: INTEGER (FK)  
isFlippingIF52: INTEGER (FK)  
isFlippingIF53: INTEGER (FK)  
isFlippingIF54: INTEGER (FK)  
isFlippingIF55: INTEGER (FK)  
isFlippingIF56: INTEGER (FK)  
isFlippingIF57: INTEGER (FK)  
isFlippingIF58: INTEGER (FK)  
isFlippingIF59: INTEGER (FK)  
isFlippingIF60: INTEGER (FK)  
isFlippingIF61: INTEGER (FK)  
isFlippingIF62: INTEGER (FK)  
isFlippingIF63: INTEGER (FK)  
isFlippingIF64: INTEGER (FK)  
isFlippingIF65: INTEGER (FK)  
isFlippingIF66: INTEGER (FK)  
isFlippingIF67: INTEGER (FK)  
isFlippingIF68: INTEGER (FK)  
isFlippingIF69: INTEGER (FK)  
isFlippingIF70: INTEGER (FK)  
isFlippingIF71: INTEGER (FK)  
isFlippingIF72: INTEGER (FK)  
isFlippingIF73: INTEGER (FK)  
isFlippingIF74: INTEGER (FK)  
isFlippingIF75: INTEGER (FK)  
isFlippingIF76: INTEGER (FK)  
isFlippingIF77: INTEGER (FK)  
isFlippingIF78: INTEGER (FK)  
isFlippingIF79: INTEGER (FK)  
isFlippingIF80: INTEGER (FK)  
isFlippingIF81: INTEGER (FK)  
isFlippingIF82: INTEGER (FK)  
isFlippingIF83: INTEGER (FK)  
isFlippingIF84: INTEGER (FK)  
isFlippingIF85: INTEGER (FK)  
isFlippingIF86: INTEGER (FK)  
isFlippingIF87: INTEGER (FK)  
isFlippingIF88: INTEGER (FK)  
isFlippingIF89: INTEGER (FK)  
isFlippingIF90: INTEGER (FK)  
isFlippingIF91: INTEGER (FK)  
isFlippingIF92: INTEGER (FK)  
isFlippingIF93: INTEGER (FK)  
isFlippingIF94: INTEGER (FK)  
isFlippingIF95: INTEGER (FK)  
isFlippingIF96: INTEGER (FK)  
isFlippingIF97: INTEGER (FK)  
isFlippingIF98: INTEGER (FK)  
isFlippingIF99: INTEGER (FK)  
isFlippingIF100: INTEGER (FK)

# REVIEW & CONCLUSIONS

# REVIEW

---

- Provenance is not easy
- Provenance is needed
- Provenance is not only a Data Model thing
- We have to show how to deal with literature provenance, and with application provenance.

# CONCLUSIONS

---

- Personal View: Data Provenance is a **huge need** of the VO
  - Vehicle for dissemination of the VO
- We have a basic agreement on the basic things
- Perhaps we should be/keep pragmatic → Collaboration opportunity DAL/apps WGs

**THANK YOU!**